

## Supplementary information

### Quantification of predictability

The main method of this paper is quantitative estimation of the predictability of a spike train from a predictor variable, which may be any combination of spatial location, theta phase, or peer activity (Figure S1). We will first explain the method of predicting the spike train solely from position.

A 10-fold cross-validation procedure<sup>1</sup> is used to repeatedly divide the recorded data into a “training set” and a “test set”. The training set is used to construct a predicted intensity as a function of space  $f(x)$ , i.e. a place map.

The place map constructed from the training set is then evaluated on the test set. The position  $x(t)$  at each moment of the test set, and the place field  $f(x)$  constructed from the training set are used to produce an estimated intensity at each time,  $f(t) = f(x(t))$ . The log-likelihood density of the actual spike train  $\{t_s\}$  under this estimated distribution can be shown to be

$$L_f = - \int f(t) dt + \sum_s \log f(t_s) \quad (1)$$

The training set is also used to construct a mean firing rate independent of position,  $f_0$ , defined as the number of spikes during the training period divided by the length of the training period. The predictability on the test set is defined to be the difference,  $L_f - L_{f_0}$ . It therefore gives the log likelihood ratio of the data under the two predicted intensity models,  $f(x)$ , and the constant  $f_0$ . The predictability of the entire data set is defined by a cross-validation procedure, where the data is divided into 10 segments, each segment is in turn used as test set, and the log likelihood ratios for each segment are summed and divided by the total time (Figure S2).

This predictability measure also has an intuitive interpretation. Suppose an observer wants to communicate whether a spike occurred at a given time instant. She or he can communicate this most concisely using a code derived from her or his best estimate of the probability of spike occurrence. The number of bits needed, on average, to communicate using this code is the expected negative  $\log_2$  likelihood under this estimated distribution<sup>2</sup>. The predictability measure estimates the number of bits saved by an observer who is allowed to use the instantaneous position of the animal to construct a code, compared to one who can only use the mean firing rate.

The use of cross-validation guards against artificial overestimation of spike train predictability. For example, if firing probability is independent of position, an observer

using position to help encode the spike train would actually use more bits than one who used mean firing rate, because the random fluctuations in the training would result in a worse code used to encode the test set spike train. However, if cross-validation were not used, random fluctuations in the data set would lead to an illusory saving of bits needed to transmit the spike train.

### Construction of Place Fields

Prediction of intensity from location is equivalent to construction of a place field. Here we construct place fields by a smoothing-based method, where we divide a smoothed spike count map by a smoothed occupancy map. The estimated intensity at a point  $x$  is given by

$$f(x) dt = \frac{\sum_t n_t w(|x - x_t|)}{\sum_t w(|x - x_t|)} \quad (2)$$

Here,  $n_t$  is the number of spikes fired in a given time bin,  $x_t$  is the position of the rat in that time bin, and  $dt$  is the time bin size. The smoothing function  $w$  is a Gaussian with a variable width parameter:

$$w(d) = \exp(-d^2 / 2\lambda^2) \quad (3)$$

By comparing prediction quality with varying values of the smoothing width  $\lambda$ , we found that the optimal value for this parameter was  $\sim 5$ cm (see figure S3 below).

We note that the above method for place field computation is a form of *locally weighted maximum likelihood estimation*<sup>3</sup>, where the intensity  $f_x$  at a point  $x$  is chosen by maximizing a weighted sum of log-likelihoods of each time bin under a Poisson distribution, with weights given by  $w(|x - x_t|)$ :

$$L(f_x) = \sum_t w(|x - x_t|) [-f_x dt + n_t \log(f_x dt)] \quad (4)$$

### Construction of Phase Fields

The preferred phase of pyramidal cell spikes with respect to the theta rhythm varies with the animal's location in space<sup>4-7</sup>. We may therefore predict the spike trains more accurately by allowing the predicted intensity to vary as a function of theta phase. To do this, we must quantify the dependence of the cell's phase preference on spatial position. We do this again using locally weighted maximum likelihood estimation. We fit the

phases of all spikes to a von Mises distribution<sup>8</sup> whose parameters  $\theta_x$  (mean phase) and  $\kappa_x$  (modulation depth) vary with position  $x$ , to minimize the weighted likelihood

$$L(\theta_x, \kappa_x) = \sum_s w(|x - x_s|) \frac{\exp(\kappa_x \cos(\theta_x - \theta_s))}{2\pi I_0(\kappa_x)} \quad (5)$$

Here, the sum is over all spikes  $s$ , and  $x_s$  and  $\theta_s$  are the position and instantaneous theta phase at the time of spike  $s$ . As with place fields, the local maximum likelihood estimate may be efficiently computed by a smoothing method, according to the following formulas:

$$\theta_x = \arg \left( \frac{\sum_s e^{i\theta_s} w(|x - x_s|)}{\sum_s w(|x - x_s|)} \right) \quad (6)$$

$$\kappa_x = A_1^{-1} \left( \frac{\sum_s e^{i\theta_s} w(|x - x_s|)}{\sum_s w(|x - x_s|) + 1} \right) \quad (7)$$

where  $A_1$  is the ratio of Bessel functions  $A_1(y) = I_1(y)/I_0(y)$ <sup>8</sup>. In order to regularize against over-fitting in areas where few spikes were fired, a constant term of 1 was added to the denominator in (7).

The predicted intensity from position and phase is the product of the place field term and a phase modulation term:

$$f_t = f(x_t) \frac{\exp(\kappa_x \cos(\theta_t - \theta_x))}{I_0(\kappa_x)} \quad (8)$$

### Prediction of unit activity from population

To predict the activity of one cell from the population of peer cells, we used a generalized linear model<sup>9</sup>. Initially, the spike trains of the peer cells are smeared in the time domain with a Gaussian function of variable width  $\sigma$  (the *peer prediction timescale*):

$$s_{t\alpha} = \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{\tau_\alpha} \exp\left(-\frac{(t - \tau_\alpha)^2}{2\sigma^2}\right) \quad (9)$$

Here the sum runs over all spikes of cell  $\alpha$ . Under the generalized linear model, the predicted intensity at a time  $t$  is given by

$$f_t = g\left(\sum_{\alpha} s_{t\alpha} w_{\alpha}\right) \quad (10)$$

The link function  $g(\eta)$  had the following form:

$$g(\eta) = \begin{cases} \exp(\eta) & \eta < 0 \\ \eta + 1 & \eta \geq 0 \end{cases} \quad (11)$$

A simple exponential was not used, because this led to excessively high predicted intensities in the case when many positively predicting peer cells were firing simultaneously.

The prediction weights  $w_{\alpha}$  were chosen to maximize the penalized log-likelihood on the training set

$$L = \sum_t [-f_t dt + n_t \log(f_t dt)] - \frac{1}{4} \sum_{\alpha} w_{\alpha}^2 \quad (12)$$

The maximization is carried out by Newton's method with an analytically calculated Hessian matrix. The penalty term helped to prevent over-fitting by reducing large weight values that did not substantially improve prediction quality on the training set.

### **Prediction of unit activity from position and population**

When spike trains were to be predicted from spatial variables, in combination with peer prediction, the peer prediction function was multiplied with the spatial prediction function:

$$f_t = f(x_t) \frac{\exp(\kappa_x \cos(\theta_t - \theta_x))}{I_0(\kappa_x)} g\left(\sum_{\alpha} s_{t\alpha} w_{\alpha}\right) \quad (13)$$

This formula was used to examine whether the prediction of the spike train, made from spatial and phase variables, may be further improved by taking into account the activity of peer cells. The weights in this case are not necessarily the same as those when activity is predicted from peer cells alone, and are recomputed using Newton's method as those that maximize the penalized likelihood of the product of spatial and peer prediction intensities on the training set.

### **Use of prediction method to estimate spatial scale of place field**

Figure S3. Computation of place fields involves the use of a spatial smoothing scale. The cross-validation method may be used to ensure the use of an optimal smoothing scale. a) Place fields of the same cell, computed for three different values of smoothing scale. At 3cm, the place field shows a high degree of spatial structure; however, this structure arises from random fluctuations, rather than reliable place preferences of the neuron (under-smoothing). At 10cm, the place field is nearly circular; however, smoothing at this scale loses does not capture the full spatial structure of the place field (over-smoothing). b) Optimal smoothing scale was estimated by predicting the cell's activity from space alone, for a range of smoothing scales. Peak predictability was at 5.6cm for this cell. If information rate was computed by a direct method<sup>10</sup>, without cross-validation, the apparent information content of the cell increased without bound as the spatial scale was lowered, indicating that cross-validation is necessary to protect against under-smoothing. c) Across the population, the median optimal smoothing scale was found to be 5cm.

### Relation of predictability to spike train characteristics

Figure S4a. Dependence of “supra-spatial” peer predictability (Peer Gain) on isolation quality of target cell. Unit isolation quality was assessed using the “isolation distance” measure<sup>11</sup>. When all initially clustered cells are considered, including those below the isolation distance threshold (shown in black), a significant correlation is found between predictability and isolation quality ( $p < 0.001$ , red line), indicating that poorly isolated cells are less predictable from peers. However, if only those cells that passed the isolation distance threshold of 20 are considered, no correlation is seen ( $p > 0.5$ , green line). We therefore only considered these cells for further analysis.

Figure S4b and c. Supra-spatial peer predictability, measured in bits/sec, is positively correlated with target cell firing rate (Fig S4b;  $p < 0.001$ ). However, if the predictability measure is normalized by the firing rate of the target cell to give a measurement of bits per spike, a negative correlation is observed (Fig S4c;  $p < 0.001$ ). This suggests that faster firing cells are more predictable simply because there are more spikes whose occurrence can be predicted. Furthermore, the negative correlation of target cell firing rate with bits per spike suggests a rule of “diminishing returns” for high-firing cells.

Figure S4d. Peer predictability increases with the number of peer predictor cells (red line; slope 0.12 bits/spike/cell;  $p < 0.001$ ). To ensure the effect was not unduly influenced by one animal with the largest number of cells (42 of the 189 passing the isolation criteria), the analysis was repeated with this animal excluded (green line). Fits were constrained to pass through the origin (zero cells provide zero predictability).

Figure S4e. Peer predictability appears to be negatively correlated with data set size (i.e. length of recording)(red line;  $p < 0.01$ ). However, this correlation is entirely due to a single animal with a large number of cells and a short recording time. If this recording is excluded, the correlation disappears (green line).

Figure S4f. Peer predictability is correlated with spike train variability, as measured by the Fano factor<sup>12</sup> ( $p=0.003$ ). The Fano factor can be defined using variable window sizes; a significant positive correlation of variability with predictability was seen for all windows greater than or equal to 25ms (the Fano factor displayed here is calculated with a window of 50ms). This observation suggests that neurons exhibiting greater spike train variability also show a higher gain in predictability using peer activity over location alone.

The timescale at which neuronal spikes were best predictable from population activity showed a clear mode at approximately 25ms (Fig 3b, main paper). Nevertheless, some cells showed optimal prediction timescale different to this value. To clarify which neuronal characteristics correlate with non-standard predictability timescale, we divided cells into two categories, those whose predictability timescale was close to the mode (in the range 10-40ms), and all other cells. We performed a multiple logistic regression analysis<sup>9</sup> to predict the category a cell belonged to using the following array of predictor variables: Isolation quality, number of predictor cells, total recording length, spike train variability (Fano factor), and predictability from space. A significant effect was found for isolation quality and number of predictor cells, both of which correlated positively with the probability that the cell would show optimal predictability in the range 10-40ms ( $p=0.01$  and  $0.0008$ , respectively). None of the other predictor variables showed a significant effect. We therefore concluded that the scatter in predictability timescales was related to properties of the extracellular recording, rather than properties of the cell itself, or its relation to network activity.

### **Relation of prediction weight to location of place fields**

Figure S5a) The phase of firing of hippocampal pyramidal cells is known to depend on the animal's location in space<sup>4</sup>, with mean phase close to the negative peak of pyramidal layer theta cycle in the center of a place field, and close to the positive peak in the periphery. One would therefore expect correlations between cells to depend on the spatial overlap of their place fields<sup>5</sup>. b) The prediction weight indicates the degree to which correlations differ from those expected if spike timing was solely determined by spatial location. Prediction weight is shown as a function of the degree of place field overlap (computed as the scalar product of the normalized place field maps), for each pair of cells. While prediction weights are highly variable, there is a weak but significant correlation between prediction weight and the degree of place field overlap ( $r=0.17$ ;  $p<0.0001$ ), indicating decreased or increased synchronization of cells beyond that predicted from simultaneous independent theta phase precession. The large scatter about the fit line suggests that non-spatial factors also play a role in determining correlation strength.

### **Relation of prediction weight to anatomical location within the CA1 region**

Figure S6. a) Example relation between anatomical location and prediction weight. Vertical cell location in the pyramidal layer was estimated from the mean spike waveform for each cell recorded by the 8-site silicon electrode shank, whereas lateral

position was determined from the intershank distance (200  $\mu\text{m}$ ). Predictor cells (triangles) are color-coded by prediction weight (red positive, blue negative) to a representative target cell (star). No cells on the target cell shank were used as predictor cells to avoid spurious synchrony caused by isolation errors<sup>13</sup>. No consistent anatomical distribution of positively or negatively weighted cells was seen. b. Across the population, no relation was seen between the anatomical spacing of cells (shank separation), and prediction weight (linear regression, red line,  $p=0.47$ ). However, because we avoided prediction of neuronal activity from extremely anatomically proximal neurons<sup>13</sup> (<100  $\mu\text{m}$ ), we cannot ascertain whether these extremely proximal neurons would show a reliable difference in prediction weight.

#### Reference List

1. Ripley, B.D. Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge (1996).
2. Cover, T.M. & Thomas, J.A. Elements of information theory. New York, N.Y. ; Chichester : John Wiley, (1991).
3. Loader, C. Local regression and likelihood. Springer-Verlag, New York (1999).
4. O'Keefe, J. & Recce, M.L. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317-30 (1993).
5. Skaggs, W.E., McNaughton, B.L., Wilson, M.A. & Barnes, C.A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149-72 (1996).
6. Harris, K.D. *et al.* Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells. *Nature* **417**, 738-741 (2002).
7. Mehta, M.R., Lee, A.K. & Wilson, M.A. Role of experience and oscillations in transforming a rate code into a temporal code. *Nature* **417**, 741-746 (2002).
8. Fisher, N.I. Statistical analysis of circular data. Cambridge University Press, New York (1993).
9. Dobson, A.J. An Introduction to Generalized Linear Models. Chapman and Hall, London (1990).
10. Skaggs, W.E., McNaughton, B.L., Gothard, K.M. & Markus, E. Advances in Neural Information Processing Systems, Vol. 5. Hanson, S., Cowan, J. & Giles, G. (eds.), pp. 1030-1037 (Morgan Kaufmann, San Mateo, 1993).

11. Harris,K.D., Hirase,H., Leinekugel,X., Henze,D.A. & Buzsaki,G. Temporal interaction between single spikes and complex spike bursts in hippocampal pyramidal cells. *Neuron* **32**, 141-149 (2001).
12. Baddeley,R. *et al.* Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. R. Soc. Lond B Biol. Sci.* **264**, 1775-1783 (1997).
13. Quirk,M.C. & Wilson,M.A. Interaction between spike waveform classification and temporal sequence detection. *J. Neurosci. Methods* **94**, 41-52 (1999).

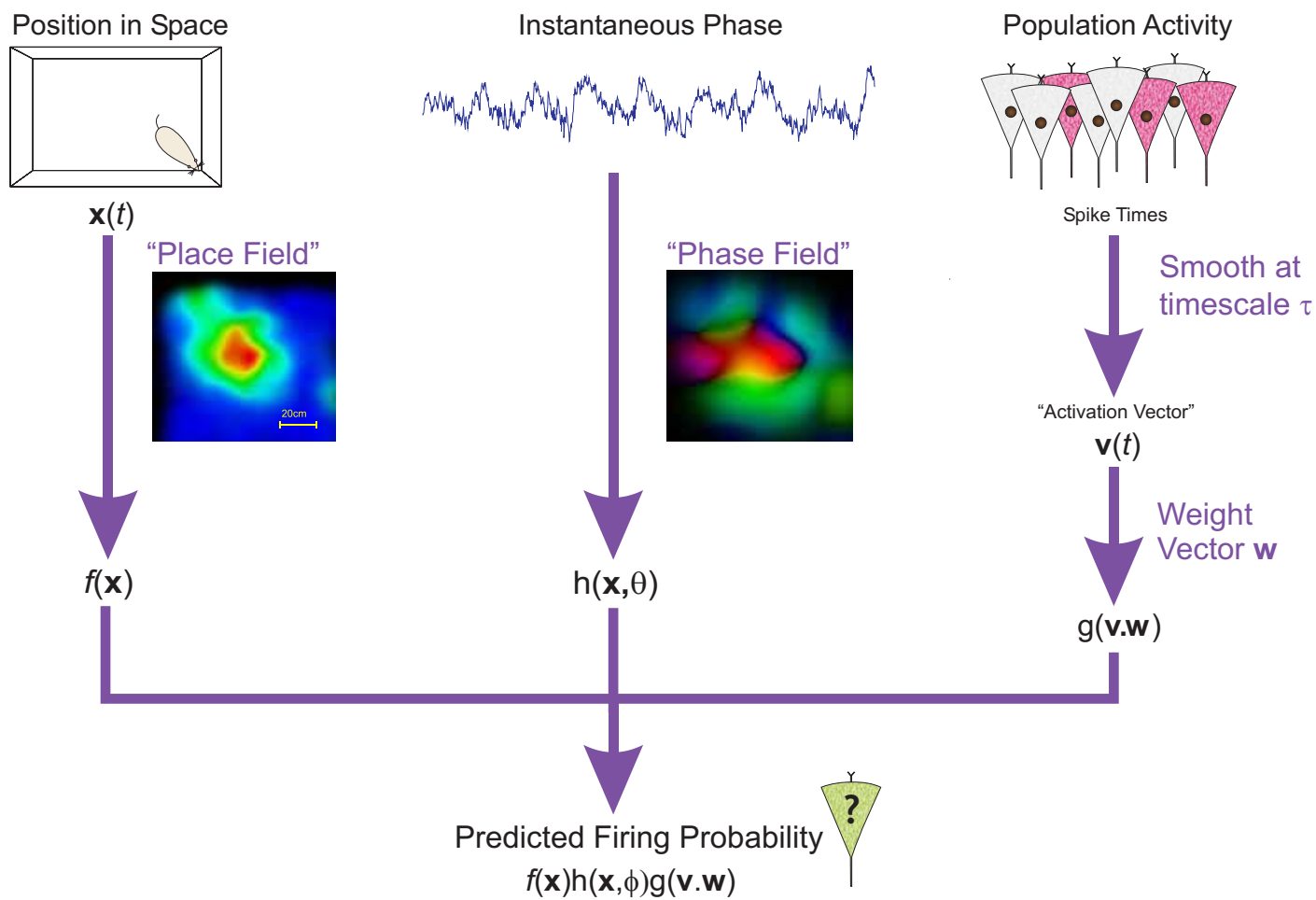


Figure S1

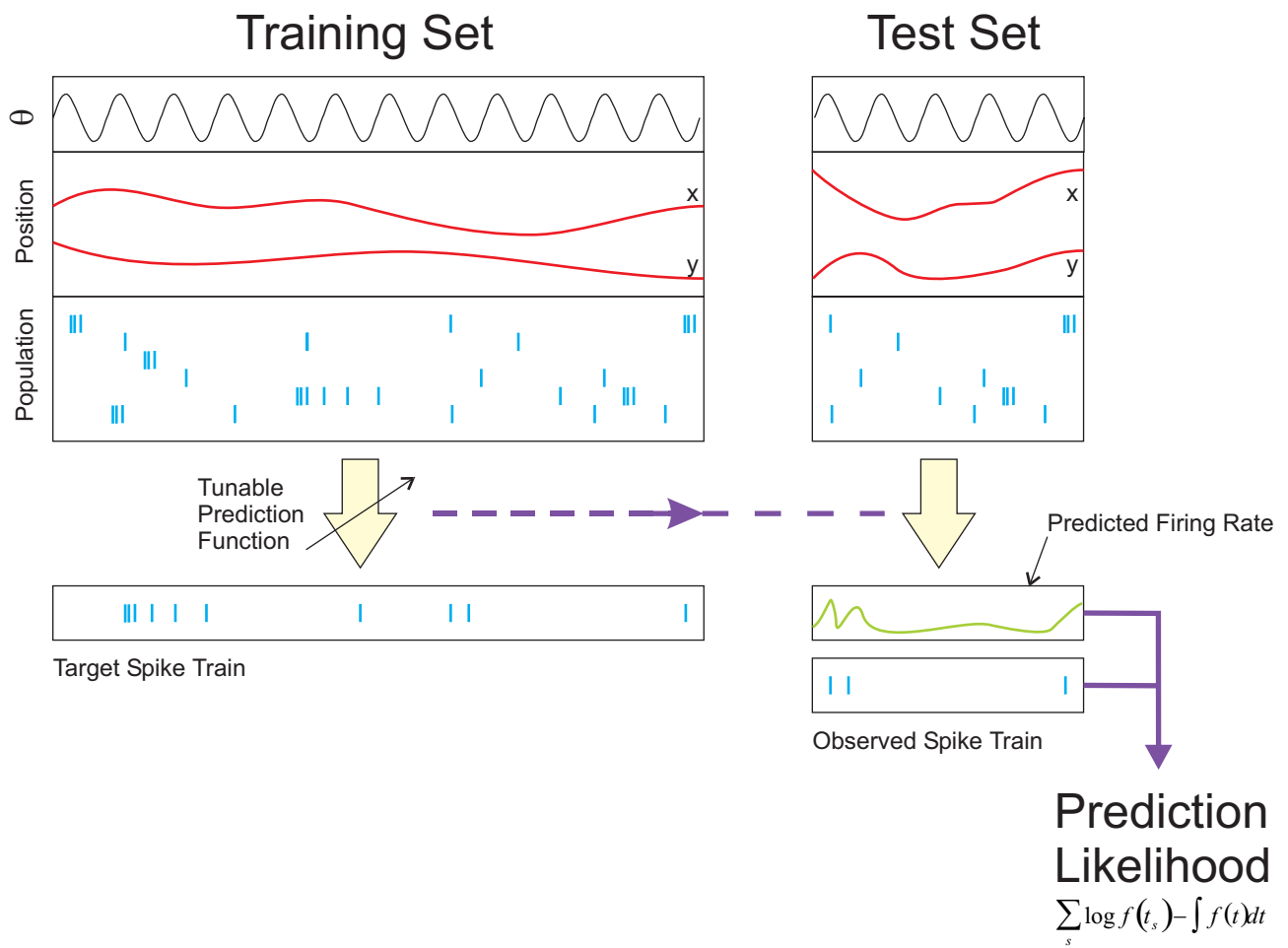


Figure S2

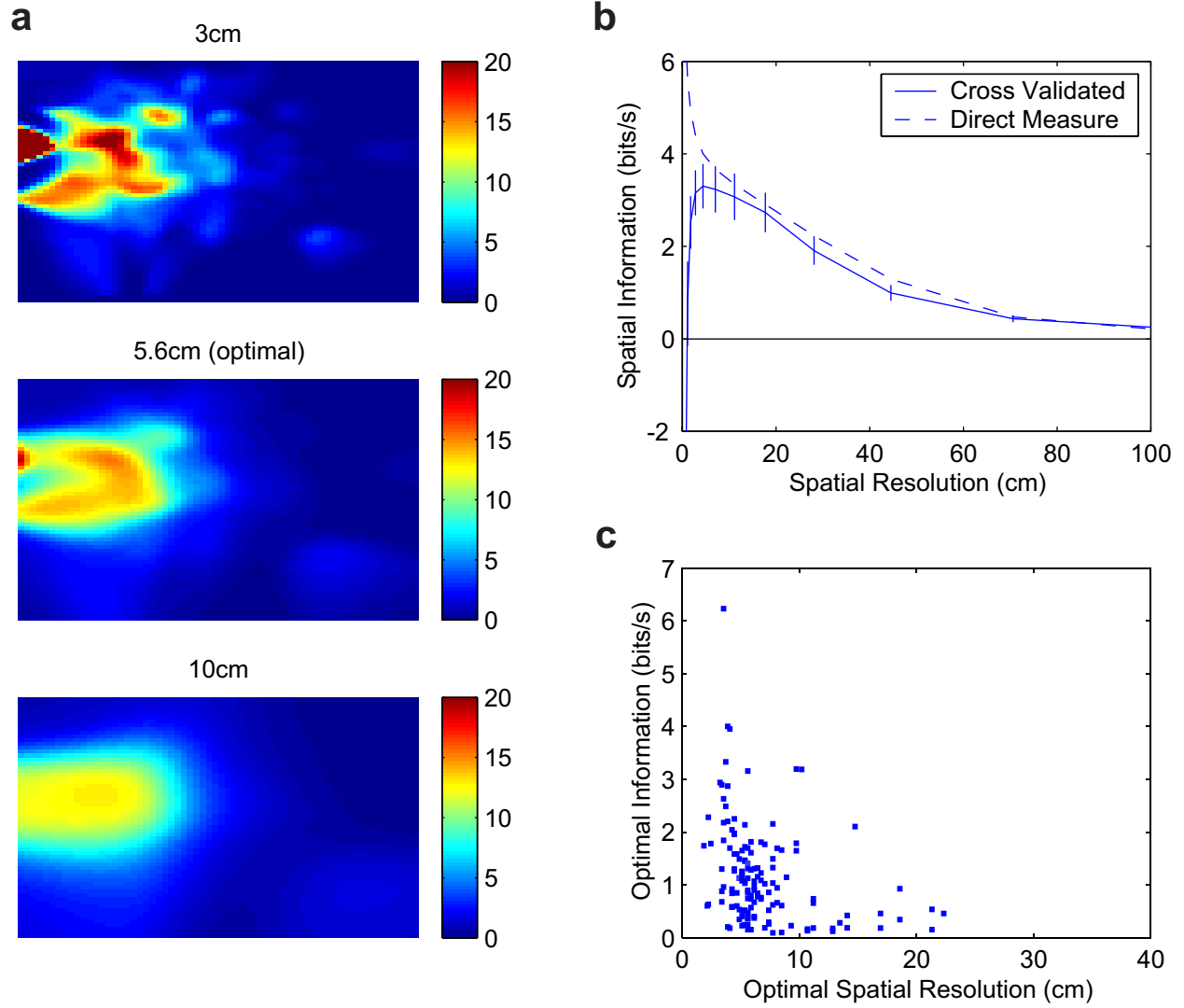


Figure S3

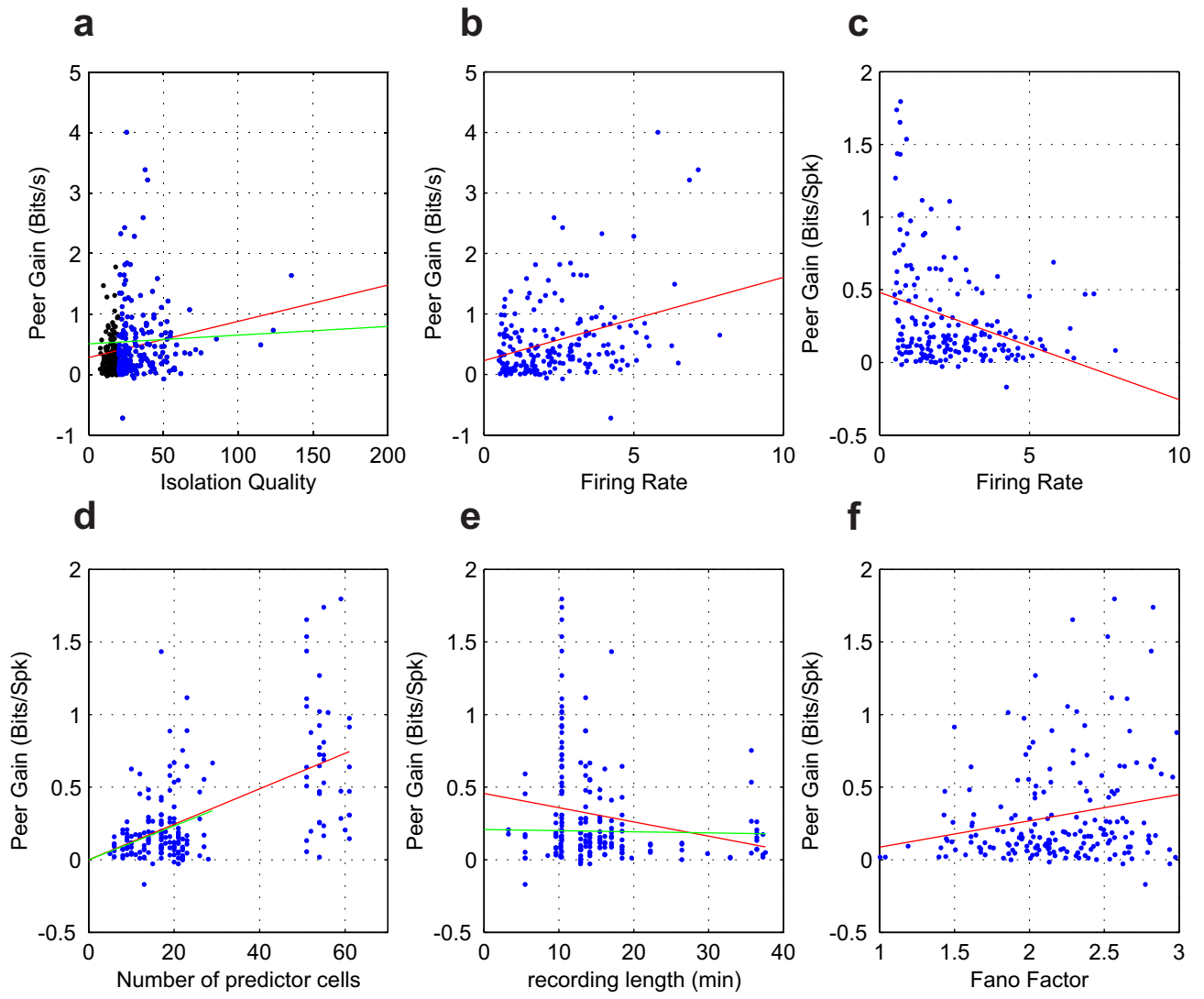


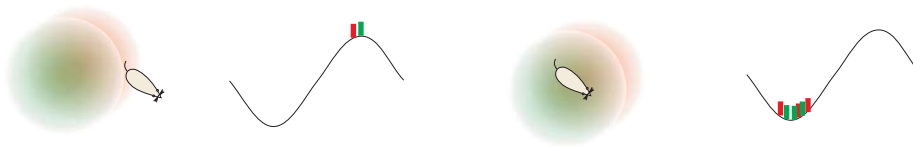
Figure S4

**a**

Non-overlapping place fields => Anticorrelated firing



Overlapping place fields => Correlated firing



**b**

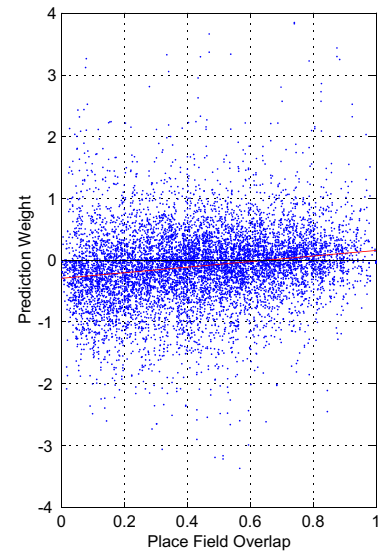


Figure S5

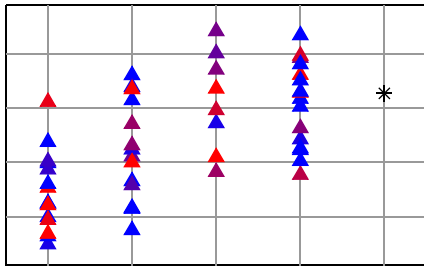
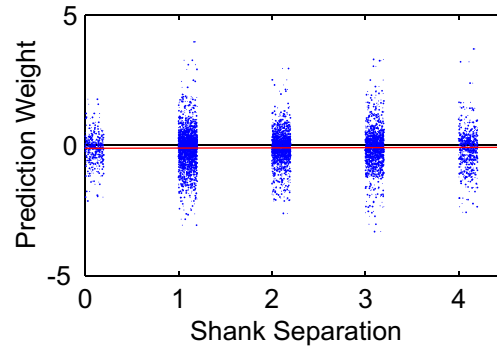
**a****b**

Figure S6