

ARTICLE

DOI: 10.1038/s41467-017-02009-x

OPEN

# A very large-scale microelectrode array for cellular-resolution electrophysiology

David Tsai<sup>1</sup>, Daniel Sawyer<sup>1</sup>, Adrian Bradd<sup>1</sup>, Rafael Yuste<sup>2</sup> & Kenneth L. Shepard<sup>3</sup>

In traditional electrophysiology, spatially inefficient electronics and the need for tissue-to-electrode proximity defy non-invasive interfaces at scales of more than a thousand low noise, simultaneously recording channels. Using compressed sensing concepts and silicon complementary metal-oxide-semiconductors (CMOS), we demonstrate a platform with 65,536 simultaneously recording and stimulating electrodes in which the per-electrode electronics consume an area of 25.5  $\mu\text{m}$  by 25.5  $\mu\text{m}$ . Application of this platform to mouse retinal studies is achieved with a high-performance processing pipeline with a 1 GB/s data rate. The platform records from 65,536 electrodes concurrently with a  $\sim 10$   $\mu\text{V}$  r.m.s. noise; senses spikes from more than 34,000 electrodes when recording across the entire retina; automatically sorts and classifies greater than 1700 neurons following visual stimulation; and stimulates individual neurons using any number of the 65,536 electrodes while observing spikes over the entire retina. The approaches developed here are applicable to other electrophysiological systems and electrode configurations.

---

<sup>1</sup>Department of Electrical Engineering, Columbia University, New York, NY 10027, USA. <sup>2</sup>Departments of Biological Sciences and Neuroscience, Columbia University, New York, NY 10027, USA. <sup>3</sup>Departments of Electrical and Biomedical Engineering, Columbia University, New York, NY 10027, USA. Correspondence and requests for materials should be addressed to K.L.S. (email: [shepard@ee.columbia.edu](mailto:shepard@ee.columbia.edu))

The ability to observe and manipulate, with single-cell precision, the activities of large neuronal populations is an essential step toward better understanding of the nervous system<sup>1</sup>. It is now possible to observe activities in large neuronal populations through calcium imaging<sup>2–4</sup>. However, intracellular calcium concentration is influenced by many simultaneously active and mutually interacting mechanisms, including voltage- and ligand-gated calcium channels, intracellular stores, and chelation by proteins and by the fluorescent calcium indicator<sup>5</sup>. Many of these are not under direct control of the experimentalist and operate over a time scale substantially longer than that of action potentials, thus complicating interpretation. In particular, inferring action potentials from calcium signals require non-trivial calibration<sup>6</sup>, particularly if one is interested in precise spike timing and not just the number of spikes.

Extracellular electrophysiology, in contrast, enables direct read out of spikes. Recent progress has substantially increased the number of simultaneously recordable neurons<sup>7</sup>. However, the observable neurons, ranging from few tens to approximately a thousand<sup>8–10</sup>, remains small relative to the number of neurons in the brain. This is primarily due to trade-offs between several competing requirements: adequate recording signal-to-noise ratio (SNR), minimal biological invasiveness, at-scale recording, and, ideally, ability to stimulate the recorded neurons with precision.

Because of spatial requirements, having a complete amplifier chain and digitizer for each electrode, with sufficient noise performance, bandwidth and dynamic range, is infeasible for at-scale electrophysiology. Consequently, time-division multiplexing has been the mainstay for increasing density and channel count<sup>9,11,12</sup>, by sharing one signal path between several inputs. Traditional time-division multiplexing constitutes a sampled system. Since the input channels are sequentially scanned (sampled) over time, per-channel low-pass filters are required to reject frequencies above half the sampling rate (Supplementary Fig. 1a). Failure to do so causes aliasing of frequencies above half the per-channel sampling rate, thus degrading the SNR. Moreover, the more channels an inadequately filtered system has, the greater the SNR degradation.

The physical dimension of these filters dictates the density limit of large-scale electrophysiology. With an ideal noise ceiling of approximately 10  $\mu\text{V}$  rms<sup>13</sup>, a parsimonious, multiplexed recorder, consisting of the electrode, an antialiasing filter and an amplifier, cannot be smaller than  $\sim 10,000 \mu\text{m}^2$  to ensure adequate noise performance, due to the capacitance density available in today's microelectronic technologies (Supplementary Fig. 1b; also note Supplementary Fig. 1 caption). With mammalian neuronal soma  $\leq 25 \mu\text{m}$  in diameter, this 20-to-1 dimensional mismatch fundamentally restricts the scalability of current approaches in electrophysiology. Due to this noise-verse-density trade-off, today's large-scale neural recorders have either low noise but low simultaneously recording channel count<sup>8,14</sup> or high channel count but high noise<sup>11,15,16</sup>.

Here we present a large-scale, high-density and low noise electrophysiological recording and stimulation platform based on CMOS electronics and an acquisition paradigm that negates the requirement for per-channel antialiasing filters, thereby overcoming scaling limitations faced by existing systems. This allows us to maintain 10  $\mu\text{V}$  rms recording noise with per-channel electronics for 65,536 channels at a 25.5- $\mu\text{m}$  electrode pitch, avoiding the common trade-off between density, channel count and noise<sup>8,11,14–19</sup>. We then demonstrate the platform's ability to record from 10 s of thousands of neurons simultaneously. In conjunction with visual stimuli and the platform's high-performance computing infrastructure, the system could functionally classify more than 1700 neurons in the retina automatically, including identification of rare cell types. Finally,

combining cellular-resolution microstimulators with dense recordings across the entire retina allowed us to re-examine how electrical stimulation recruits neurons in a network, and importantly, how focal activation could be achieved.

## Results

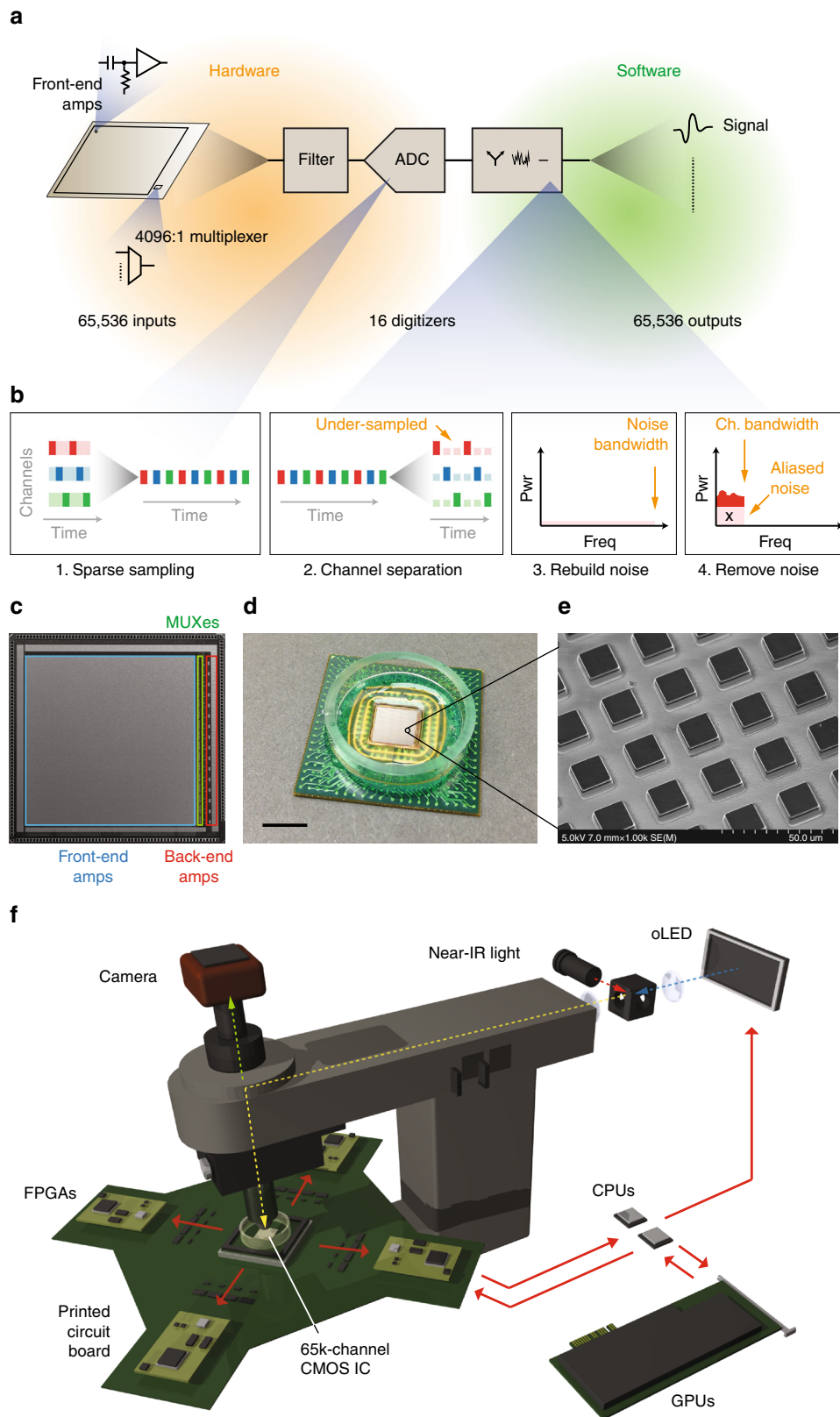
**Compressed sensing-inspired electrophysiology.** We constructed a 65,536-channel, multiplexed, extracellular electrophysiology system consisting of a recording and stimulation array based on a custom integrated circuit (IC), circuit-board-level analog and digital circuits and custom software libraries (Fig. 1a). Traditional implementations of such multiplexed systems incorporate per-channel low-pass filters to prevent noise aliasing (Supplementary Fig. 1a). Our approach (Fig. 1b, Supplementary Fig. 1b) does not use these filters, but instead, avoids aliasing using concepts from compressed sensing<sup>20,21</sup>. This is made possible by noting several characteristics of extracellular recordings and thermal noise aliasing.

First, electrophysiological recording is dominated by thermal noise at frequencies above a few kHz. It is a stationary process with a Gaussian time-domain amplitude distribution and uniform frequency distribution, up to the recording channel's bandwidth. Therefore, this thermal noise can be described, and generated computationally, with only two parameters, its variance and bandwidth. Second, thermal noise aliasing offers two averaging properties, which greatly simplify the reconstruction, and subsequent removal, of its spectral contribution in the under-sampled, per-channel data. The power of thermal noise is approximately uniform. As the thermal noise powers are folded down into the first Nyquist zone (Supplementary Fig. 3b) during aliasing, the slight variations in power between frequencies are averaged out. This allows us to compute the power contributed by aliasing using the expected average thermal noise power, multiplied by the number of folded Nyquist zones. Similarly, the spectral angles of thermal noise have a uniform distribution with zero mean. The angle variation between frequencies converge to zero as the aliased thermal noise are folded down into the first Nyquist zone.

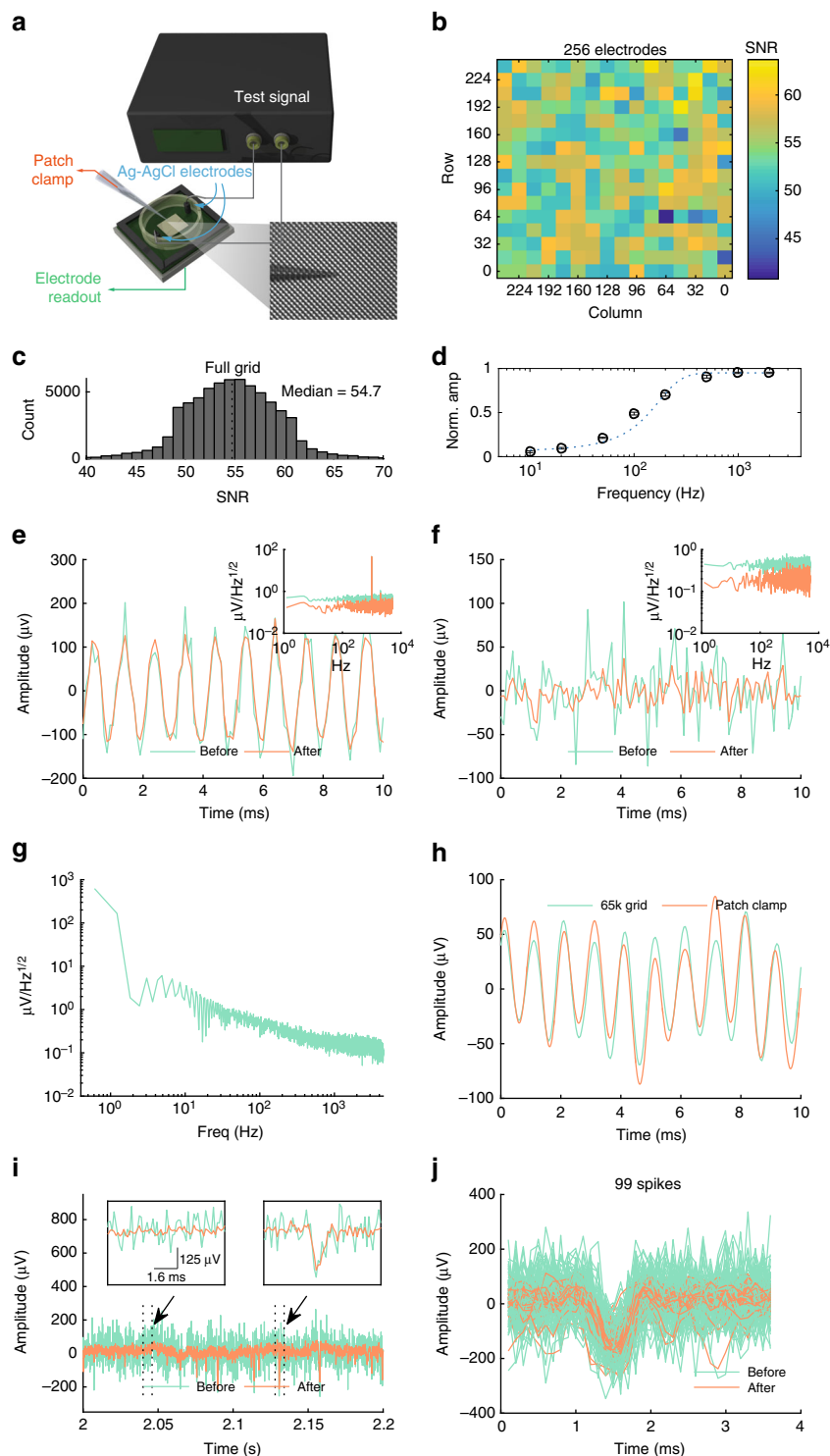
Taking advantage of the foregoing characteristics, we can digitally reconstruct the spectral contributions originating from the under-sampled thermal noise, then remove them from the sparsely-sampled channel data, thereby minimizing the effects of aliasing, without using per-channel anti-aliasing filters.

This acquisition strategy allows us to pack 65,536 channels (Fig. 1c, d) into an area of 42.6 mm<sup>2</sup>, with 25.5  $\mu\text{m}$  spacing between channels (Fig. 1e), using CMOS IC processes. Each channel can be sampled at 10 kHz during full-grid recordings, with higher sampling rates achievable by reducing the recording area. Importantly, this platform does not have the noise-verse-density trade-off of classical large-scale electrophysiology. We constructed an electrophysiological platform based on this acquisition paradigm. It consists of the aforementioned 65,536-channel CMOS IC, custom circuit boards with filters and field programmable gate arrays (FPGAs), CPUs, graphical processing units (GPUs), and an OLED display for generating visual patterns (Fig. 1f).

**Achieving 65,536-channel recordings with minimal noise.** To test the recording performance, we applied test signals through a pair silver-silver chloride electrodes into the recording chamber filled with physiological saline (Fig. 2a). The median SNR across the array was 54.9 with 200  $\mu\text{V}$  test signals (Fig. 2b, c). Our system uses a capacitive recording interface<sup>15,22,23</sup>, formed by a 6-nm thick HfO<sub>2</sub> dielectric deposited above each electrode and a pseudo-resistor constructed from a p-type MOSFET. The corner



**Fig. 1** Platform for dense, large-scale electrophysiology. **a** Overview of acquisition paradigm based on compressed sensing. **b** Sparse sampling by the ADC, followed by digital reconstruction and removal of the spectral contribution by the aliased thermal noise. **c** A 65,536-channel recording and microstimulation grid based on CMOS-integrated circuits (IC). **d** Packaged IC. **e** Scanning electron micrograph of the  $14 \times 14 \mu\text{m}$  electrodes, spaced  $25.5 \mu\text{m}$  apart. **f** Illustration of the experimental platform consisting of: the IC; supporting circuits; data processing pipeline containing CPUs and GPUs; and optics for near-infrared visualization and visual pattern delivery. ADC, analog-to-digital-converter; MUX, multiplexer



**Fig. 2** System performance characterization. **a** Test setup. Recordings from the 65,536-channel grid are compared against patch clamp recordings. Inset photo: pipette above the grid. **b** SNR variation across the array, measured at every sixteenth row and column. The test signal was a bath-applied 1-kHz, 200  $\mu\text{V}$  sine wave. **c** SNR distribution for all electrodes in the array. **d** Frequency response of the capacitive recording front-end (mean  $\pm$  SEM, 8 electrodes). **e, f** Comparison of per-channel data, before (green) and after (orange) removal of aliased thermal noise, for a recording with 1 kHz signal **e** and a baseline recording **f**. Insets in **e** and **f**: PSD plots of time-domain data in **e** and **f**, respectively. The dc component has been removed to better illustrate the linear reduction in noise floor across frequencies. **g** The system's input referred noise was  $\sim 10$   $\mu\text{V}$  rms over 100–10k Hz, in saline. **h** Comparison to patch clamp amplifier recordings. The test signal was a 1 kHz, 100  $\mu\text{V}$  peak-to-peak sine wave. Both traces have been bandpass filtered between 300–3k Hz for clarity. **i** Recordings before and after removing aliased thermal noise. Inset: expanded view of segment without and with spike, respectively. **j** Overlaid traces for 99 spike segments before and after removing aliased noise. Traces in **i** and **j** are unfiltered

frequency is user-tuneable, and is nominally set to 100 Hz (Fig. 2d).

Each channel is typically observed at a rate well below the channel bandwidth. To minimize thermal noise aliasing, the spectral contribution of the under-sampled thermal noise is computed, then removed, from the channel data (Fig. 1b, steps 3–4). Figure 2e compares, for a 1 kHz signal applied in the saline-filled recording bath, the channel data before (green) and after (orange) removal of the aliased thermal noise. The SNR improvement is also apparent in the spectral domain (Fig. 2e inset). The noise floor was reduced uniformly across frequencies. Fig. 2f illustrates the effects of our processing strategy on a segment of baseline recording without test signal. Here, the noise was reduced from 21.7  $\mu\text{V}$  rms (green) to 10.02  $\mu\text{V}$  rms (orange) over the 100–3 kHz bandwidth following signal processing.

When recording in physiological saline, the 65k-electrode grid had  $\sim 10\ \mu\text{V}$  rms input referred noise over the 100–5 kHz bandwidth, encompassing the spike frequency range of 300–3 kHz (Fig. 2g). Finally, the per-channel signal from the 65k-electrode grid closely resembled those from patch clamp recordings—the gold standard in electrophysiology—performed adjacent to the test electrode (Fig. 2h). For clarity, both traces were bandpass filtered between 300–3 kHz. The low-frequency fluctuation was due to noise pick-up by the wires connecting the signal generator to the bath electrodes.

To assess the performance of the de-noising procedure on biological recordings, Fig. 2i compares a recording before (green) and after (orange) removal of aliased noise. In particular, the inset plots in Fig. 2i show that the procedure reduced noise fluctuations without degrading the action potential waveforms. This is further illustrated in Fig. 2j for 99 spikes recorded from one electrode. The average variance of these raw waveforms is 8307.6. After removing the aliasing-induced spurious fluctuations, the average variance of the processed waveforms is 3179.0. While the amplitudes of signal and noise are both reduced, the SNR is improved by reducing the spectral contribution of aliased noise from the first Nyquist zone (Supplementary Fig. 3b versus Supplementary Fig. 3c).

Collectively, these results demonstrate the ability of this system to acquire, with high SNR, weak signals having amplitudes typical of mammalian extracellular recording, and to do so at spatial resolutions down to 25.5  $\mu\text{m}$ , while simultaneously providing observable spatial coverage of 42.6  $\text{mm}^2$  with 65,536 electrodes.

### Simultaneous recordings from more than 34,000 electrodes.

We tested the ability of the platform to carry out at-scale, cellular-resolution recordings with single-spike sensitivity by placing a piece of mouse retina, retinal ganglion cell (RGC) side down, on the recording grid (Fig. 3a). We began by observing the neurons' spontaneous activities under scotopic conditions (Fig. 3b). Spikes were readily apparent, with 34,187 electrodes picking up spiking activities. This exceeded the best existing attempts at across-retina spike recordings by an order of magnitude in channel count<sup>24</sup> and the best calcium imaging efforts in the retina by approximately two orders of magnitude<sup>25,26</sup>.

Spikes from each neuron were observed on multiple adjacent electrodes (Fig. 3c, Supplementary Fig. 4e) and each electrode acquired spikes from more than one neuron (Fig. 3d, Supplementary Fig. 4c). Spike sorting accuracy is substantially improved by combining spatially dependent waveforms from adjacent electrodes<sup>27</sup> (Supplementary Fig. 4), made possible by the dense 25.5  $\mu\text{m}$  electrode pitch. Figure 3d, e illustrate the sorted spike waveforms and rasters, respectively, from one electrode in Fig. 3b. Note the lack of inter-spike intervals (ISIs) less than or equal to 2.5 ms (Fig. 3f), the typical absolute refractory period of

mammalian neurons. The presence of such intervals would be indicative of incorrect clustering. A similar absence of ISI violation was also observed when we sorted electrodes with an order of magnitude more events (Supplementary Fig. 4a–d). The high electrode density also allowed us to triangulate the putative location of each observed neuron on the basis of small changes in waveform amplitude over space (Fig. 4a).

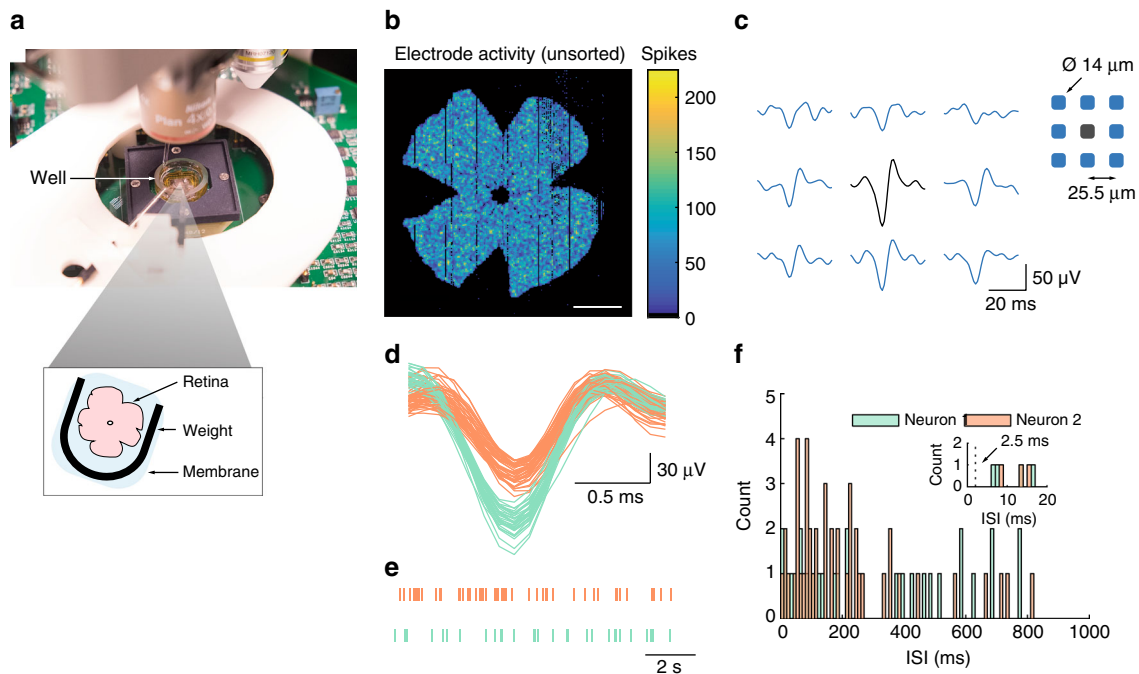
**Functional classification of neurons.** There have been intense debates over the number of functional types of RGCs, the retina's output neurons, and hence the number of information channels connecting the eyes to the brain. In the mouse, it has been variously estimated to be:  $>12$  by traditional sparse electrophysiology<sup>28</sup>,  $\sim 12$  types on morphological basis<sup>29</sup>,  $\geq 16$  by genetic markers<sup>30</sup>, and, most recently,  $\geq 30$  types with two-photon calcium imaging<sup>26</sup>. The crucial requirement for successful functional-type accounting is unbiased sampling, over large distances and at cellular granularity. Dense, high-channel-count recording systems are well suited for such applications. Importantly, electrophysiological recordings read out spikes—the transmission protocol between the retina and the brain—rather than somatic calcium fluxes, which are secondary to spike generation.

As a proof of concept, in Fig. 4 we flashed 1122- $\mu\text{m}$  diameter light spots of 1-s duration over the retina while simultaneously recording spikes. Using a high-throughput stream processing pipeline (Supplementary Fig. 5), we sorted and functionally classified the evoked spikes from 1750 neurons in response to the light stimuli. The RGCs were classified according to changes in spiking activities during a period spanning 3-s around the 1-s visual stimuli. Upon satisfying appropriate response characteristics (Methods section), the neurons were assigned one of several classical functional types<sup>31,32</sup>: ON transient, ON sustained, OFF transient, OFF sustained, ON–OFF, and SbC RGCs. Notably, large-scale recordings allowed us to routinely identify and record from the so-called suppression-by-contrast (SbC) RGCs, also known as uniformity detectors. These neurons were first described<sup>33</sup> in 1967, but seldom studied electrophysiologically<sup>34</sup>, presumably due to rare encounters in low-channel-count recordings. Example spike rasters, for five stimulus repetitions, of each functional class are illustrated in Fig. 4b–g.

The RGC types were homogeneous throughout the field of analyses (Fig. 4a), consistent with known RGC spatial distribution properties<sup>35</sup>. We further analysed the population distributions of the six RGC types (Fig. 4h). Approximately 14% of the neurons were the ON–OFF class, in agreement with anatomical accounting<sup>36</sup>. There was a slight excess of OFF-type neurons comparing to the ON-types, a consequence of some OFF neurons' smaller dendritic arbor and hence higher density<sup>37</sup>. More sophisticated light stimuli will permit further sub-division of RGC types. Nevertheless, these results illustrate the ability of the processing pipeline and algorithms to analyse the data accurately and automatically. These are important attributes for at-scale analytics.

**Simultaneous microstimulation and recording.** Electrical microstimulation has a long application history in neuroscience<sup>38,39</sup>. It provides a method for perturbing the neurons and/or network being studied. Furthermore, neural stimulation at scale may be useful in medical applications, as demonstrated by a 1500-electrode implantable photodiode array, which enabled blind patients to read and navigate<sup>40</sup>. With the exception of a recent design<sup>16</sup>, existing at-scale ( $>1000$  channels) electrophysiological tools have either lacked microstimulation features<sup>11</sup>





**Fig. 3** Large-scale recordings in the retina. **a** Photo of setup with mouse retina. **b** Spontaneous spiking activities recorded simultaneously from 34,187 electrodes over 12 s, at 10 kHz per electrode. Scale bar, 1275  $\mu\text{m}$ . **c** Unit activities were observed concurrently on several adjacent electrodes spaced 25.5  $\mu\text{m}$  apart. These spikes are single-trial waveforms. Sorted spike waveforms **d** and spike time raster **e** for one electrode in **b**. There are 33 and 47 spikes for the first and second neuron, respectively. **f** Inter-spike interval (ISI) plot for the two neurons in **d**. Inset: zoomed-in view of the first 20 ms

or offered limited simultaneously operable stimulation sites (up to approximately three dozen) despite high electrode number<sup>8,19</sup>.

Space saving achieved by removing the per-channel antialiasing filters allowed us to implement a stimulator within each recording site. The stimulators are individually programmable. Stimulus artifacts are reduced with two circuit features. First, routing-associated parasitic capacitance is minimized by integrating the stimulation circuit beneath each electrode. The charging and discharging of this capacitance during stimulation manifest as transient artifacts in the recordings. Second, the MOSFET pseudo-resistor in series with the electrode (Supplementary Fig. 2a) is disabled during and immediately after stimulation to quickly restore the first recording transistor's biasing voltage. Figure 5a illustrates, for ten trials, the electrically evoked spikes of a RGC following a single pulse. These spikes are easily distinguished from the artifacts. Further artifact suppression was achieved by subtracting recordings without neurons from those with neurons (Fig. 5b). The evoked spikes could be detected automatically in these post-processed data using the platform's stream processing pipeline (Fig. 5c). In this example, the neuron responded in 8 of 10 trials. To assess the reliability of electrical stimulation, we stimulated and calculated the response rate (over 10 trials) of 46 RGCs in two retinas using single 1.6 v pulses (Fig. 5d). More than half of these neurons responded to each trial, while the remaining neurons responded with  $\geq 50\%$  probability.

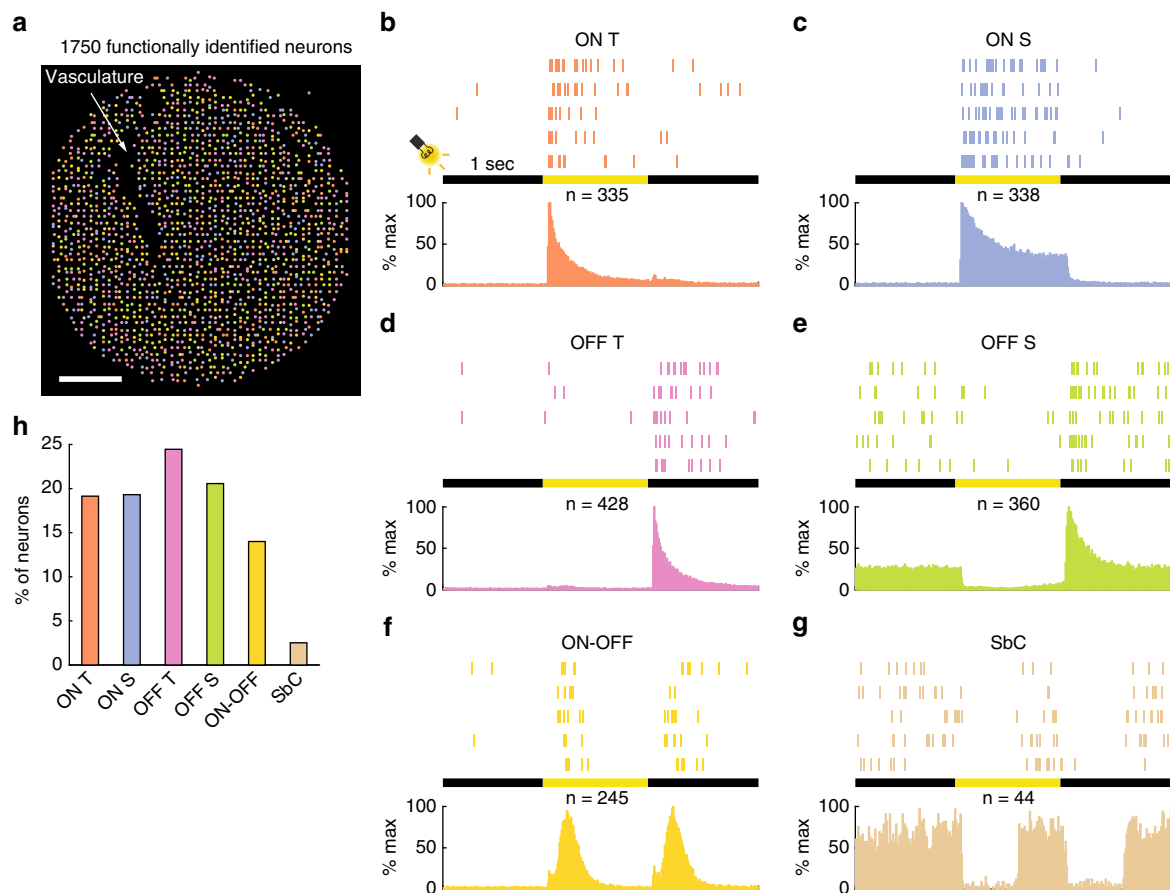
To ensure that these short-latency spikes were not spontaneous activities, we examined the quiescent firing rate of 22 neurons from a single retina (Supplementary Fig. 6a). The mean firing rate was 2.6 Hz. In contrast, when electrically stimulated, these same neurons spiked with a mean response rate of 74.1% within 3.0 ms of stimulus delivery, over 10 trials (Supplementary Fig. 6b). Therefore, the increased spiking probability following electrical stimulation was unlikely to be of spontaneous origin.

**Loss of focal activation by high-strength microstimulation.** An important goal of microstimulation, and indeed for neuronal

manipulation in general, is achieving spatiotemporally precise activation. Some studies have found confined activation with single-neuron precision<sup>41,42</sup>, while others observed wide-spread neuronal activation<sup>39,43</sup>. This discrepancy could be a consequence of shortcomings in existing recording tools. First, the inability to record from every neuron, or nearly so, across sufficient area could lead one to incorrectly conclude that activation is spatially confined, because signals from the recruited neurons were not completely accounted for. Second, techniques with limited single-spike sensitivity, such as calcium reporters<sup>44,45</sup>, require large stimuli capable of eliciting multiple spikes to reach detectability threshold. Electric field size increases with stimulus amplitude, influencing a larger neuronal population, giving rise to the alternative, incorrect conclusion of wide-spread neuronal activation, caused by the use of excessive stimuli due to poor spike sensitivity.

A key advantage of cellular-resolution, at-scale electrophysiology is the ability to simultaneously observe activities over the entire retina with single-spike sensitivity. We re-examined the spatial confinement of microstimulation in the retina while stimulating at one location (Fig. 5e). With moderate stimulus strength, we observed a 70% (7 out of 10 trials) response rate from one neuron at the stimulation site (Fig. 5i, red circle). A number of distant neurons were also recruited (i.e., responded in  $\geq 50\%$  of trials). The response latencies increased with distance from the stimulation location (Fig. 5f, g). The axon of RGCs converges at the optic nerve head near the central retina, where they exit the eye. The spatial distribution and response latency of these activated distant neurons were consistent with retrograde axonal stimulation<sup>46,47</sup>, as the axon from these neurons passed in close proximity to the stimulation site.

The number of distant neurons recruited by electrical stimulation was strongly dependent on the stimulus strength. Weak stimuli recruited one neuron at the stimulating electrode (i.e., the targeted neuron; Fig. 5h). The number of distant neurons elicited, as well as their response probability, increased as the



**Fig. 4** Functional classification of neurons using automated processing pipeline. The stimuli were 1-s, 1122- $\mu\text{m}$  diameter, light spots. **a** Data from a sensing area of  $1377 \times 1275 \mu\text{m}$  were automatically analysed and, upon satisfying appropriate response characteristics, classified into one of six functional classes. Scale bar, 225  $\mu\text{m}$ . **b–g** Raster plot for a representative ON transient RGC **b**, ON sustained RGC **c**, OFF transient RGC **d**, OFF sustained RGC **e**, ON-OFF RGC **f**, and suppression-by-contrast (SbC) RGC **g**, to five repetitions of one-second light flash. The peri-stimulus time histogram (10 ms bins) beneath each raster shows the normalized firing rate for all detected neurons in each class. The n-numbers denote the neurons considered in each histogram. All neurons were recorded simultaneously. **h** Distribution of functional classes for the identified neurons in **a**

stimulus strength increased (Fig. 5h–j). Therefore, focal activation depended critically on the stimulus strength. It should be sufficiently powerful to recruit the close proximity, target neuron, but should not be excessive, to avoid stimulating distant neurons with axons passing near the stimulating electrode.

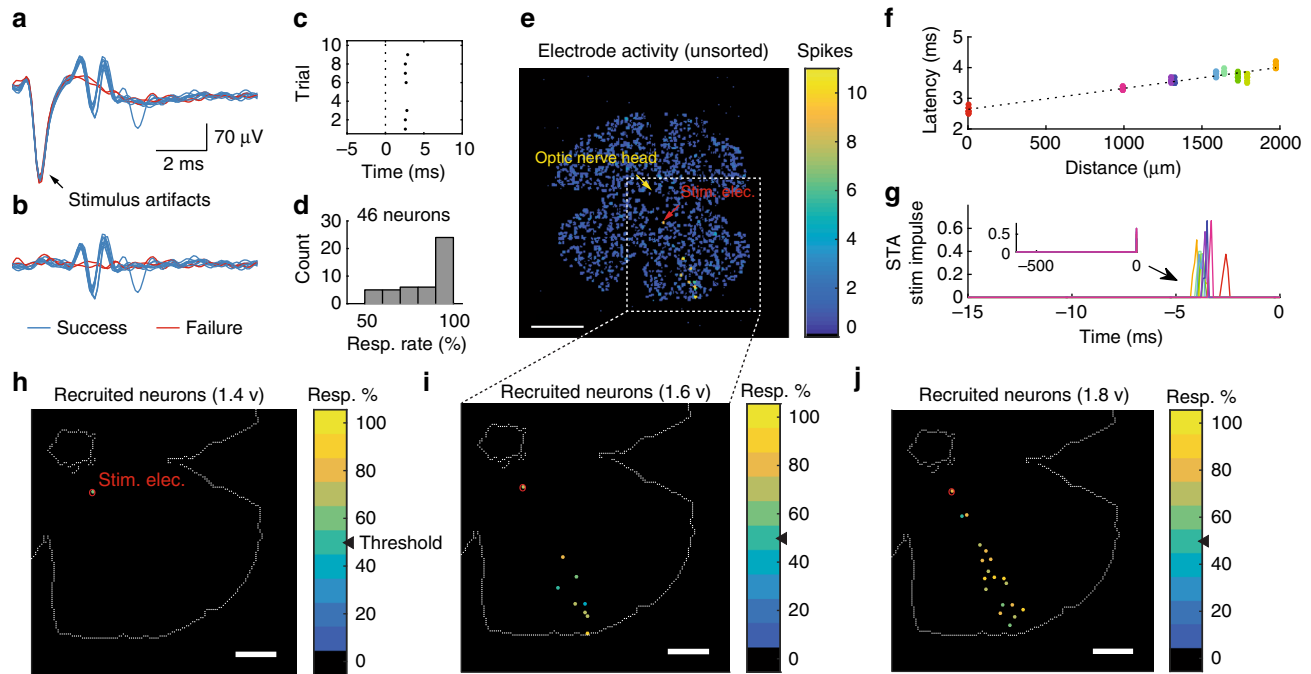
## Discussion

Traditional multichannel electrophysiology is limited in its ability to simultaneously realize low noise, dense and large-scale recordings, due to the need for per-channel antialiasing filters. We presented an acquisition paradigm that does not require these scalability-limiting elements. A platform based on this paradigm allowed us to record spiking activities in the mouse retina across more than 34,000 electrodes with high SNR. In conjunction with the platform's high-performance computing infrastructure, we were able to sort and functionally classify more than 1700 neurons following light stimulation. Finally, recording at cellular-resolution, across large area and with single-spike sensitivity, allowed us to examine the dynamics of microstimulation in greater spatiotemporal resolution than previously possible.

Our acquisition paradigm is inspired by compressive sensing (CS), based on the central CS notion that, if we know something about the frequency contents of the signal being acquired, it may be possible to recover the signal without sampling at the classical Nyquist rate. In electrophysiology, we know how the aliased thermal noise is manifested in the under-sampled, per-channel

data. Conventional CS approach would attempt recovery by some form of iterative, optimization algorithm, which is generally quite slow. In electrophysiology, we could instead exploit the statistical prior of thermal noise for computationally efficient recovery. Thus, while the implementations differ, the general concept is identical. There are further conceptual similarities. CS often uses irregular/random under-sampling to achieve incoherent aliasing. That is, the spectral power of the aliased content is evenly spread out across frequencies constituting the under-sampled data. This is notionally similar to our approach, where the aliased thermal noise is folded down into the first Nyquist zone (Supplementary Fig. 3b), averaging out variations in the original thermal noise power spectra and spectral angles.

Several multichannel electrophysiological systems have recently reached simultaneously recording channel counts in the thousands<sup>11,16,23</sup> or even up to 16,000 channels<sup>48</sup>, at the expense of noise performance. The noise in these tools are several times higher than traditional systems with at most a few hundred electrodes ( $\leq 10 \mu\text{V}$  rms versus 25–250  $\mu\text{V}$  rms or more). High SNR is critical for spike sorting, where neurons are distinguished on the basis of minute differences in spike waveform. This is particularly relevant for the mammalian nervous system. For example, extracellular signals in the mouse retina generally do not exceed much more than 150  $\mu\text{V}$  peak-to-peak, while reliable spike sorting requires at least 100  $\mu\text{V}$  peak-to-peak signals under optimal SNR conditions<sup>49</sup>. Indeed, current high-channel-count



**Fig. 5** Spatiotemporal effects of microstimulation revealed by large-scale electrophysiology. A grid of  $2 \times 2$  electrodes was used throughout. Besides **h** and **j**, 1.6 v pulses were used. **a** Simultaneous stimulation and recording at one electrode. Superimposed traces from 10 trials. **b** Same data as **a**, after removing the stimulus artifacts. **c** Raster plot for the evoked spikes in **b**. **d** Response rate (10 trials) of 46 RGCs in two retinas. **e** Events detected within 3 ms of stimulus onset. The stimulation site is marked by the red arrow. Aggregated data from 10 repetitions. Scale bar, 1275  $\mu\text{m}$ . **f** Latency of the evoked spikes increased with distance from the stimulation site. Each dot represents a spike (69 in total). The colors denote different neurons. Dotted line is linear fit ( $R^2 = 0.9541$ ,  $p < 0.0001$ , F-test). **g** Time course of spike-triggered average (STA) stimulus for each (color-matched) neuron in **f**. **h–j** Spike-sorted activity maps of identified neurons. The number of distant neurons activated by electrical stimulation increased with stimulus strength. The colors indicate response rate over 10 trials. Dotted lines are approximate outline of the retina. Scale bar, 765  $\mu\text{m}$

implementations have used higher-than-typical spike detection thresholds<sup>24</sup> (7.5 SD vs.  $\sim 4.0$  SD) to avoid misinterpreting noise as spikes; have detected few neurons despite the large number of electrodes<sup>48</sup> (126 neurons in 16,384 electrodes); or have identified putative events at locations that apparently did not correspond with neurite positions<sup>23</sup>. In general, because of noise, a substantial fraction of neurons may be unobservable when using these systems, potentially diminishing the benefits of high channel count and/or high electrode density. The systems designed by the Litke and Chichilnisky groups<sup>27,50</sup>, and the Roska and Hierlemann groups<sup>8</sup> have achieved input referred noise as low as 5  $\mu\text{V}$  rms and 2.4  $\mu\text{V}$  rms, respectively. The higher SNR offers several advantages, including improved spike sorting performance and the ability to detect dendritic spikes. However, the superior noise performance limits the simultaneously recording channels to 1024 or less, due to the need for per-channel anti-aliasing filters in these classical multiplexed systems.

Here we demonstrated the multichannel acquisition paradigm in a 65,536-channel ex vivo recording grid, fabricated using commercial CMOS-integrated circuit (IC) technology. The strategy is generalizable to any dense, high-channel count electrophysiological systems, including implantable, long-term in vivo recording tools. In these, multiplexing is important not only for density and channel count scaling, but also to reduce wiring, power consumption and heat dissipation. Indeed, the sampling paradigm will work for any big-data acquisition applications where the spectral and statistical characteristics of the high frequency components, above twice the per-channel observation rate and below the recording channel bandwidth, can be reasonably approximated.

Another advantage of this data acquisition approach is that the signal processing steps (channel separation and aliased noise

removal) are all implemented in the digital domain. The throughput of these procedures will improve with technological advancements in electronics, allowing the approach to continue scaling beyond the tens of thousands of simultaneously recording and stimulating channels presented here.

## Methods

**Recording and stimulation architecture.** The architecture for our platform is summarized in Fig. 1a. Supplementary Fig. 2a shows the schematic overview for the recording and stimulating circuits. The platform is constructed from a combination of custom IC, circuit-board-level components, synthesized digital logic in field programmable gate arrays (FPGAs) and algorithms running on  $\times 86$  CPUs and NVIDIA CUDA processors. The IC (Supplementary Fig. 2b) contained 65,536 front-end elements, divided into 16 blocks of 4096 elements each. Each block is connected to a back-end circuit for additional amplification through a 4096:1 multiplexer. We bandpass filter the outputs from these back-ends to confine spectral content between 50 Hz and 40 MHz with a Sallen-Key filter, implemented on the printed circuit board, then digitize the resulting signals using 12-bit analog-to-digital converters (ADCs). The ADCs' data streams are captured by a FPGA and transferred to a computer. There are four FPGAs in the system, each handling the outputs of four ADCs. Each front-end element contains programmable registers to enable or disable voltage-based, electrical microstimulation, via the capacitive-coupled HfO<sub>2</sub> dielectric interface.

The system is powered by a 6 V supply, and uses approximately 24.7 W when in operation. The power consumption is dominated by the four Xilinx FPGAs and, to a lesser extent, the board-level bandpass filters. The IC consumes less than 0.6 % (i.e.,  $\leq 148$  mW) of the total power budget.

**Electrical stimulation.** The electrical stimulation strategy is based on our previous architecture<sup>51,52</sup>. Thirty-two elements are configured at a time during the programming phase, which takes 60 ns. Only elements needing change of stimulation status require programming. Asserting a global digital signal triggers stimulus delivery on all electrodes programmed to do so. User programmable voltage stimuli ranging from 0 to 2.0 volt, are generated by a voltage source on the circuit board (Fig. 1f), fed into the IC, shared by all stimulating electrodes, and capacitive-coupled to the neurons via the electrodes' HfO<sub>2</sub> dielectric interface. The MOSFET pseudo-resistor is turned off during stimulation to minimize source resistance. For



all electrical stimulation results presented here, we delivered the stimuli with four neighboring electrodes arranged in a  $2 \times 2$  pattern, we found the larger electric field so generated to more consistently elicit spikes in the mouse retina than that from one electrode. The stimulus threshold is defined as the voltage required to elicit electrically evoked responses within 3 ms of stimulus onset in 50% of trials (out of 10), at the  $2 \times 2$  stimulating electrodes.

**Integrated circuit fabrication and post-processing.** The top metal layer of the CMOS IC serves as the base material for the sensing electrodes. This is achieved by etching away the foundry-deposited passivation layers (polyimide, silicon nitride and silicon dioxide) by inductively coupled plasma/reactive ion etching (ICP/RIE) using a mixture of  $\text{SF}_6$  and  $\text{O}_2$  plasma. We restrict etching to within the sensing region by protecting all other area with a  $\sim 16 \mu\text{m}$  layer of AZ-4620 photoresist, patterned with standard UV photolithography.

The naturally occurring aluminum oxide on the top metal is stripped by ion milling. Next, we deposit 6 nm of  $\text{HfO}_2$ , a high-K dielectric, by atomic layer deposition (ALD) at  $150^\circ\text{C}$  on top of the metal. This serves two purposes. First, it creates a capacitive sensing and stimulation interface; and second, it provides a passivation layer for the underlying aluminum. This  $\text{HfO}_2$  layer provides a capacitance of 5.8 pF over the  $14 \times 14 \mu\text{m}$  electrode. The capacitance is ascertained by building test structures (Supplementary Fig. 2c) consisting of a metal- $\text{HfO}_2$ -metal stack on a  $\text{SiO}_2$  substrate, followed by measurements with a semiconductor parameter analyser (Agilent B1500).

A 160-nm-thick film of conductive polymer<sup>10</sup> is spun over the die surface to reduce the electrode-to-electrolyte interfacial impedance<sup>53</sup> (between the conformal ALD  $\text{HfO}_2$  layer and saline). This is followed by a 220-nm PMMA A4 (MicroChem) barrier film. The PEDOT:PSS+ PMMA stack is patterned using UV lithography with Shipley S1813 photoresist, followed by  $\text{O}_2$  ICP etching, such that only the electrodes are covered by PEDOT:PSS. Finally, the remaining PMMA is stripped with PG Remover (MicroChem).

Each post-processed die is attached to a custom ball grid array (BGA) with thermally conductive epoxy, wire-bonded, and then encapsulated with medical grade epoxy (OG-116-31, Epoxy Technology, Inc.). In the final step, we attach a polycarbonate ring around the IC using Sylgard 184 (Dow Corning) to serve as the perfusion chamber.

**Sparse sampling and data recovery.** Multiplexing causes each channel to be observed at a rate ( $f_{\text{visit}}$ ), through the multiplexer, considerably lower than the channel bandwidth ( $f_{\text{BW}}$ ). Unless the content spanning  $f_{\text{visit}}/2 \dots f_{\text{BW}}$  is removed from the per-channel data, aliasing occurs. The problem of per-channel data recovery is thus two-fold. First, the channel data must be extracted from the ADC data stream (Fig. 1b, step 2); and second, the spectral contribution of contents in  $f_{\text{visit}}/2 \dots f_{\text{BW}}$  has to be computed (Fig. 1b, step 3) and removed from the channel data (Fig. 1b, step 4).

The first task, per-channel data extraction, is achieved by keeping a history of the scanned channels during recording. In this manner, each sampled value from the ADC can be assigned to the channel from which it originates by examining the history at the corresponding time point.

The goals of the second task are to preserve the spectral contents of neural signal and to prevent aliasing of contents in  $f_{\text{visit}}/2 \dots f_{\text{BW}}$ , for data sampled at only  $f_{\text{visit}}$  (Supplementary Fig. 3b), with  $f_{\text{visit}} \ll f_{\text{BW}}$ . We begin by setting the multiplexers' per-channel visit rate ( $f_{\text{visit}}$ ) to be sufficiently high, such that the spike bandwidth (300–3k Hz) is entirely encompassed by  $f_{\text{visit}}/2$  and that the range  $f_{\text{visit}}/2 \dots f_{\text{BW}}$  is dominated by thermal noise. We typically set  $f_{\text{visit}}$  to 10 kHz to achieve these requirements.

Several statistical and spectral characteristics of thermal noise make its aliased image amenable to reconstruction in the frequency domain. This thermal noise, which comes from the electrodes and from the amplifier transistors, is a stationary random process, with a flat spectrum and a Gaussian time-domain amplitude distribution<sup>54</sup> of zero mean and variance  $\sigma^2$ . The probability density function for such a process is

$$N(x|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

We can easily determine every channel's  $\sigma^2$  for thermal noise calculation by recording each channel without multiplexer interruption (i.e. conventional sampling) at full system bandwidth, thereby completely specifying the thermal noise characteristics of the channel up to  $f_{\text{BW}}$ . With the noise variance  $\sigma^2$  and bandwidth  $f_{\text{BW}}$  now known for every recording channel, we computationally construct the thermal noise  $n_i$  for each channel  $i$ .

As the under-sampled thermal noise is folded down into the First Nyquist zone, in the per-channel data, fluctuations in power and spectral angle, from frequency to frequency, are averaged out (Supplementary Fig. 3b). We can compute the power contributed by aliasing using the expected average thermal noise power, multiplied by the number of folded Nyquist zones. Similarly, the spectral angles converge to zero in the aliased version of the thermal noise. We construct vectors in the frequency space to represent the aliased thermal noise, subtracting these from the per-channel data, thereby reversing the effects of aliasing.

The effects of thermal noise aliasing, between  $f_{\text{visit}}/2 \dots f_{\text{BW}}$ , in the per-channel data (Supplementary Fig. 3b) can be readily reproduced by decimating the

generated noise  $n_i$  to a lower rate,  $f_{\text{visit}}$ . We denote this aliased sequence  $a_i$ :

$$a_i : \text{decimate}(n_i, f_{\text{visit}})$$

Next we construct another sequence  $b_i$ , a decimated version of  $n_i$  without aliasing. This is accomplished by first low-pass filtering  $n_i$  at  $f_{\text{visit}}/2$ , followed by decimation to the new rate  $f_{\text{visit}}$ :

$$b_i : \text{decimate}(\text{lowpass}(n_i, f_{\text{visit}}/2), f_{\text{visit}})$$

The power contributed by the aliased thermal noise at each frequency, for a system with bandwidth  $f_{\text{BW}}$  but sampled at only  $f_{\text{visit}}$ , is, therefore, the power difference between the deliberately aliased sequence  $a_i$  and the anti-aliased sequence  $b_i$ :

$$P_i = |\mathcal{F}(a_i)| - |\mathcal{F}(b_i)|$$

where  $\mathcal{F}$  denotes Fourier transform. By removing the contribution of  $P_i$  in the per-channel data, we avoid aliasing. Because thermal noise is a stochastic process, for any finite-length segment there will be slight fluctuations in power from frequency to frequency, and no two finite-length segments are exactly identical. These uncertainties are minimized with increased length for  $n_i$ , and by computing  $P_i$  from the averaged power, which converges to the true value as the number of analysed frequencies increases:

$$P_i = \text{mean}(|\mathcal{F}(a_i)|) - \text{mean}(|\mathcal{F}(b_i)|)$$

We then construct a set of vectors describing the aliased contents in the frequency domain:

$$\vec{V}_i = P_i e^{j\text{arg}(\mathcal{F}(a_i))}$$

In the last step, we remove these aliased contents  $\vec{V}_i$  from the per-channel data  $d_i$ . In doing so, we recover the data  $e_i$  without aliasing (Supplementary Fig. 3c):

$$e_i = \mathcal{F}^{-1}(\mathcal{F}(d_i) - \vec{V}_i)$$

In practice, we perform the thermal noise parameter estimation procedure separately in physiological saline prior to the biological experiments with 50–100 ms recordings at full sampling rate. This can be computed for 16 pixels in parallel, taking advantage of the 16 parallel read-outs on the IC. The noise parameters are saved for each pixel and reused in subsequent biological experiments.

**Data analysis pipeline.** The four FPGAs, each collecting digitized data from four ADCs, are connected to a high-performance computer with separate USB3 links (Supplementary Fig. 5), with a combined transfer capacity of approximately 1 GB/s. Low-level drivers and custom libraries store the data to RAID0 hard drives and arbitrate interactions with near real-time processing algorithms written in C++, running on Intel x86 CPUs (Xeon E5-2623 3 GHz) and NVIDIA GPUs (Quadro K5200).

**Characterization of acquisition paradigm.** The perfusion well is filled with PBS at physiological concentration. We generate 1-kHz sine waves from a function generator (AFG3102C, Tektronix) and attenuate the signal amplitude down to 100–200  $\mu\text{V}$  peak-to-peak. The test signals are applied in the PBS bath through a pair of large (hence low impedance) Ag-AgCl electrodes. The recording SNR is measured by applying a 1-kHz sine wave into the saline bath, directly above the test electrode. After bandpass filtering the data between 300 and 3 kHz, we then compare its variance against a similarly bandpass filtered quiescent recording with a grounded bath. We determine the corner frequency of the high-pass filter at each electrode by measuring the amplitude attenuation of a bath-applied sine wave, at different frequencies.

We also compare the performance of our system to that of conventional, low-noise patch clamp recordings. A  $\sim 950\text{-k}\Omega$ , PBS-filled borosilicate glass pipette, connected to a commercial patch clamp recording setup (MultiClamp 700B, Digidata 1550, pClamp 10, all Molecular Devices), is brought within 50  $\mu\text{m}$  of the test electrode in the 65,536-electrode grid using a micromanipulator (MP-285, Sutter Instruments). The digitally recovered data from the test electrode within the CMOS recording grid is compared against the patch clamp amplifier's recordings. Both signals are bandpass filtered between 300–3k Hz.

**Mouse retina preparation.** WT mice  $>P40$ , of either sex, are dark adapted for one hour and deeply anaesthetized with isoflurane in  $\text{O}_2$ . Following euthanasia, the eyes are rapidly enucleated under dim red light then placed in oxygenated Ames' medium (Sigma Aldrich) with 1.9 g/L of  $\text{NaHCO}_3$  and equilibrated with 95%  $\text{O}_2$ /5%  $\text{CO}_2$ , at room temperature. Under a near-infrared illuminated dissection microscope we hemisect the eyes, remove the anterior chamber, the vitreous and the posterior eye cup, then place the isolated retina in equilibrated Ames' medium at room temperature, in darkness.

For recordings, an intact retina is flattened by several small incisions around the periphery, transferred onto a transparent dialysis membrane, then placed retinal ganglion cell side down, on top of the 65,536-channel CMOS recording / stimulation grid. A small, custom-made platinum harp, with (Supplementary Fig. 7) or without nylon threads, is placed over the membrane to maintain retina-

to-electrode contact. The retina is kept alive by perfusing with equilibrated Ames' medium, heated to 33–35 °C, at a rate of ~4.5 mL/min. We allow at least 30-minutes recovery in the warm solution before recordings. All experiments are performed in the dark. Visualization of the retina under a fixed-stage upright microscope (Nikon FN1) is achieved with near-infrared illumination (≥850 nm) and an IR-sensitive CCD camera.

Visual stimuli are generated on an OLED display with built-in digital signal processor and memory, then projected onto the retina via custom optics attached to the modified epi-fluorescent light path of a fixed-stage microscope. Visual patterns are triggered via a TTL digital input. The light spot stimuli are 1-s in duration, of 1122 μm diameter, white on black background at 100 % contrast, with intensity of  $11.4 \times 10^{11}$  photons/s/cm<sup>2</sup>.

**Data analysis.** Eight mouse retinas were used for the biological results presented here. After per-channel data separation and noise removal (Fig. 1b), the data are bandpass filtered between 100 Hz and 3 kHz. Putative spikes are detected by threshold crossing over 4.5 standard deviation of mean. We use an event window of 0.7 ms and 1.0 ms, before and after, the spike peak, respectively. To facilitate spike sorting, whenever an event is detected, we also collect the waveforms from the electrode's eight adjacent neighbors over the corresponding timeframe. These waveforms are concatenated and saved to a database, mapping from electrode address to event data (peak time, waveform data and recording sweep ID).

We compute the principal components for each waveform by singular value decomposition, then sort the waveforms using the first four scores by expectation maximization (EM) with Gaussian mixture model. We repeat the EM procedure ten times to avoid suboptimal, local-maxima solutions. The repetitions are performed concurrently using parallel CPUs/GPUs to reduce run time. The clustering with the highest fitness metric, namely, maximal between-cluster separation, minimal within-cluster spread, lack of singleton clusters and zero interspike interval violation, are deemed the correct/best solution. Conventional EM algorithms require a priori the number of clusters, an impractical requirement for at-scale spike sorting. We implement automatic cluster number detection using the foregoing fitness metric. Specifically, the number is increased incrementally from a minimum of two, up to maximum of 10. The lowest cluster number without interspike interval violation and having the highest, or equal highest, fitness metric is used. These procedures are implemented on parallel hardware to reduce run time. All sorted spikes are saved to a database, mapping from electrode address to a list of waveforms, their associated spike time, sweep ID and cluster assignment.

To classify RGCs into functional types<sup>31,32</sup>, we flash a 1-s light spot over the region of interest in the retina. The recording duration is three seconds, for 1-s pre-stimulus and 1-s post-stimulus periods. We divide the 3-s recording interval into six equal segments of 500 ms each, numbered 1–6. ON-type neurons spike predominantly in segments 3 and 4, while OFF-type neurons spike predominantly in segments 5 and 6. Neurons with high segment 3 or segment 5 rates relative to segments 4 and 6 are designated as "Transient", otherwise they are designated as "Sustained." ON-OFF RGCs are distinguished by simultaneously strong responses in segments 3 and 5, and low baseline spiking rates. SbC RGCs are distinguished by their high baseline rate, with little or no spikes in segments 3 and 5. Units with ambiguous spiking profile, without the foregoing characteristics, are not assigned a functional class. Units with low spike rate ( $\leq 2$  Hz) during the 3-s recording period are also not functionally classified, because the low spike counts preclude accurate classification. The pseudocode for the functional classification procedure can be found in Supplementary Note 1.

In Fig. 5, the spike-triggered average (STA) stimulus is defined as the average stimulus preceding a spike from a neuron. Specifically, it is the sum of stimuli (voltage impulses through the HfO<sub>2</sub> dielectric) that preceded each spike, divided by the number of spikes.

To create the spike latency versus distance plot (Fig. 5f), we begin by collapsing, for each electrode, the spike-sorted events over 10 stimulus repetitions into a time-invariant plot. From this plot, at each electrode, we select all spike-sorted unit(s), with at least 5 events within 5 ms following stimulus delivery (i.e., 5 successes out of 10 trials). The spike time of these units are then plotted as a function of distance from the stimulation site.

**Data availability.** The data sets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 3 August 2017 Accepted: 1 November 2017

Published online: 27 November 2017

## References

1. Alivisatos, A. P. et al. The brain activity map project and the challenge of functional connectomics. *Neuron* **74**, 970–974 (2012).
2. Yuste, R. & Katz, L. C. Control of postsynaptic calcium influx in developing neocortex by excitatory and inhibitory neurotransmitters. *Neuron* **6**, 333–344 (1991).

3. Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M. & Keller, P. J. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nat. Methods* **10**, 413–420 (2013).
4. Prevedel, R. et al. Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nat. Methods* **11**, 727–730 (2014).
5. Grienberger, C. & Konnerth, A. Imaging calcium in neurons. *Neuron* **73**, 862–885 (2012).
6. Yuste, R. & Konnerth, A. *Imaging in neuroscience and development: a laboratory manual*. (Cold Spring Harbor Laboratory Press, 2011).
7. Buzsáki, G. et al. Tools for probing local circuits: High-density silicon probes combined with optogenetics. *Neuron* **86**, 92–105 (2015).
8. Ballini, M. et al. A 1024-Channel CMOS microelectrode array with 26,400 electrodes for recording and stimulation of electrogenic cells in vitro. *IEEE J. Solid State Circuits* **49**, 2705–2719 (2015).
9. Lopez, C. M. et al. An implantable 455-active-electrode 52-channel CMOS neural probe. *IEEE J. Solid State Circuits* **49**, 248–261 (2013).
10. Khodagholy, D. et al. NeuroGrid: recording action potentials from the surface of the brain. *Nat. Neurosci.* **18**, 310–315 (2015).
11. Berdoncini, L. et al. Active pixel sensor array for high spatio-temporal resolution electrophysiological recordings from single cell to large scale neuronal networks. *Lab. Chip* **9**, 2644–2651 (2009).
12. Du, J., Blanche, T. J., Harrison, R. R., Lester, H. A. & Masmanidis, S. C. Multiplexed, high density electrophysiology with nanofabricated neural probes. *PLoS ONE* **6**, e26204 (2011).
13. Buzsáki, G., Anastassiou, C. A. & Koch, C. The origin of extracellular fields and currents - EEG, ECoG, LFP and spikes. *Nat. Neurosci.* **13**, 407–420 (2012).
14. Ficsella, M. et al. Recording from defined populations of retinal ganglion cells using a high-density CMOS-integrated microelectrode array with real-time switchable electrode selection. *J. Neurosci. Methods* **211**, 103–113 (2012).
15. Eversmann, B. et al. A 128x128 CMOS biosensor array for extracellular recording of neural activity. *IEEE J. Solid State Circuits* **38**, 2306–2317 (2003).
16. Bertotti, G. et al. in *Biomedical Circuits and Systems Conference (BioCAS)* (IEEE, Lausanne, 2014).
17. Muller, R. et al. A minimally invasive 64-channel wireless uECOG implant. *IEEE J. Solid State Circuits* **50**, 344–359 (2015).
18. Johnson, L. J. et al. A novel high electrode count spike recording array using an 81,920 pixel transimpedance amplifier-based imaging chip. *J. Neurosci. Methods* **205**, 223–232 (2012).
19. Yuan, X. et al. in *Proceedings of the IEEE Symposium on VLSI Circuits* 1–2 (Honolulu, HI, USA, 2016).
20. Candes, E., Romberg, J. & Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **59**, 1207–1223 (2006).
21. Aldroubi, A. & Gröchenig, K. Nonuniform sampling and reconstruction in shift-invariant spaces. *SIAM Rev.* **43**, 585–620 (2001).
22. Viventi, J. et al. Flexible, foldable, actively multiplexed, high-density electrode array for mapping brain activity in vivo. *Nat. Neurosci.* **14**, 1599–1605 (2011).
23. Lambacher, A. et al. Identifying firing mammalian neurons in networks with high-resolution multi-transistor array (MTA). *Appl. Phys. B* **102**, 1–11 (2011).
24. Maccione, A. et al. Following the ontogeny of retinal waves: pan-retinal recordings of population dynamics in the neonatal mouse. *J. Physiol.* **592**, 1545–1563 (2014).
25. Briggman, K. L., Helmstaedter, M. & Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471**, 183–188 (2011).
26. Baden, T. et al. The functional diversity of retinal ganglion cells in the mouse. *Nature* **529**, 345–350 (2016).
27. Litke, A. M. et al. Large-scale imaging of retinal output activity. *Nucl. Instrum. Methods Phys. Res. A* **501**, 298–307 (2003).
28. Farrow, K. & Masland, R. H. Physiological clustering of visual channels in the mouse retina. *J. Neurophysiol.* **105**, 1516–1530 (2011).
29. Kong, J.-h., Fish, D. R., Rockhill, R. L. & Masland, R. H. Diversity of ganglion cells in the mouse retina: Unsupervised morphological classification and its limits. *J. Comp. Neurol.* **489**, 293–310 (2005).
30. Sümbül, U. et al. A genetic and computational approach to structurally classify neuronal types. *Nat. Commun.* **5**, 3512 (2014).
31. Wässle, H. Parallel processing in the mammalian retina. *Nat. Rev. Neurosci.* **5**, 747–757 (2004).
32. Zeck, G. M. & Masland, R. H. Spike train signatures of retinal ganglion cell types. *Eur. J. Neurosci.* **26**, 367–380 (2007).
33. Levick, W. Receptive fields and trigger features of ganglion cells in the visual streak of the rabbits retina. *J. Physiol.* **188**, 285–307 (1967).
34. Sivyver, B., Taylor, W. R. & Vaney, D. I. Uniformity detector retinal ganglion cells fire complex spikes and receive only light-evoked inhibition. *Proc. Natl Acad. Sci. USA* **107**, 5628–5633 (2010).
35. Rodieck, R. W. *The First Steps in Seeing*. (Sinauer Associates, 1998).
36. Vaney, D. I., Sivyver, B. & Taylor, W. R. Direction selectivity in the retina: symmetry and asymmetry in structure and function. *Nat. Rev. Neurosci.* **13**, 194–208 (2012).

37. Ratliff, C. P., Borghuis, B. G., Kao, Y.-H., Sterling, P. & Balasubramanian, V. Retina is structured to process an excess of darkness in natural scenes. *Proc. Natl Acad. Sci. USA* **107**, 17368–17373 (2010).
38. Cohen, M. R. & Newsome, W. T. What electrical microstimulation has revealed about the neural basis of cognition. *Curr. Opin. Neurobiol.* **14**, 169–177 (2004).
39. Logothetis, N. K. et al. The effects of electrical microstimulation on cortical signal propagation. *Nat. Neurosci.* **13**, 1283–1291 (2010).
40. Zrenner, E. et al. Subretinal electronic chips allow blind patients to read letters and combine them to words. *Proc. R. Soc. B* **278**, 489–1497 (2011).
41. Sekirnjak, C. et al. Electrical stimulation of mammalian retinal ganglion cells with multielectrode arrays. *J. Neurophysiol.* **95**, 3311–3327 (2006).
42. Sekirnjak, C. et al. High-resolution electrical stimulation of primate retina for epiretinal implant design. *J. Neurosci.* **28**, 4446–4456 (2008).
43. Histed, M. H., Bonin, V. & Reid, R. C. Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation. *Neuron* **63**, 508–522 (2009).
44. Tian, L., Hires, S. A. & Looger, L. L. Imaging neuronal activity with genetically encoded calcium indicators. *Cold Spring Harb. Protoc.* **1012**, 647–656 (2012).
45. Harris, K. D., Quiroga, R. Q., Freeman, J. & Smith, S. L. Improving data quality in neuronal population recordings. *Nat. Neurosci.* **19**, 1165–1174 (2016).
46. Tehovnik, E. J., Tolia, A. S., Sultan, F., Slocum, W. M. & Logothetis, N. K. Direct and indirect activation of cortical neurons by electrical microstimulation. *J. Neurophysiol.* **96**, 512–521 (2006).
47. Fried, S. I., Lasker, A. C. W., Desai, N. J., Eddington, D. K. & Rizzo, J. F. Axonal sodium channel bands shape the response to electric stimulation in retinal ganglion cells. *J. Neurophysiol.* **101**, 1972–1987 (2009).
48. Menzler, J., Channappa, L. & Zeck, G. Rhythmic ganglion cell activity in bleached and blind adult mouse retinas. *PLoS ONE* **9**, e106047 (2014).
49. Tikidji-Hamburyan, A. et al. Retinal output changes qualitatively with every change in ambient illuminance. *Nat. Neurosci.* **18**, 66–74 (2015).
50. Li, P. H. et al. Anatomical identification of extracellularly recorded cells in large-scale multielectrode recordings. *J. Neurosci.* **35**, 4663–4675 (2015).
51. Lei, N. et al. High-resolution extracellular stimulation of dispersed hippocampal culture with high-density CMOS multielectrode array based on non-Faradaic electrodes. *J. Neural Eng.* **8**, 044003 (2011).
52. Lei, N., Shepard, K. L., Watson, B. O., MacLean, J. N. & Yuste, R. in *Digest of Technical Papers, International Solid State Circuits Conference* (San Francisco, CA, USA, 2008).
53. Samba, R., Herrmann, T. & Zeck, G. PEDOT–CNT coated electrodes stimulate retinal neurons at low voltage amplitudes and low charge densities. *J. Neural Eng.* **12**, 016014 (2015).
54. Barry, J. R., Lee, E. A. & Messerschmitt, D. G. *Digital Communication*. (Kluwer Academic Publisher, 2003).

## Acknowledgements

This work was supported in part by the U. S. Army Research Laboratory and the U. S. Army Research Office under Contract W911NF-12-1-0594, by the Defense Advanced Research Project Agency (DARPA) under Contract N66001-17-C-4002, by the National Institutes of Health under Grants U01NS0099697 and U01NS0099726, and by the National Health and Medical Research Council of Australia CJ Martin Fellowship (APP1054058).

## Author contributions

D.T. built the system with assistance from D.S. and A.B. D.T. carried out the biological experiments. D.T., D.S., A.B., R.Y. and K.S. contributed to the system design, experimental discussions, and wrote the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-017-02009-x>.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017