An amplitude code transmits information at a visual synapse

Ben James^{1,2}, Léa Darnet^{1,2}, José Moya-Díaz^{1,2}, Sofie-Helene Seibel^{1,2} and Leon Lagnado ¹

Most neurons transmit information digitally using spikes that trigger the release of synaptic vesicles with low probability. The first stages of vision and hearing are distinct in that they operate with analog signals, but it is unclear how these are recoded for synaptic transmission. By imaging the release of glutamate in live zebrafish, we demonstrate that ribbon synapses of retinal bipolar cells encode contrast through changes in both the frequency and amplitude of release events. Higher contrasts caused multiple vesicles to be released within an event, and such coding by amplitude often continued after the rate code had reached a maximum frequency. Glutamate packets equivalent to five vesicles transmitted four times as many bits of information per vesicle compared with those released individually. By discretizing analog signals into sequences of numbers up to about 11, ribbon synapses can increase the dynamic range, temporal precision and efficiency with which visual information is transmitted.

he spike code of neurons has been studied in detail ^{1,2} but much less is known about the vesicle code transmitting information across the synapse³. Since the early 1950s⁴, the general view has been that changes in presynaptic potential are communicated by modulating the mean rate of a Poisson process in which vesicles at the active zone are triggered to fuse independently^{5,6}. At most synapses, the voltage signal controlling this process is digital, arriving in the form of an all-or-none spike, and the output is also digital, in the form of a vesicle releasing a fixed packet of neurotransmitter. But in the first stages of vision and hearing, the neural signal is in analog form; that is, continuous changes in membrane potential that are graded with the strength of the stimulus⁷. How are these sensory signals recoded for transmission across a synapse?

Synapses driven by graded changes in membrane potential are distinguished from those driven by spikes by the presence of a specialized structure, the ribbon, that holds tens of vesicles just behind the active zone^{8,9}. It has been generally assumed that ribbon synapses represent the strength of the incoming stimulus in the form of a rate code generated by changes in the frequency of release events composed of one quantum of neurotransmitter contained within a vesicle^{5,10-12}. But it is also known that when these synapses are activated strongly, several vesicles can be released within a few milliseconds^{13–15}. This process, termed multivesicular release (MVR), is now recognized to be a feature of synaptic connections in several regions of the brain¹⁶. In ribbon synapses of mechanosensitive hair cells and retinal bipolar cells, the fusion of two or more vesicle equivalents can even be synchronized to within 100 µs¹⁷⁻²⁰. This type of process is termed coordinated multivesicular release (CMVR) to highlight the deviation from a Poisson process in which vesicles fuse independently of each other. Although the mechanisms underlying MVR are not understood, these observations suggest that the output of a ribbon synapse may not be a binary set of symbols consisting of zero or one vesicle but rather a number of symbols composed of different numbers of vesicle equivalents21. The role that MVR might play in coding sensory information has been unclear because it has not yet been observed in response to light or sound.

To understand the functional role of MVR in the retina, we used the fluorescent glutamate reporter iGluSnFR 22 to image the synaptic

output of bipolar cells in larval zebrafish. Optimizing the signal-tonoise ratio (SNR) allowed us to count vesicles released from individual active zones in vivo and to investigate how they are used to
represent a stimulus. We found that these ribbon synapses employ
a hybrid strategy for encoding visual contrast that involves changes
in both the frequency of release events and the amplitude of those
events. We term these two components the rate and amplitude
codes, respectively. Changes in the amplitude code confer several
advantages, including the ability to signal contrast beyond the range
at which the rate code has reached a saturating frequency, improving the temporal precision of transmission and increasing the efficiency of communication by transmitting more bits of information
per vesicle.

Results

Optical detection of MVR in vivo. All visual information flows through bipolar cells, which in turn drive the ganglion cells that deliver the results of retinal processing to other parts of the brain. To investigate how bipolar cells transmit this information across the synapse, we drove expression of iGluSnFR using the ribeye promoter²³ (Fig. 1a). A periodic and spatially uniform stimulus (5 Hz) triggered a train of glutamate transients that were sampled by performing line scans through the terminal at 1 kHz (Fig. 1b,c; Supplementary Videos 1 and 2). These signals had Gaussian intensity profiles along the line scan constrained to distances of $\sim 1.5 \,\mu\text{m}$, making it possible to distinguish fusion events at neighboring active zones by spatial demixing (Fig. 1c; Supplementary Figs. 1 and 2). Glutamate release at adjacent active zones did not always coincide in time, reflecting the stochastic nature of vesicle fusion (Fig. 1d). Most notably, the size of glutamate transients varied widely, both within and between active zones providing the output of one bipolar cell.

The time course of iGluSnFR transients were similar for events of different amplitude (Fig. 2a), a property that allowed us to improve event detection by using Wiener deconvolution⁶. The steps in the analysis procedure are described in detail in the Methods and are summarized in Fig. 2b, which shows an example of a raw iGluSnFR trace and its deconvolution using a temporal kernel with the shape

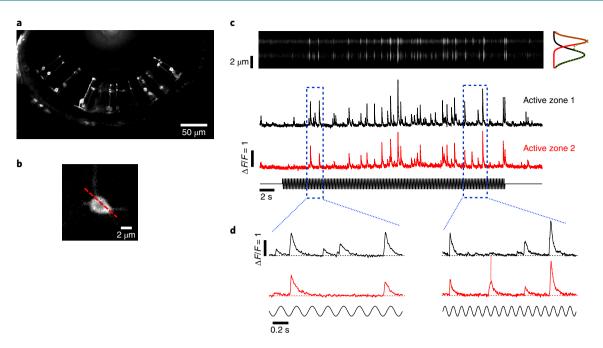


Fig. 1 | Glutamate transients of varying amplitude imaged at individual active zones. **a**, Multiphoton section through the eye of a zebrafish larva (7 d.p.f.) expressing iGluSnFR in a subset of bipolar cells. **b**, Line scan through a single terminal. No other terminals were in the vicinity. **c**, The kymograph (top) shows the intensity profile along the broken red line in **b** as a function of time. The broken green trace to the side of the kymograph shows the time-averaged intensity along the line, and the red and black traces are the two Gaussians that sum to best describe this spatial profile. The amplitude of each Gaussian at each time point was used to quantify the signal at each active zone, and is plotted in the traces below. The modulation in light intensity (20% contrast, 5 Hz) is shown immediately below. Note the large variations in the amplitude of glutamate transients. **d**, Expansion of the records within the boxes outlined in broken blue lines in **c**. Sometimes both active zones release glutamate, while on other occasions only active zones 1 or 2 are active (red dotted line). Qualitatively similar iGluSnFR signals were observed in 187 independent experiments.

shown in Fig. 2a. This approach significantly improved the SNR, but it was still necessary to set a threshold to distinguish events from noise, and this was chosen to be at least 3 standard deviations above the baseline (red line in Fig. 2b; detailed examples shown in Supplementary Figs. 4 and 5). After recording responses to stimuli lasting tens of seconds, we constructed histograms of event amplitudes for individual active zones, and these displayed several peaks (Fig. 2c). These peaks were spaced at equal intervals, indicating that transients of different sizes were composed of multiples of an underlying unit of transmission, very likely corresponding to a single vesicle (see also Supplementary Fig. 5). Based on these amplitude distributions, we partitioned events of different amplitudes into numbers of quanta using a Gaussian mixture model (Fig. 2b, lower trace).

Does the iGluSnFR reporter provide a linear read-out of glutamate release? Evidence that it does is provided in Fig. 2d, which plots the relative distance between the peaks in amplitude histograms of the type shown in Fig. 2c. These measurements were collected from six active zones in which large numbers of events of different size were recorded by applying stimuli of contrasts between 20% and 100%. The interpeak distances were not significantly different up to the ninth peak, the largest number that could be detected routinely within a distribution. These results indicate that the iGluSnFR signal provided a linear read-out of MVR up to at least nine vesicles, and that these estimates were not skewed by saturation of the reporter. In comparison, the largest events we observed from a sample of 51 synapses were equivalent to 11 quanta.

The electrical signals that generate release events of different amplitude remain to be established, but are likely to reflect the fact that the synaptic terminal is not a passive electrical compartment. Regenerative depolarizations can be generated within bipolar cell terminals by the same voltage-sensitive calcium channels that open

to trigger vesicle fusion²⁴. Moreover, calcium spikes can phase-lock to visual stimuli with millisecond precision^{24–26}.

Multiquantal events were a ubiquitous property of bipolar cells in both the On and Off channels of the retina (75 On synapses and 112 Off). We could not, however, clearly differentiate between strict CMVR (defined as events synchronized within a 100- μs time window) and a transient and very large increase in release probability (P_r), whereby vesicles are nonetheless released independently (MVR). The best SNR achievable with iGluSnFR only allowed us to distinguish events occurring at intervals greater than $\sim\!10\,\mathrm{ms}$ (Supplementary Figs. 6–8). Nonetheless, multiple vesicles released in a time window of $100\,\mu s$ or $10\,\mathrm{ms}$ will summate similarly on a postsynaptic ganglion cell because these have membrane time constants in the range of $10-40\,\mathrm{ms}^{27-29}$. Glutamate release events of different amplitude can therefore be considered as different synaptic symbols (Figs. 1c,d and 2b).

Reverse-correlation at an active zone: the transmitter-triggered average. Do synaptic events of varying amplitude play a role in encoding a visual stimulus? We began investigating this question by adapting an approach that has been widely used to explore the information represented by spikes; that is, calculation of the spike-triggered average (STA). The STA is obtained by reverse-correlating the spike train generated by a random (white noise) stimulus to the stimulus itself² and provides an estimate of the neurons tuning in the form of its linear filter in time and/or space. By reverse-correlating events measured using iGluSnFR in response to full-field noise, we instead constructed the transmitter-triggered average (TTA) to describe the output of an individual active zone. This approach to understanding the information encoded by vesicles at a synapse was first proposed in 2004 (ref. ³), but has not been realized until now. There is, however, a fundamental distinction

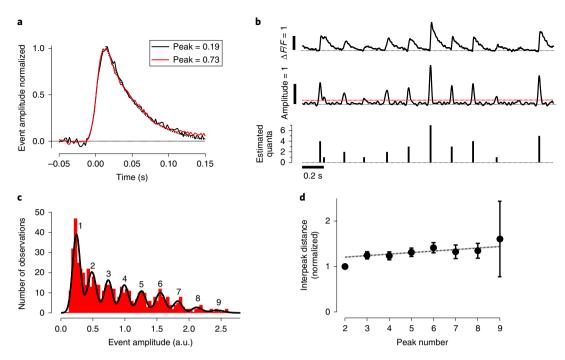


Fig. 2 | Glutamatergic events of different amplitudes were composed of varying numbers of quanta. a, The average of 15 iGluSnFR events with a peak amplitude of $\Delta F/F = 0.19 \pm 0.001$ (mean ± s.d.) superimposed on the average of 15 events approximately four times as large (peak $\Delta F/F = 0.73 \pm 0.01$). After normalization, the small and large events are superimposed. The decline from the peak occurs with a time constant of 44 ms (broken line). **b**, An example of the basic steps in the analysis by which events were counted and quantified. The raw iGluSnFR signal is shown in the top panel (stimulus at 5 Hz, 80% contrast; Savitsky–Golay filter, 21 ms). The middle trace shows the results of Wiener deconvolution using a kernel of unitary area and the shape shown in **a**. The time and amplitude of each event was obtained from the local maxima above a threshold (broken red line). Similar analyses could be carried out in 187 independent experiments. **c**, A histogram of event amplitudes for the active zone featured in **b** (n = 547 events accumulated using stimulus contrasts of 30, 80 and 100%). The black line is a fit of nine Gaussians, identified using a Gaussian mixture model. Note that the variance of successive Gaussians did not increase in proportion to the peak number. The first peak had a value of 0.24, and the distance between peaks averaged 0.26, indicating the existence of a quantal event equivalent to ~0.25. The amplitude of the quantal event averaged 0.23 ± 0.01 (mean ± s.e.m., n = 20 synapses from independent experiments). **d**, The mean distance between successive peaks (normalized to the distance between the first and second peaks) plotted as a function of the peak number. Collected results from n = 6 synapses. Points show the mean ± s.e.m. The broken line is a linear fit with a slope of 0.03 ± 0.12 (mean ± s.e.m.), which is not significantly different from zero. There were no signs of saturation of the iGluSnFR signal for events composed of up to nine quanta.

between the STA and TTA. The spike output of a neuron contains just two symbols (0 and 1), whereas in the case of bipolar cells, we are dealing with a vesicle code involving a number of symbols composed of one, two or three quanta, and so on. We therefore calculated the TTA separately for synaptic events composed of different numbers of vesicles (Fig. 3).

The TTA revealed that the more quanta within an event, the higher, on average, the temporal contrast driving it (Fig. 3b). The relationship between the number of quanta released per event (Q_e) and the contrast in the temporal filter (C) could be described by a first-order saturation in the form $C = C_{\max} \times Q_e / (Q_e + Q_{1/2})$, where $C_{\max} = 19\%$ and $Q_{1/2} = 2.4$ vesicles (Fig. 3c). Thus, MVR encodes one of the most fundamental properties of a visual stimulus, temporal contrast, in a direct way.

The TTA revealed a second distinction between events of different amplitude. That is, uniquantal events were characterized by temporal filters that were monophasic through both On and Off channels, corresponding to a low-pass filter. In contrast, the TTAs from On synapses were biphasic, corresponding to bandpass characteristics with a peak transmission at ~4 Hz (Fig. 3b). In other words, a period of lower light intensity immediately preceding higher intensity favored larger synaptic events through the On channel. Such antagonism in the time domain is commonly observed in the temporal receptive fields of neurons early in the visual system³⁰, where it has been proposed to underlie the suppression

of redundant signals³¹. Synchronizing the release of multiple vesicles is therefore expected to amplify the output of a bandpass filter to enhance the signaling of positive temporal contrast.

A hybrid rate and amplitude code. Given that bipolar cells can transmit visual information as changes in both the frequency and amplitude of release events, what are the relative contributions of these two coding strategies? To make such a comparison across synapses, it was important to take into account variations in the contrast-response function and to compare the modulation of rate and amplitude around similar parts of the operating range of each of the synapses. To achieve this, we proceeded in two steps. First, we constructed the contrast-response function of individual active zones using a 5 Hz stimulus in which contrast was increased in steps of 10% (Fig. 4a). The first measure of response that we used was the total number of quanta per cycle of the stimulus (Qc), from which we estimated the contrast generating the half-maximal response, $C_{1/2}$ (Fig. 4b). The parameter $C_{1/2}$ occurs at the steepest part of the contrast-response function and therefore also defines the range over which the synapse signaled a change in contrast with the highest sensitivity2. The second step was to make more detailed measurements of responses in a stimulus range of $\pm 10\%$ around $C_{1/2}$, for which different contrasts were applied in a pseudo-random order (Fig. 4c). Each contrast was applied for 2 s and again at 5 Hz, keeping the mean luminance constant throughout.

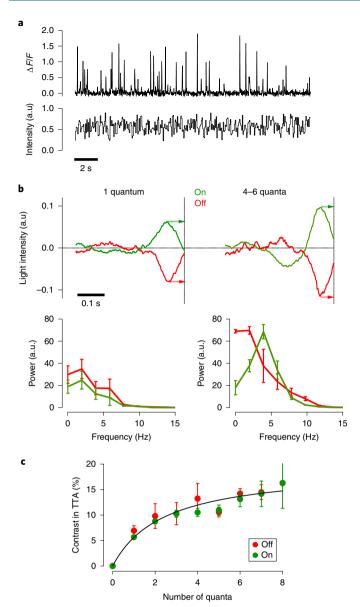


Fig. 3 | The TTA depends on the number of quanta in a release event.

a, Example of iGluSnFR signals (top) elicited by a white noise stimulus (bottom). Qualitatively similar responses were elicited in 17 independent experiments. **b**, Top: linear filters extracted by reverse-correlation of responses composed of one quantum (left) and four to six quanta (right) from On synapses (n=9) and Off synapses (n=8). Multiquantal events encoded larger modulations in intensity. The plots below show the power spectra of the linear filters (each point is the mean \pm s.e.m.). Through the Off channel, both uniquantal and multiquantal events were driven through low-pass filters, but in the On channel, multiquantal events were driven through a band-pass filter peaking at -4 Hz. **c**, Relation between the Michelson contrast represented in the TTA and the number of quanta released within the events used for reverse correlation in On and Off terminals (n=9 and 8, respectively; each point is the mean \pm s.e.m.). The relation could be described by a first-order saturation in the form $C = (C_{max})N/(N+N_{1/2})$, where $C_{max} = 19\%$ and $N_{1/2} = 2.4$.

A first inspection of the responses (Fig. 4a,c) indicated that although the amplitude of events fluctuated throughout each 2-s exposure to a contrast step, higher contrasts tended to increase both the frequency and amplitude of glutamatergic events. To assess how strongly these two components of the synaptic code were modulated,

 Q_c was factorized into the number of events per cycle (E_c) representing the contribution of the rate code) and the number of quanta per event (Q_e , the amplitude code). The relative change in E_c and Q_e varied across our sample of 55 active zones, and this could be illustrated by splitting the population according to a simple criterion: is the relative change in the E_c more or less than the Q_e ? In 38 out of 55 active zones, an increase in contrast caused the E_c to rise more steeply than the Q_e , and the average behavior of this subset is shown in Fig. 4d. In the remaining 17 synapses, the amplitude code was dominant and the Q_e rose more rapidly than the E_c (Fig. 4e). This functional hetreogeneity may reflect the fact that different bipolar cells vary in their intrinsic electrophysiological properties as well as in the retinal microcircuits in which they are embedded.

Active zones in which the amplitude of synaptic events was modulated more strongly displayed a second notable property. When the average frequency of events reached a maximum (~15 Hz), further increases in contrast were represented wholly as increases in the number of quanta released per event (indicated by the double-arrow line in Fig. 4e). MVR therefore extends the range of contrasts that can be signaled beyond those allowed by the rate code alone.

The amplitude code improves the temporal resolution of synaptic transmission. Synaptic events of larger amplitude improved the temporal precision with which the visual signal was transmitted. An example of this phenomenon is shown in Fig. 5a, which compares the timing of events relative to the phase of a 5-Hz stimulus for contrasts of 20% and 100%. Uniquantal events displayed a standard deviation ('temporal jitter') of 24-28 ms over a range of contrasts, while events composed of 7 or more quanta jittered by as little as 2.5 ms (Fig. 5b). The degree to which synaptic events were consistent in time also depended on the contrast of the stimulus eliciting the event. For instance, the largest packets of glutamate consistently observed at 20% contrast contained 8 quanta with a jitter of 12.1 ± 1.4 ms, while 8 quantal events at 100% contrast jittered by just 4.6 ± 0.6 ms (Fig. 5b). The temporal precision of the largest events was similar to that displayed by the spike trains in postsynaptic ganglion cells, which also show precisions down to a few milliseconds when responding to high contrast³². MVR may therefore be one of the mechanisms that cause the spike output of ganglion cells to become more reliable as contrast is increased.

Multivesicular events transmit more information per vesicle.

The responses of sensory neurons can be highly variable, but they nonetheless provide information because the distribution of possible responses alters for different stimuli^{2,33}. To begin exploring how synapses of bipolar cells provided information through MVR, we measured the change in the distribution of synaptic event amplitudes at high and low contrasts (20% and 100%). Figure 6a shows that unitary events predominated at low contrast, while an event composed of two quanta was equally likely whether the contrast was 20% or 100%. At the other extreme, events composed of more than 8 quanta indicated a contrast above 20% with a very high degree of certainty. The largest events observed at 100% contrast were composed of 11 quanta.

A more systematic investigation of changes in the distribution of events was made by applying information theory^{33,34}. Using the stimulus protocol shown in Fig. 4c, we quantified the specific information between a set of stimuli of varying contrasts and release events containing different numbers of quanta. Vesicles released individually carried an average of 0.125 bits of information, which is, as expected, significantly less than the 1–3.6 bits transmitted per spike in postsynaptic ganglion cells^{32,35}. Larger events transmitted progressively more information (Fig. 6b) because they were rarer overall and correlated with higher contrasts rather than occurring randomly (Figs. 3–5). The relation between the specific information (I_2 , bits) and Q_e , the number of vesicles comprising the event, could be described as a power function with an exponent of ~3, indicating

ARTICLES

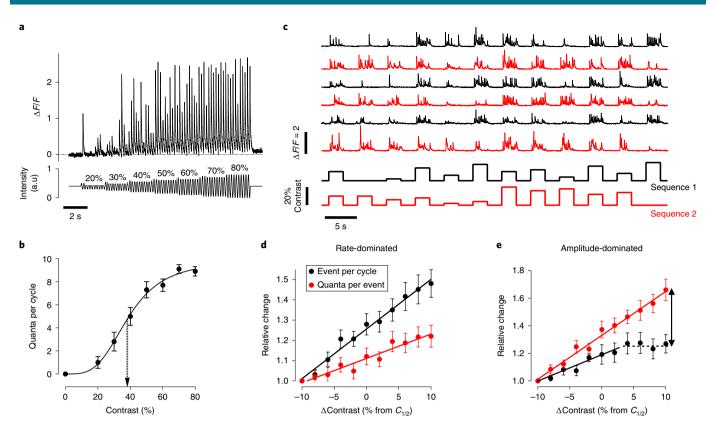


Fig. 4 | The relative contributions of coding by rate and amplitude. a, Example of the iGluSnFR signal (top) elicited by a full-field stimulus of increasing contrast delivered at 5 Hz (bottom). This protocol was used to quickly assess the half-point of the contrast-response function ($C_{1/2}$) in 38 independent experiments. **b**, Contrast-response function extracted from the synapse in **a**, with response (R) quantified as the total number of quanta per cycle of the stimulus. The curve is a Hill equation of the form $C = (R_{max})(C^h/C^h + C^h_{1/2})$, where $R_{max} = 9.8$ quanta per cycle (49 quanta s⁻¹), h = 4.5 and $C_{1/2} = 39\%$ (broken arrow). Each point shows the mean \pm s.e.m. for n = 10 cycles of the stimulus. **c**, A stimulus set for quantifying the modulation of the rate and amplitude of synaptic events in 55 independent experiments. Stimuli of varying contrasts were selected to span $\pm 10\%$ of the range around which the contrast sensitivity was highest in steps of 2%. Each stimulus lasted 2 s, and all were at the same mean luminance and a frequency of 5 Hz. Note the variations in the frequency and amplitude of events at different contrast levels. Stimuli were applied in two different sequences, which were repeated alternately. The same stimulus set was used to estimate mutual information (Fig. 6). **d**, The relative change in synaptic activity around $C_{1/2}$. The average number of events per cycle (E_c) is compared with the average number of quanta per event (E_c). Stimuli were delivered in 2-s episodes with a 2-s rest, as shown in Fig. 4a. In these E_c which then saturated (broken line) such that further increases in contrast were signaled only by increases in the number of quanta per event (vertical arrow). Points in **d** and **e** show the mean \pm s.e.m.

that larger synaptic events transmitted more information per vesicle. To quantify this idea more directly, we divided the amount of information in an event by the number of vesicles it contained to provide a measurement of 'vesicle efficiency', after which we averaged across synapses to quantify the trend. The change in this quantity is plotted in Fig. 6c after normalizing to the value measured in the same synapse for one vesicle. Synaptic events composed of five vesicles carried, on average, four times as much information per vesicle compared with unitary events.

Discussion

Achieving single-vesicle resolution with iGluSnFR provided the opportunity to investigate the synaptic code by counting quanta released at individual active zones, similar to how electrophysiology allows the counting of spikes to investigate the neural code^{1,2}. This in vivo approach demonstrated that the release of two or more vesicles within a time window of 10 ms or less is a fundamental aspect of the strategy by which ribbon synapses of bipolar cells recode an analog signal for transmission across the synapse. MVR discretizes the graded signal arriving down the axon into about 12 output values, with larger modulations of intensity increasing the average

number of vesicles released within an event (Fig. 6a). As a result, these synapses implement a hybrid coding strategy that represents a stimulus as changes in both the rate and amplitude of synaptic events (Figs. 3–4). The amplitude code complements the rate code by increasing the temporal accuracy (Fig. 5) and operating range (Fig. 4) of synapses transmitting the visual signal to retinal ganglion cells, as well as the efficiency of information transmission measured as bits per vesicle (Fig. 6).

An amplitude code at ribbon synapses. Neurons in the brain can be characterized as 'leaky integrators' in which inputs converging on the dendrites summate increasingly less effectively the further apart they arrive in time and space². The time window for effective summation is set by the membrane time constant, which varies between 10 ms and 40 ms in retinal ganglion cells²⁴⁻²⁶. In this context, MVR can be viewed as a mechanism by which glutamatergic vesicles are summed presynaptically and at one site over a time window of one time constant or less, causing them to act more effectively on the postsynaptic neuron.

One factor determining how efficiently the presynaptic summation of glutamate converts into a postsynaptic current is whether

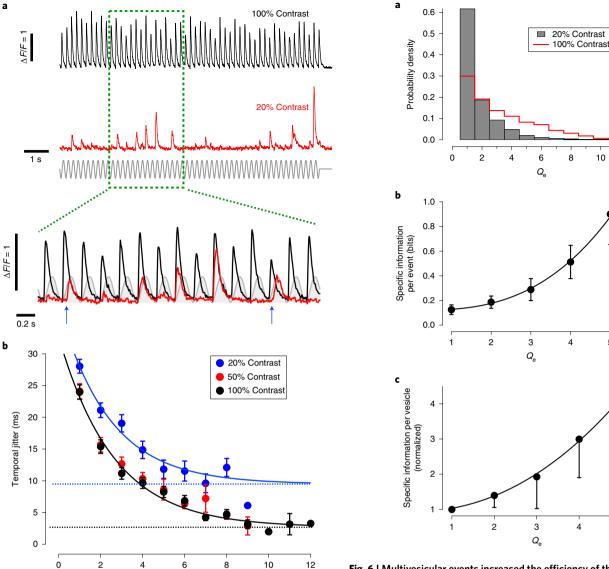


Fig. 5 | Multivesicular events increased the temporal precision of synaptic transmission. a, Top: iGluSnFR signals from a terminal stimulated at 100% and 20% contrast. Note the large variation in event amplitudes at lower contrast. Bottom: expanded time scale showing responses at 20% and 100% contrast superimposed on the phase of the stimulus (gray). Events composed of fewer quanta are less synchronized than multiquantal events. Two examples of events occurring at phases different to the majority are indicated by the blue arrows. Qualitatively similar responses were observed in 70 independent experiments. **b**, Temporal jitter (ms) as a function of the number of quanta per event (Q_e) at contrasts of 20% (n=57 synapses), 50% (n=61) and 100% (n=63). Points show the mean \pm s.e.m. Events composed of more quanta exhibit lower jitter at a given contrast, and events of a given amplitude exhibit lower jitter at higher contrasts. The relationship between s and $Q_{\rm e}$ could be described as $s = s_{\rm min} + (s_{\rm max} - s_{\rm min})$ $(e^{-Q}_{e/k})$. At 100% contrast, $s_{min} = 2.7 \,\text{ms}$ (broken black line), $s_{max} = 30 \,\text{ms}$ and k = 2.6 quanta. At 20% contrast, $s_{min} = 9.5$ ms (broken blue line), $s_{max} = 28 \text{ ms}$ and k = 2.4 quanta.

Quanta

postsynaptic receptors saturate. Saturation does not appear to occur at all amacrine cells postsynaptic to rod-driven bipolar cells¹⁸ or auditory nerve fibers postsynaptic to hair cells²⁰, but may occur at AMPA receptors on retinal ganglion cells³⁶. However, multivesicular events

Fig. 6 | Multivesicular events increased the efficiency of the vesicle code.

5

a, Changes in the distribution of event amplitudes elicited by low (20%) and high (100%) contrast. Events composed of more than 8 quanta were not observed at 20% contrast. The largest events observed at 100% contrast were composed of about 11 quanta. **b**, Specific information per event (i, bits) as a function of Q_e , the number of vesicles comprising the event. Each point shows the mean \pm s.e.m., and the curve describing the points is a power function of the form $i=i_0+A\times Q_e^{\times}$, with $i_0=0.12$ bits, A=0.008 and x=2.8. Results pooled from n=17 synapses. **c**, Specific information per vesicle normalized to the value measured for a uniquantal event (i'). Each point shows the mean \pm s.e.m., and the curve describing the points is a power function of exponent 2.1. Results from the same 17 synapses in **b**.

activate extrasynaptic NMDA receptors on ganglion cell dendrites much more effectively than single vesicles³⁷. We therefore require experimental measurements that explore how varying numbers of vesicles released within an event alter the input current and probability of spiking of a postsynaptic cell. Such measurements will be difficult using electrophysiology alone because a patch pipette on the cell body cannot easily distinguish true MVR from the coincidental arrival of vesicles released from different synapses. A direct and reliable assessment of the postsynaptic effects of MVR has, however, been possible in the auditory system of the bullfrog, in which afferents contacting only one hair cell ribbon can be patched. Measuring

the synaptic currents injected into the afferent relative to the currents needed to depolarize the fiber beyond the threshold has demonstrated that larger glutamatergic events can reliably trigger spikes²⁰.

Cellular mechanisms underlying MVR. The cellular mechanisms synchronizing the release of two or more vesicles remain to be discovered. It is, however, notable that the synaptic terminals of bipolar cells in zebrafish^{27,28} and mice²⁹ can convert the analog signal arriving down the axon into regenerative calcium spikes that cause large increases in presynaptic calcium. It is also known that strong activation of presynaptic calcium channels can release a large fraction of the readily releasable pool (RRP) of vesicles within a few milliseconds¹³⁻¹⁵. For instance, the RRP at an individual active zone has been estimated to be about 18 vesicles in goldfish bipolar cells³⁸, indicating that MVR events ranging up to 11 could be generated through the fast release of vesicles that are already docked and primed. Variations in the degree to which the rate and amplitude codes were modulated in different synapses (Fig. 4) might reflect intrinsic factors such as the complement of ion channels in the terminal membrane or the size of the RRP, or the extrinsic influence of the microcircuit in which the bipolar cell is embedded, including inhibition from amacrine cells.

The ionic mechanisms that generate neural signals and the synaptic processes that transmit them constitute a major energetic cost to the brain^{39,40}. Meanwhile, the evolutionary pressure exerted by the need to transmit information in an energy-efficient manner has provided a general principle by which to understand the design of sensory circuits^{12,41,42}. This principle provides an explanation for the use of analog signaling to transmit early visual and auditory information, whereby graded changes in membrane potential transmit information more efficiently than spikes^{12,43}. It would be interesting to carry out an energy budget for MVR at synapses of bipolar cells where the efficiency of transmission could be quantified as bits per unit energy^{44,45}, but this will require a better understanding of the underlying electrical events.

The time resolution of measurements with iGluSnFR do not rule out the possibility that the release of multiple vesicles is synchronized to submillisecond time scales by coordinated MVR. The mechanisms that might underlie CMVR are far from clear^{17–19}, but appear distinct from more common mechanisms of synaptic transmission by deviating from a Poisson process in which vesicles fuse independently of each other. A third suggestion, based on experiments on ribbon synapses of hair cells, is that variable amounts of glutamate are released from a single vesicle because of the dynamics of a fusion pore⁴⁶. This idea runs counter to evidence that bipolar cells transmit by full collapse of vesicles into the membrane surface^{47,48} and to the clear quantization of glutamate transients that we have observed with iGluSnFR (Fig. 2).

The presence of ribbon structures at synapses driven by analog signals correlates with a second functional specialization, a continuous mode of operation that allows the transmission of sensory information to be maintained over prolonged periods^{9,49} (Figs. 4 and 5). The ribbon is thought to support continuous release by capturing vesicles from a mobile pool in the cytoplasm³⁸ and then transporting them to the active zone¹⁹. The maximum rates of continuous release that we observed using iGluSnFR were in the range of 50–100 vesicles s⁻¹ per active zone, which is in agreement with measurements made using the membrane dye FM1–43 (refs. ^{14,49}) and the fluorescent reporter protein sypHy²³. The efficient replenishment of vesicles within the RRP is an essential aspect of ribbon synapse operation in the retina and allows MVR to continue to operate over prolonged periods, as shown in Figs. 3–5.

Amplitude and rate coding in relation to the output of the retina. For vision to be useful, information about important stimuli must be transmitted over an appropriately short time window. Multiquantal

events were rare and strongly dependent on temporal contrast, so their arrival provided more information about a preceding stimulus than individually released vesicles (Fig. 6). Crucially, MVR also encoded the timing of a stimulus more precisely, whereby the largest glutamatergic events jittered by just a few milliseconds relative to a stimulus (Fig. 5), which is similar to the spike responses observed in postsynaptic ganglion cells. The spike trains generated by naturalistic stimuli often consist of brief increases in firing frequency from longer background periods of silence, making it difficult to describe activity as a time-varying rate. Rather, it has been suggested that "firing events containing single spikes or bursts of spikes are elicited precisely enough to convey distinct packets of visual information, and hence may constitute the fundamental symbols in the neural code of the retina"³². It seems possible that these symbols originate in the amplitude code of bipolar cell synapses.

To understand why a coding strategy based on amplitude might have arisen, it is useful to think about the temporal requirements of a simple rate code. If a Poisson synapse releasing all vesicles independently encodes an event by changing the rate of vesicle release by a factor k (from R to kR), the SNR achieved over an observation time Δt will be as follows:

$$SNR = \frac{(kR\Delta t - R\Delta t)}{\sqrt{(kR\Delta t + R\Delta t)}}$$

(Online Methods, equation 4). A change in contrast from a $C_{1/2}$ of -10% to a $C_{1/2}$ of +10% increased the rate of vesicle release by an average factor of k=1.8 ($Q_e \times E_c$; Fig. 4d), from a basal rate R of no more than 20 vesicles s⁻¹. To detect such a change with a SNR of 3 (a fairly good reliability) would require an observation time of ~2 s. In comparison, any release event with amplitudes greater than 8 quanta would signal an increase in contrast beyond 20% within milliseconds (Fig. 6a). This simple comparison illustrates one of the potential advantages of recoding an analog signal using symbols varying in amplitude rather than by varying the frequency of digital events. That is, an unexpected symbol immediately imparts new information, while a stochastic rate code must be observed over a time window sufficiently long to establish a significant change relative to the background. Further analyses of the statistics of MVR will probably shed more light on the properties of the vesicle code transmitting visual information.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41593-019-0403-6.

Received: 22 May 2018; Accepted: 9 April 2019; Published online: 20 May 2019

References

- Rieke, F., Bialek, W., Warland, D. & de Ruyter van Steveninck, R. Spikes: Exploring the Neural Code (MIT Press, 1999).
- Dayan, P. & Abbott, L. R. Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems (MIT Press, 2005).
- Abbott, L. F. & Regehr, W. G. Synaptic computation. *Nature* 431, 796–803 (2004).
- Del Castillo, J. & Katz, B. Quantal components of the end-plate potential. J. Physiol. 124, 560–573 (1954).
- Choi, S. Y. et al. Encoding light intensity by the cone photoreceptor synapse. Neuron 48, 555–562 (2005).
- Malagon, G., Miki, T., Llano, I., Neher, E. & Marty, A. Counting vesicular release events reveals binomial release statistics at single glutamatergic synapses. J. Neurosci. 36, 4010–4025 (2016).
- Masland, R. H. The neuronal organization of the retina. Neuron 76, 266–280 (2012).

- 8. Fuchs, P. A. Time and intensity coding at the hair cell's ribbon synapse. *J. Physiol.* **566**, 7–12 (2005).
- 9. Lagnado, L. & Schmitz, F. Ribbon synapses and visual processing in the retina. *Annu. Rev. Vis. Sci.* 1, 235–262 (2015).
- Jackman, S. L. et al. Role of the synaptic ribbon in transmitting the cone light response. *Nat. Neurosci.* 12, 303–310 (2009).
- Freed, M. A. Quantal encoding of information in a retinal ganglion cell. J. Neurophysiol. 94, 1048–1056 (2005).
- 12. Sterling, P. & Laughlin, S. B. Principles of Neural Design (MIT Press, 2015).
- Mennerick, S. & Matthews, G. Ultrafast exocytosis elicited by calcium current in synaptic terminals of retinal bipolar neurons. *Neuron* 17, 1241–1249 (1996).
- Neves, G. & Lagnado, L. The kinetics of exocytosis and endocytosis in the synaptic terminal of goldfish retinal bipolar cells. *J. Physiol.* 515, 181–202 (1999).
- Burrone, J. & Lagnado, L. Synaptic depression and the kinetics of exocytosis in retinal bipolar cells. J. Neurosci. 20, 568–578 (2000).
- Rudolph, S., Tsai, M. C., von Gersdorff, H. & Wadiche, J. I. The ubiquitous nature of multivesicular release. *Trends Neurosci.* 38, 428–438 (2015).
- Glowatzki, E. & Fuchs, P. A. Transmitter release at the hair cell ribbon synapse. *Nat. Neurosci.* 5, 147–154 (2002).
- Singer, J. H., Lassova, L., Vardi, N. & Diamond, J. S. Coordinated multivesicular release at a mammalian ribbon synapse. *Nat. Neurosci.* 7, 826–833 (2004).
- Mehta, B., Snellman, J., Chen, S., Li, W. & Zenisek, D. Synaptic ribbons influence the size and frequency of miniature-like evoked postsynaptic currents. *Neuron* 77, 516–527 (2013).
- Li, G. L., Cho, S. & von Gersdorff, H. Phase-locking precision is enhanced by multiquantal release at an auditory hair cell ribbon synapse. *Neuron* 83, 1404–1417 (2014).
- DeWeese, M. R. & Meister, M. How to measure the information gained from one symbol. Netw. Comput. Neural Syst. 10, 325–340 (1999).
- Marvin, J. S. et al. An optimized fluorescent probe for visualizing glutamate neurotransmission. *Nat. Methods* 10, 162–170 (2013).
- Odermatt, B., Nikolaev, A. & Lagnado, L. Encoding of luminance and contrast by linear and nonlinear synapses in the retina. *Neuron* 73, 758–773 (2012).
- 24. Dreosti, E., Esposti, F., Baden, T. & Lagnado, L. In vivo evidence that retinal bipolar cells generate spikes modulated by light. *Nat. Neurosci.* 14, 951, 952 (2011)
- Baden, T., Esposti, F., Nikolaev, A. & Lagnado, L. Spikes in retinal bipolar cells phase-lock to visual stimuli with millisecond precision. *Curr. Biol.* 21, 1859–1869 (2011).
- Baden, T., Berens, P., Bethge, M. & Euler, T. Spikes in mammalian bipolar cells support temporal layering of the inner retina. *Curr. Biol.* 23, 48–52 (2013).
- Taylor, W. R., Mittman, S. & Copenhagen, D. R. Passive electrical cable properties and synaptic excitation of tiger salamander retinal ganglion cells. *Vis. Neurosci.* 13, 979–990 (1996).
- Robinson, D. W. & Chalupa, L. M. The intrinsic temporal properties of alpha and beta retinal ganglion cells are equivalent. *Curr. Biol.* 7, 366–374 (1997).
- O'Brien, B. J., Isayama, T., Richardson, R. & Berson, D. M. Intrinsic physiological properties of cat retinal ganglion cells. *J. Physiol.* 538, 787–802 (2002).
- Meister, M. & Berry, M. J. II. The neural code of the retina. *Neuron* 22, 435–450 (1999).
- Srinivasan, M. V., Laughlin, S. B. & Dubs, A. Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 216, 427–459 (1982).
- Berry, M. J., Warland, D. K. & Meister, M. The structure and precision of retinal spike trains. *Proc. Natl Acad. Sci. USA* 94, 5411–5416 (1997).
- 33. Stone, J. V. Principles of Neural Information Theory: Computational Neuroscience and Metabolic Efficiency (Sebtel Press, 2018).
- Strong, S. P., Koberle, R., de Ruyter van Steveninck, R. & Bialek, W. Entropy and information in neural spike trains. *Phys. Rev. Lett.* 80, 197–200 (1998).

- Koch, K. et al. How much the eye tells the brain. Curr. Biol. 16, 1428–1434 (2006).
- Sagdullaev, B. T., McCall, M. A. & Lukasiewicz, P. D. Presynaptic inhibition modulates spillover, creating distinct dynamic response ranges of sensory output. *Neuron* 50, 923–935 (2006).
- Chen, S. & Diamond, J. S. Synaptically released glutamate activates extrasynaptic NMDA receptors on cells in the ganglion cell layer of rat retina. J. Neurosci. 22, 2165–2173 (2002).
- Holt, M., Cooke, A., Neef, A. & Lagnado, L. High mobility of vesicles supports continuous exocytosis at a ribbon synapse. *Curr. Biol.* 14, 173–183 (2004).
- Laughlin, S. B., de Ruyter van Steveninck, R. & Anderson, J. C. The metabolic cost of neural information. *Nat. Neurosci.* 1, 36–41 (1998).
- Attwell, D. & Laughlin, S. B. An energy budget for signaling in the grey matter of the brain. J. Cereb. Blood Flow. Metab. 21, 1133–1145 (2001).
- Barlow, H. B. in Sensory Communication (ed. Rosenblith, W. A.) 217–234 (MIT Press, 1961).
- 42. Niven, J. E. & Laughlin, S. B. Energy limitation as a selective pressure on the evolution of sensory systems. *J. Exp. Biol.* **211**, 1792–1804 (2008).
- 43. de Ruyter van Steveninck, R. & Laughlin, S. B. The rate of information transfer at graded-potential synapses. *Nature* **379**, 642 (1996).
- 44. Harris, J. J., Jolivet, R. & Attwell, D. Synaptic energy use and supply. *Neuron* 75, 762–777 (2012).
- Harris, J. J., Jolivet, R., Engl, E. & Attwell, D. Energy-efficient information transfer by visual pathway synapses. Curr. Biol. 25, 3151–3160 (2015).
- Chapochnikov, N. M. et al. Uniquantal release through a dynamic fusion pore is a candidate mechanism of hair cell exocytosis. *Neuron* 83, 1389–1403 (2014).
- Llobet, A., Beaumont, V. & Lagnado, L. Real-time measurement of exocytosis and endocytosis using interference of light. *Neuron* 40, 1075–1086 (2003).
- Zenisek, D., Davila, V., Wan, L. & Almers, W. Imaging calcium entry sites and ribbon structures in two presynaptic cells. J. Neurosci. 23, 2538–2548 (2003).
- Lagnado, L., Gomis, A. & Job, C. Continuous vesicle cycling in the synaptic terminal of retinal bipolar cells. *Neuron* 17, 957–967 (1996).

Acknowledgements

The authors express many thanks to J. Johnston for discussions and to H. Smulders and N. Bashford for looking after the zebrafish. Thanks are also given to M. Meyer for the gift of the constructs. This work was supported by grants to L.L. from the Wellcome Trust (102905/Z/13/Z) and an EU International Training Network (H2020-MSCA-ITN-2015-674901).

Author contributions

B.J. wrote software, conceived and designed the experiments and helped prepare the manuscript. L.D. performed experiments and analyses. J.M.-D. carried out two-photon imaging experiments and analyses and helped prepare the manuscript. S.-H.S. performed molecular biology experiments, fish transgenesis and initial functional analyses. L.L. conceived the project, designed experiments, wrote software, analyzed results and prepared the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-019-0403-6.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to L.L.

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

Methods

Zebrafish husbandry. Fish were raised and maintained under standard conditions on a 14h light–10h dark cycle²³. To aid imaging, fish were heterozygous or homozygous for the *casper* mutation, which results in hypopigmentation, and they were additionally treated with 1-phenyl-2-thiourea (200 µM final concentration; Sigma) from 10 h post-fertilization (h.p.f.) to further reduce pigmentation. All animal procedures were performed in accordance with the Animal Act 1986 and the UK Home Office guidelines, and with the approval of the University of Sussex Animal Welfare and Ethical Review Board. More information about experimental design and reagents is available in the Nature Research Reporting Summary.

Molecular biology. The zebrafish *ribeye a* (*ctbp2*) promoter was used to drive the expression of iGluSnFR in all neurons with ribbon synapses. Tg(-1.8ctbp2:Gal4VP16_BH) fish that express the transcriptional activator protein Gal4VP16 were generated by co-injection of I-SceI meganuclease and endofree purified plasmid into wild-type zebrafish with a mixed genetic background. A myocardium-specific promoter that drives the expression of mCherry protein was additionally cloned into the plasmid to allow for phenotypic screening of founder fish. Tg(10×UAS:iGluSnFR_MH) fish that express the glutamate sensor iGluSnFR²² under the regulatory control of the 10× UAS enhancer elements were generated by co-injection of purified plasmid and tol2 transposase RNA into offspring of AB wild-type fish outcrossed to *casper* wild-type fish. The sequences for the myocardium-specific promoter driving the expression of enhanced green fluorescent protein (GFP; mossy heart) were added to the plasmid to facilitate the screening process. Supplementary Table 1 shows plasmid and primer information.

Multiphoton imaging in vivo. Zebrafish larvae (7–9 days post-fertilization (d.p.f.)) were immobilized in 3% low-melting-point agarose (Biogene) in E2 medium on a glass coverslip (0 thickness) and mounted in a chamber where they were superfused with E2 (ref. 23). Imaging was carried out using a two-photon microscope (Scientifica) equipped with a mode-locked titanium-sapphire laser (Chameleon, Coherent) tuned to 915 nm and an Olympus XLUMPlanFI ×20 water immersion objective (numerical aperture (NA) of 0.95). To prevent eye movements, the ocular muscles were paralyzed via the injection of 1 nl of α -bungarotoxin (2 mg ml $^{-1}$) behind the eye. The SNR for imaging was optimized by collecting photons through both the objective and a sub-stage oil condenser (Olympus; NA of 1.4). Emission was filtered through GFP filters (HQ 525/50, Chroma Technology) before detection using GaAsP photomultipliers (H7422P-40, Hamamatsu). The signal from each detector passed through a current-to-voltage converter, and then the two signals were added by a summing amplifier before digitization. Scanning and image acquisition were controlled using the software ScanImage (v.3.6)50. Images were typically acquired at 10 Hz (128×100 pixels per frame, 1 ms per line) while line scans were acquired at 1 kHz.

Full-field light stimuli were generated by an amber light-emitting diode (peak wavelength = $590\,\mathrm{nm}$, Thorlabs), filtered through a $590/10\,\mathrm{nm}$ bandpass filter (Thorlabs), and delivered through a light guide placed close to the eye of the fish. Stimuli were normally delivered as modulations around a mean intensity of $\sim 320\,\mathrm{nW\,mm^{-2}}$, and the microscope was synchronized to visual stimulation.

To minimize the possibility of recording iGluSnFR signals from adjacent synapses, we used zebrafish in which only a fraction of bipolar cells were expressing iGluSnFR and chose terminals that were spatially isolated from others expressing the reporter, as shown in Fig. 1a and Supplementary Figs. 9 and 10. These transients were generated by the electrical signal arriving from the cell body rather than by glutamate spillover from neighboring neurons because they were destroyed by ablating the soma of the cell (Supplementary Fig. 9).

Counting ribbons. To assess the probability of conflating signals from different active zones within one terminal, we measured the numbers and distribution of synaptic ribbons that held vesicles close to the sites of fusion¹². Larvae were fixed at 7 d.p.f. with 4%paraformaldehyde (in PBS) for 35 min. The eyeballs were removed and sliced into two equal parts, which were then permeabilized for 10 min in PBS containing 0.5% Triton X-100. The pieces of eyecup were incubated for 3 days in primary antibodies specific for ribeye A (1/200 dilution) and GFP (1/500 dilution, targeting iGluSnFR; Torrey Pines Biolabs, TP401, lot number 42704), followed by incubation overnight in a secondary antibody, all at 4 °C on a shaker. The primary antibody against zebrafish ribeye A ([C]-YNQGYLDRPDPRNIRK-[N]) was an isolated chicken IgY fragment produced by Cambridge Research Biochemicals (clone 2575) and has been validated in the laboratories of L. Lagnado, R. Wong and L. Godinho. The secondary antibody for ribeye A was Alexa-546 antichicken (Molecular Probes, catolog no. A11040, lot no. 682609, dilution 1:1,000), and for GFP we used Alexa-647 anti-rabbit (Molecular Probes, A21244, lot no. A21244, dilution 1:1,000). A confocal microscope (Leica SP8) was used to image through the retina using z-steps of 0.25 µm, and ImageJ was used to make threedimensional (3D) reconstructions of individual terminals using the segmentation plugin 3D Blob. The number of ribbons in a sample of 27 terminals averaged 4.6 ± 2.4 (mean \pm s.d.). The distance from one ribbon to its neighbors averaged $0.96 \pm 0.4 \,\mu\text{m}$. The probability of conflating signals from two active zones was estimated to be less than 8% (see below).

Statistics. No statistical methods were used to predetermine sample sizes. Experiments were repeated until trends in results were clear, and this resulted in sample sizes at least equivalent to previous publications $^{15,18-20}$. All data are given as the mean \pm s.e.m. unless otherwise stated in the figure legends. All statistical tests were calculated using in-built functions in IgorPro (Wavemetrics) and met the assumptions of the statistical tests used. When data were not normally distributed, we used nonparametric tests. All tests were two-sided and significance was defined as $P\!<\!0.05$. Data collection was not randomized because all experiments were carried out within one set of animals. Delivery of different stimuli was randomized where appropriate. Data collection and analyses were not carried out blind to the conditions of the experiments. Data were only excluded from the analysis if the SNR of the iGluSnFR signals elicited at a given synapse was not sufficient to detect unitary responses to visual stimuli with a SNR of at least three.

Calculation of temporal jitter. To compute the temporal jitter of the glutamatergic events, we first calculated the vector strength (VS) as follows:

$$VS_{q} = \frac{1}{N_{q}} \sqrt{\left(\sum_{i=1}^{N_{q}} \cos\left(\frac{2\pi t_{q_{i}}}{T}\right)\right)^{2} + \left(\sum_{i=1}^{N_{q}} \sin\left(\frac{2\pi t_{q_{i}}}{T}\right)\right)^{2}}$$
(1)

where t_{qi} is the time of the ith q-quantal event, T is the stimulus period, and N_q is the total number of events of composed of q-quanta. The temporal jitter (TJ) can then be computed using equation (2):

$$TJ_{q} = \frac{\sqrt{2(1 - VS_{q})}}{2\pi f} \tag{2}$$

where f is the stimulus frequency.

The SNR associated with a change in the rate of a Poisson process. Imagine the mean rate of a Poisson process, R, changes by a factor k. The signal, S, generated by comparing two observation times Δt will be the change in the mean number of events counted in each period $(kR\Delta t - R\Delta t)$, and the variance in that signal will be the sum of the number of events counted in each period $(kR\Delta t + R\Delta t)$. Defining the SNR in the same way as the discriminability (d') used in signal detection theory 45 , we have the following:

$$SNR = \frac{S}{\sqrt{\text{variance}}}$$
 (3)

which yields

$$SNR = \frac{(kR\Delta t - R\Delta t)}{\sqrt{(kR\Delta t + R\Delta t)}}$$
(4)

which can be rearranged to obtain the time period Δt required to obtain a given SNR. as follows:

$$\Delta t = \frac{\text{SNR}^2}{R} \frac{(k+1)}{(k-1)^2}$$
 (5)

Calculation of the TTA. The TTA for an event containing a specific number of quanta (s_q) was calculated by averaging the stimuli that preceded each release event of that type as follows:

$$s_q = \frac{1}{n_q} \sum_{i=1}^{n_q} s(t_i)$$
 (6)

where q is the quantal event type, n_q is the number of n-quantal events within the recording, and s(t) is the stimulus window ending at time t. We used Gaussian white noise in which intensities were drawn from a Gaussian distribution updated at a specific stimulus frame rate (varied between 10 and 30 Hz). The Gaussian white noise was then discretized into eight equally spaced bins to produce a stimulus that approximates a Gaussian distribution while effectively driving synaptic responses.

Calculations based on information theory. To quantify the amount of information about a visual stimulus that is contained within the sequence of release events from an active zone, we first needed to convert bipolar cell outputs into a probabilistic framework from which we could evaluate the specific information (I_2), a metric that quantifies how much information about one random variable is conveyed by the observation a specific symbol of another random variable²¹. The time series of quantal events was converted into a probability distribution by dividing them into time bins of 20 ms, such that each bin contained either zero events or one event of an integer amplitude. We then counted the number of bins containing events of amplitude one, two or three, and so on. By dividing the number of bins of each type by the total number of bins for each different stimulus, we obtained the conditional

NATURE NEUROSCIENCE

distribution of Q given S, $p(Q \mid S)$, where Q is the random variable representing the quanta per bin, and S is the random variable representing the stimulus contrasts presented throughout the course of the experiment. We then computed the joint probability distribution by the chain rule for probability (given the experimentally defined uniform distribution of stimuli S) as follows:

$$p(S,Q) = p(Q \mid S)p(S)$$
(7)

To convert this distribution into the conditional distribution of S given Q, we used the definition of the conditional distribution as follows:

$$p(S \mid Q) = \frac{p(S, Q)}{p(Q)} \tag{8}$$

From these distributions, we computed the specific information as the difference between the entropy of the stimulus S minus the conditional entropy of the stimulus given the observed symbol in the response *q*:

$$I_2(S, q) = H(S) - H(S \mid q)$$
 (9)

$$I_2(S, q) = -\sum_{s \in S} p(s) \log p(s) + \sum_{s \in S} p(s \mid q) \log p(s \mid q)$$
(10)

representing the amount of information observing each quantal event type $q \in Q$ carries about the stimulus distribution S.

Measuring entropy and mutual information from neural responses can be a challenging problem^{33,51}. Estimates require sampling from an unknown discrete probability distribution and, in many cases, recording sufficient samples to observe all non-zero probability events is neither tractable nor practical. The biases introduced by undersampling can be a particular problem when the full support of the distribution (all values that map to non-zero probabilities) is high. Within the past few decades, various approaches to correcting biases in information theoretic analyses have been developed⁵¹. However, as the distributions of interest in this work have both a small support and are well sampled, we have opted to use standard estimates for the quantities of interest.

Analysis sequence for the quantal decomposition of iGluSnFR signals. An analysis package was created within Igor Pro (Wavemetrics) to detect and quantize glutamatergic events, comprising six major steps:

- 1. Separation of regions of interest (ROIs) by spatial decomposition
- Time series extraction by weighted averaging 2.
- 3. Baseline correction and calculation of $\Delta F/F$
- 4. Identification of events by Wiener deconvolution
- 5. Extraction of events
- Amplitude clustering to create a time series of quantized events 6.

The analysis begins with an $x \times L$ matrix in which x is the number of pixels per line and L is the number of lines, and provides a final output of a set of E event times and AQ estimated quanta for each event in each active zone. An example of the analysis proceeding from steps 1 to 6 is shown in Supplementary Fig. 1.

ROI detection by spatial decomposition. To define ROIs within line scans, we first noted Fick's second law of diffusion describing the spatiotemporal diffusion of a substance as follows:

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} \tag{11}$$

where the change in concentration (C) can be described in one dimension (x) over time (t) with a diffusion coefficient (D). If N is the initial number of glutamate molecules released instantaneously at a point (corresponding to a synaptic release site), a solution can be in the form as follows:

$$c(x,t) = \frac{N}{\sqrt{2\pi\sigma^2(t)}} \times \exp\left[-\frac{x^2}{2\sigma^2(t)}\right]$$
 (12)

where

$$\sigma^2(t) = 2Dt \tag{13}$$

This function is recognizable as a normal distribution, with variance as a monotonically increasing function of time. We only sampled each location at intervals of 1 ms, which made an accurate analytical fit to this function unreliable. We therefore measured the spatial profile as a temporal average of the fluorescence signal along the line scan and fit this average to a sum of Gaussians, whereby each Gaussian component can be considered its own point source corresponding to an active zone. One of these fluorescence profiles is shown in Supplementary Fig. 2 defining two nearby active zones. Using a graphical user interface, the user specifies the peak (or peaks) in the profile by placing one or more cursors, and then the built-in curve fitting routines in IgorPro is used to fit the temporal average to the function k(x) as follows:

$$k(x) = \sum_{i=1}^{n} \frac{A_i}{\sqrt{(2\pi\sigma_i^2)}} \times \exp\left[\frac{-(x-\mu_i)^2}{2\sigma_i^2}\right]$$
 (14)

under the constraints

$$\mu_i \in [c_i - \delta, c_i + \delta]$$

$$A_i > 0$$

where c is the set of cursor locations, μ the set of component means, A the set of component amplitudes, δ is a small value allowing for errors in the cursor placement by the user, and n is the number of placed cursors. The potential problem of overfitting the intensity profile with multiple Gaussians was avoided by restricting the number of components to the number of placed cursors (Supplementary Fig. 2). This process could be repeated with different initial estimates of the locations of the peaks (that is, the number of cursors and their positions) until the error function reaches a threshold or until the fit was acceptable. User input at this stage allowed us to limit the analysis to active zones with distinct peaks and full width at half maximum (FWHM) values of <1.5 μm, thereby reducing the possibility of conflating signals from multiple active zones (see "Isolating iGluSnFR signals from individual active zones" below).

Once each spatial component had been defined, a time series for that component, F(t), was computed as the weighted average of the raw linescan matrix with the spatial filter estimated in step 1:

$$F(t) = \sum_{x} F(x, t)k(x)$$
(15)

where F(x,t) is the raw linescan matrix and k(x) is the Gaussian component for the ROI. Increasing the weight of pixels located toward the center of the spatial profile (Supplementary Fig. 2) allows for significant denoising.

Bleaching of iGluSnFR sometimes occurred during an observation episode and was usually corrected using a linear function of time F(t). The iGluSnFR signal used for all analyses was the relative change in fluorescence, $\Delta F/F_0$, calculated from the bleach-corrected signals. The most frequent value (that is, the baseline) of the trace was used as F_0 .

Identification of events by Wiener deconvolution. Release events within an active zone were identified by their characteristic kinetics using a Wiener filter. The time course of 101 iGluSnFR transients that were clearly separated in time from other events are shown in Supplementary Fig. 3 and could be described by a function of the following form:

$$h(t) = A \times \exp\left[-\frac{t}{\tau_f}\right] \times \left[1 - \exp\left[-\frac{t}{\tau_r}\right]\right]$$
 (16)

where A describes the amplitude of the event and τ_r and τ_f and are the time constants for rise and fall in the signal, respectively. We found that transients at most synapses could be accurately described using a kernel with parameters of $\tau_{\rm f} = 0.06$ s and $\tau_{\rm r} = 0.001$ s. These parameters were relatively invariant for transients of different amplitudes, indicating that the reporter operated linearly over the range of glutamate concentrations that we observed (Supplementary Fig. 3; see also Fig. 2). iGluSnFR signals therefore reflected a linear time-invariant system, fulfilling the assumptions required for the use of Wiener deconvolution. The result of the Wiener deconvolution was a time series in which glutamate release events were described approximately as Dirac- δ impulse functions of varying amplitudes, as shown by the traces in Supplementary Fig. 1d.

Extraction of events. Although the use of Wiener deconvolution significantly improved the SNR (Fig. 2), it was still necessary to set a threshold to distinguish events from noise. A second example analysis highlighting this key step is provided in Supplementary Figs. 4 and 5. Supplementary Figure 4a shows a kymograph of a line scan through a terminal in which there were two sources of glutamate (active zone 1 in red and active zone 2 in black), together with the activity time series for each obtained after spatial demixing. The corresponding traces after Wiener deconvolution are shown in Supplementary Fig. 4b, where it can again be seen that glutamatergic events varied widely in amplitude. The baseline in the deconvolved traces was not, however, noiseless, making it necessary to set a threshold for counting a deviation in this signal as an event. To choose this threshold, we first examined the distribution of values in the deconvolved trace. Across all experiments, these distributions were consistently Gaussians centered at or very close to zero, except for a small tail of positive values. We therefore used the standard deviation of a Gaussian fit to the distribution to set the threshold at which

positive values in the deconvolved trace were considered to be significant (that is, to reflect iGluSnFR events). The thresholds we used were 3–4 standard deviations above the baseline, as shown by the broken blue lines in Supplementary Fig. 4c. Events were then timed at the local maximum in the deconvolved trace above this threshold, as shown by the broken vertical lines in the expanded traces shown in Supplementary Fig. 5. The activity within an active zone could then be described by a vector **E** of event times and **A** of event amplitude

To provide some examples of how this procedure was performed, the parts of the records highlighted by the large box outlined in a broken green line in Supplementary Fig. 4a are shown expanded in Supplementary Fig. 5a. It can be seen, for instance, that the Wiener deconvolution allowed us to distinguish 4 events closely spaced around 43 s in active zone 1 and two overlapping events around 44 s in active zone 2. Positive deviations in the deconvolved trace that did not exceed the threshold were not counted as, for instance, around 46 s in active zone 1. Further examples of small, slow deviations of the baseline that were not counted as events are shown in Supplementary Fig. 5b, which is an expansion of the activity in active zone 2 over the period shown by the smaller green box in Supplementary Fig. 4a. Of the last three upward deviations in the raw iGluSnFR signal after 29 s, only the second caused a deviation in the deconvolved trace large enough to cross the threshold for significance. Again, the setting of a threshold after deconvolution rejected small, slowly rising bumps in the record that did not have the same shape as the Wiener kernel obtained by averaging large numbers of individual events (Supplementary Fig. 3). A histogram of event amplitudes extracted from this active zone is shown in Supplementary Fig. 5c. The black trace is a fitted sum of six Gaussians. The quantal amplitude, as defined by the average interpeak distance (0.21), is very similar to the amplitude of the first peak (0.23), indicating that the first peak in the distribution represents individually released vesicles.

Amplitude clustering and quantal time series. Based on the evidence that glutamate transients of varying amplitude were integer multiples of a unitary event or quantum (Fig. 2), we partitioned events into numbers of quanta using a Gaussian mixture model. Under this framework, the probability p of a value x is given as follows:

$$p(x) = \sum_{i=1}^{K} \phi_i \mathcal{N}(x \mid \mu_i, \sigma_i^2)$$
(17)

where N is the normal probability density function and Φ represents the mixing probability and sums to one. Several algorithms for clustering were tested, and we found that expectation–maximization provided a quick and efficient approach. For each synapse, the algorithm was run up to 15 times, each with a different number of components and with the mean and variance parameters randomly initialized. The algorithm was iterated until acceptable convergence had been reached (defined by a user-set threshold). Following completion of the 15 runs, the output clusters were plotted with a histogram of the extracted events to allow the user to select the optimal partition based on these outputs, as well as a plot of data likelihood for each run. This defines the vector \mathbf{AQ} of the estimated number of quanta for each event, an example of which is shown in Supplementary Fig. 1d and in Fig. 2 of the main text. Defining a time series as the number of quanta within each event allowed for computation of vesicle release rates and information theoretic measures.

Assessment of time resolution. The detection of events relied on finding local maxima in the deconvolved traces. Noise within our records created the potential problem of mislabeling two individual events as a single, higher quantal event, as shown by the simulation in the left panel of Supplementary Fig. 7. We therefore assessed the 'temporal discrimination window' describing the minimum time between events required before events can be reliably discriminated. First we measured the SNRs within our experimental data and then simulated a series of pulses of increasing inter-event intervals with matching SNR values. We then ran each of these simulations through the analysis sequence described above to estimate the temporal discrimination window.

The SNR within a recording was defined as the average amplitude of a uniquantal event divided by the standard deviation of the baseline noise. To compute the standard deviation of the noise signal, we plotted the distribution of fluorescence values of the raw $\Delta F/F$ trace, found the first peak and then fit all values to the left of this peak (thus removing any possible contamination by the signal) with a Gaussian, as shown in Supplementary Fig. 6. Note that this is not the same trace used to compute the threshold for event detection: that threshold was applied after deconvolution. In a sample of ten synapses, the SNR estimated in different traces ranged from three to eight, with the large majority greater than four. In 100 simulations using SNRs ranging from 3 to 8 (Supplementary Fig. 7), a SNR of 4 provided a temporal discrimination window of 10–15 ms (Supplementary Fig. 8).

A test for the possibility of glutamate spillover. Do glutamate transients recorded by iGluSnFR on the surface membrane of a bipolar cell terminal reflect glutamate release from that same cell or might they also reflect glutamate released from nearby cells? To investigate this question, we tested whether iGluSnFR transients

required an electrical signal to arrive from the soma. Supplementary Figure 9 shows responses from the terminals of two nearby bipolar cells. Destroying the soma of cell 1 using an infrared laser (800 nm) destroyed the transients in the connected terminal but not in the nearby terminal of cell 2 (Supplementary Fig. 9b). In other words, iGluSnFR molecules on terminal 1 did not generate any significant responses to glutamate released from neighboring terminals, but did require electrical drive arriving from the soma.

Testing the linearity of the iGluSnFR signal. Glutamate transients within the synaptic cleft can reach millimolar concentrations within the first hundreds of microseconds after a vesicle fuses⁵². If a substantial fraction of iGluSnFR molecules providing the signal bind glutamate when a single vesicle is released, it might be expected that the relationship between the iGluSnFR signal and the number of vesicles within an event will gradually saturate. The dissociation constant (K_d) of the iGluSnFR variant used in this study is ~4 µm²², but it is difficult to judge from such an equilibrium measurement what fraction of iGluSnFR molecules on a spatial scale of ~1 μm become occupied during an iGluSnFR transient (Fig. 1; Supplementary Fig. 2). We therefore took an experimental approach to assess whether iGluSnFR transients might begin to saturate in response to larger MVR events. Amplitude histograms from individual active zones were constructed using stimuli of both low contrast, when the distribution of event amplitudes is shifted toward smaller numbers of quanta, and high contrast, when there are more large events (Fig. 5). An example of such a histogram is shown in Fig. 2c, where eight distinct peaks are evident. We then measured the interpeak distances from a sum of Gaussians fit and tested whether the distance between peaks might be reduced for events containing larger numbers of quanta, as would be expected if there was significant saturation of the reporter. Collected results from six active zones are shown in Fig. 2d. The change in interpeak distance was not significantly different from zero up to events composed of nine quanta, indicating almost perfect linearity over this range. In comparison, the largest events we observed from a sample of 51 synapses were equivalent to 11 quanta (Fig. 5c). It therefore seems unlikely that our estimates of quantal number were skewed by saturation of the iGluSnFR reporter.

Isolating iGluSnFR signals from individual active zones. The point-spread function of the microscope had a FWHM_{xy} of $0.7 \,\mu m$ in the x-y, allowing active zones separated by 1 µm to be easily distinguished in the iGluSnFR signal along a single line scan (Fig. 1; Supplementary Fig. 2). The point-spread function in the z dimension was, however, significantly larger (FWHM_z = $2.2 \,\mu\text{m}$), raising the possibility that two or more active zones at different z depths might be considered as one if they coincided closely enough in the x dimension of the line scan. To assess the probability of conflating signals from different active zones, we measured the numbers and distribution of synaptic ribbons that hold vesicles close to the active zone^{9,53}. Ribbons within individual terminals were visualized in fixed samples using an antibody specific for ribeye A, as shown in Supplementary Fig. 10a,b. The ImageJ tool DiAna was used for object-based 3D colocalization and distance analyses⁵⁴. The distance between ribbons was measured between their centers of mass, but we did not count 'floating' ribbons, defined as those that were more than 0.5 mm from the surface membrane (Euclidean distance). Floating ribbons that are not attached to the surface are a common feature of bipolar cells⁵³, and accounted for 28% of 58 ribbons in a sample of 4 terminals. The density of ribbons attached to the surface membrane averaged $0.16\,\mu\text{m}^{-2}$, which was similar to a previous measurement of 0.12 ribbons µm⁻² made in regions of flattened membrane in bipolar cells from goldfish⁵³. The distance between nearest ribbons averaged $0.96 \pm 0.4 \,\mu\text{m}$, and the distribution of values is shown in Supplementary Fig. 11b.

To compute the probability of 'collapsing' the signal from two separate ribbons, we first calculated the lateral and axial resolutions of our microscope (0x) and 0x):

$$\omega_z = \frac{\text{FWHM}_z}{2\sqrt{\ln 2}} = 1.3 \,\mu\text{m} \tag{18}$$

$$\omega_{xy} = \frac{\text{FWHM}_{xy}}{2\sqrt{\ln 2}} = 0.42 \,\mu\text{m} \tag{19}$$

We then constructed a 'resolution' volume by creating an ellipsoid with a major axis equal to the axial resolution and minor axes equal to the lateral resolution, given by equation (20):

$$\frac{x^2}{\omega_{rv}^2} + \frac{y^2}{\omega_{rv}^2} + \frac{z^2}{\omega_r^2} = 1 \tag{20}$$

Thus, a ribbon at the origin of this ellipsoid would not be discriminated from any other ribbon lying within this volume. A section through this 'resolution volume is illustrated in Supplementary Fig. 11a, on which is superimposed the 'nearest neighbor volume' defined by a second ribbon at a distance m from the first, in

any direction. Assuming that all ribbons are randomly distributed relative to each other, the distribution of ribbons is uniform over the surface of the sphere defined by radius m. The probability of collapsing two ribbons at distance m can then be calculated as the surface area of the intersection of the two volumes divided by the surface area of the sphere. The resolution ellipsoid defined in equation (20) is a prolate spheroid, so this probability can be computed analytically or numerically using spherical caps. Supplementary Figure 11b shows the probabilities of collapsing two ribbons as a function of m. For the sample of 4 terminals and 42 non-floating ribbons, we computed an average probability of collapsing two nearest ribbons as 8%. These measurements indicate that the large majority of multiquantal events originate from single active zones.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

The code used to analyze the data in this study is available at https://github.com/lagnadoLab/glueSniffer.

References

- Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* 2, 13 (2003).
- Pola, G., Schultz, S. R., Petersen, R. S. & Panzeri, S. in Neuroscience Databases: A Practical Guide (ed. Kötter, R.) 139–154 (Springer, 2003).
- Budisantoso, T. et al. Evaluation of glutamate concentration transient in the synaptic cleft of the rat calyx of Held. J. Physiol. 591, 219–239 (2013).
- Zenisek, D., Horst, N. K., Merrifield, C., Sterling, P. & Matthews, G.
 Visualizing synaptic ribbons in the living cell. J. Neurosci. 4, 9752–9759 (2004).
- 54. Gilles, J. F., Dos Santos, M., Boudier, T., Bolte, S. & Heck, N. DiAna, an ImageJ tool for object-based 3D co-localization and distance analysis. *Methods* 115, 55–64 (2017).
- Zipfel, W. R., Williams, R. M. & Webb, W. W. Nonlinear magic: multiphoton microscopy in the biosciences. *Nat. Biotechnol.* 21, 1369 (2003).



Corresponding author(s):	Leon Lagnado
Last updated by author(s):	2/27/2019

Reporting Summary

Statistics

X Life sciences

Behavioural & social sciences

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see <u>Authors & Referees</u> and the <u>Editorial Policy Checklist</u>.

For all statistical analys	es, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.					
n/a Confirmed						
☐ ☐ The exact sam	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement					
A statement of	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly					
The statistical Only common to	The statistical test(s) used AND whether they are one- or two-sided Only common tests should be described solely by name; describe more complex techniques in the Methods section.					
A description	A description of all covariates tested					
A description	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons					
	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)					
For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted Give P values as exact values whenever suitable.						
For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings						
For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes						
Estimates of e	effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated					
	Our web collection on <u>statistics for biologists</u> contains articles on many of the points above.					
Software and o	code					
Policy information abo	ut <u>availability of computer code</u>					
Data collection	Multiphoton microscopes were controlled using Scanlmage v3.8 (Vidrio Technologies).					
Data analysis	Custom code used to analyze the data in this study is available at https://github.com/lagnadoLab/glueSniffer. Analysis was carried out using IgorPro (Wavemetrics) and Matlab.					
For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.						
Data						
- Accession codes, un - A list of figures that	ut <u>availability of data</u> include a <u>data availability statement</u> . This statement should provide the following information, where applicable: ique identifiers, or web links for publicly available datasets have associated raw data restrictions on data availability					
The data that support the findings of this study are available from the corresponding author upon reasonable request.						
Field-speci	fic reporting					
Please select the one b	elow that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.					

Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.				
Sample size	Sample sizes were chosen to be as large as possible and allow statistical tests where appropriate.			
Data exclusions	Data were only excluded from the analysis if the signal-to-noise ratio of the signals was not sufficient to detect unitary responses to visual stimuli.			
Replication	All experiments were repeated and replicated multiple times.			
Randomization	Not relevant to study as we were not testing for effects specific to a manipulation.			
Blinding	Not relevant to study as we were not testing for effects specific to a manipulation.			

Reporting for specific materials, systems and methods

N 4 = + | = = = | =

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

iviateriais & experimental systems		Methous		
n/a	/a Involved in the study		Involved in the study	
	Antibodies	\boxtimes	ChIP-seq	
\boxtimes	Eukaryotic cell lines	\boxtimes	Flow cytometry	
\boxtimes	Palaeontology	\boxtimes	MRI-based neuroimaging	
	Animals and other organisms			
\boxtimes	Human research participants			
\boxtimes	Clinical data			

Antibodies

Validation

Antibodies used

The primary antibody against zebrafish Ribeye A [C]-YNQGYLDRPDPRNIRK-[N] was an isolated chicken IgY fragment produced by Cambridge Research Biochemicals (clone 2575). The secondary antibody for ribeye A was Alexa-546 anti-chicken (Molecular Probes, Cat. A11040, Lot. 682609, dilution 1:1000) and for GFP we used Alexa-647 anti-rabbit (Molecular Probes, A21244, Lot.

A21244, dilution 1:1000).

The antibody to ribeye A has been validated in the laboratories of Leon Lagnado, Rachel Wong and Leanne Godinho.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

Zebrafish (Danio rerio), male and female, at 7-10 days post-fertilization.

Wild animals

This study did not involve wild animals.

Field-collected samples This study did not involve samples collected from the field.

Ethics oversight All animal procedures were performed in accordance with the Animal Act 1986 and the UK Home Office guidelines and with the

approval of the University of Sussex local ethical committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.