Structured internal cortical states deduced from fixed tactile input patterns

Authors: Johanna Norrlid[1†], Jonas M.D. Enander[1†], Hannes Mogensen[1] and Henrik Jörntell[1]*

Affiliations:

[1] Neural Basis of Sensorimotor Control, Department of Experimental Medical Science, Lund University, Lund, Sweden.

*Correspondence to: Henrik Jörntell, henrik.jorntell@med.lu.se

† These authors share first authorship based on equal contributions.

Abstract:

The brain has a never-ending internal activity, whose spatiotemporal evolution impacts how we perceive external inputs and generate illusions. The spatiotemporal evolution depends on the neuronal network structure, which forms such a rich dynamic system that the internal interaction with external inputs has remained poorly understood. We used reproducible touch-related spatiotemporal inputs and recorded intracellularly from rat neocortical neurons to characterize this interaction at the circuitry level. Although repeated presentations of the same input generated variable responses, they tended to sort into a set of preferred response states, unique for each neuron. This suggests that sensory inputs combine with internal brain network dynamics to cause it to fall into one out of many possible local minima solutions with disparate instantiations in the subnetworks connected to each neuron.

One Sentence Summary: Responses to identical sensory inputs provide detailed indications on the structure of internal cortical network processing.

Short title: Cortical state structure to fixed input

Behavioral, mental and perceptual functions of the neocortex depend on internal state control. A state in the brain can be described as the combination of activity in all of its neurons (*1*). Since the neurons make synaptic connections with each other, their activity are not independent, which is reflected in reports of constrained 'realms' of possible response combinations in populations of neurons (*2*, *3*). External inputs to the neocortical circuitry, which generate spatiotemporal patterns of activation arising in the multitude of sensors throughout our bodies, further constrain the space of possible neuronal responses (*2*). An important aspect of perception, and the foundation of illusions, is that a response is not only determined by the quality, or spatiotemporal pattern, of the sensory input but also depends on the current internal state of the cortex (*4*-*7*). Although by definition a high-dimensional latent state (*8*), a more specific embodiment or physiology of the circuitries generating this type of constraint has so far been difficult to identify. This is not surprising as a direct demonstration requires a precise estimate of what the experimental subject is thinking of, and how it is instantiated in the circuitry, at the time of the stimulus delivery. Otherwise, the internally generated constraints become an uncontrolled variable, which will appear as internal system noise in sensory-evoked responses. Here we aimed to deduce information about the character of these internally generated response constraints using fixed spatiotemporal patterns of tactile sensory activation.

To overcome an underestimated methodological problem of neurophysiology, that is to achieve exactly reproducible sensor activation patterns in a living organism where the sensors are located in compliant or movable tissue and their exact location or tuning relative to external stimuli are subject to uncontrollable brain efferent control, we used electrical intracutaneous stimulation to deliver a set of reproducible but also richly resolvable sensory input patterns to the brain (*9*) (Fig. 1A). In order to further minimize system noise caused by uncontrollable movements and internal thought processes unrelated to the stimuli, and in order to make the rats accept long-term stimulation of the skin electrodes, we used light anesthesia. We made intracellular, whole-cell patch clamp *in vivo* recordings from single neocortical pyramidal cells (putative layer III-V pyramidal neurons in the somatosensory cortex (S1), Fig. 1B), three of which were morphologically verified to be layer III pyramids (Fig. 1C). The intracellularly recorded signal represents the summed synaptic input from 10,000's of neurons. As these neurons by definition are part of the same subnetwork within the cortex (which in the rat contains 25,000,000 neurons (*10*)), the intracellular signal is a read-out of the instantiation of the current brain state that is specific to the subnetworks connecting to that neuron.

The time-varying states of the subnetworks connected to the recorded neuron generated a rich a background activity of spontaneous up and down states, mixed with episodes of intermediate states, against which the responses evoked by the sensor input patterns could often stand out as distinctly different (Fig. 1D). Responses evoked by the same stimulation pattern appeared to be impacted by the preceding state, as reflected in the spontaneous activity (Fig. 1D, top two traces). In some cases, the spontaneous activity could even resemble the responses evoked by the stimulation (Fig. 1D, bottom trace), in agreement with another study (*5*). A striking feature of the responses to isolated single pulse stimulations were their highly variable nature (Fig. 1E-G). Since electrocutaneous primary afferent sensory activation has a constant response latency time and absence of failures (*11*), the variability (Table S1 for statistics and Fig. S1 for further details) must be due to factors in the networks of the central nervous system.

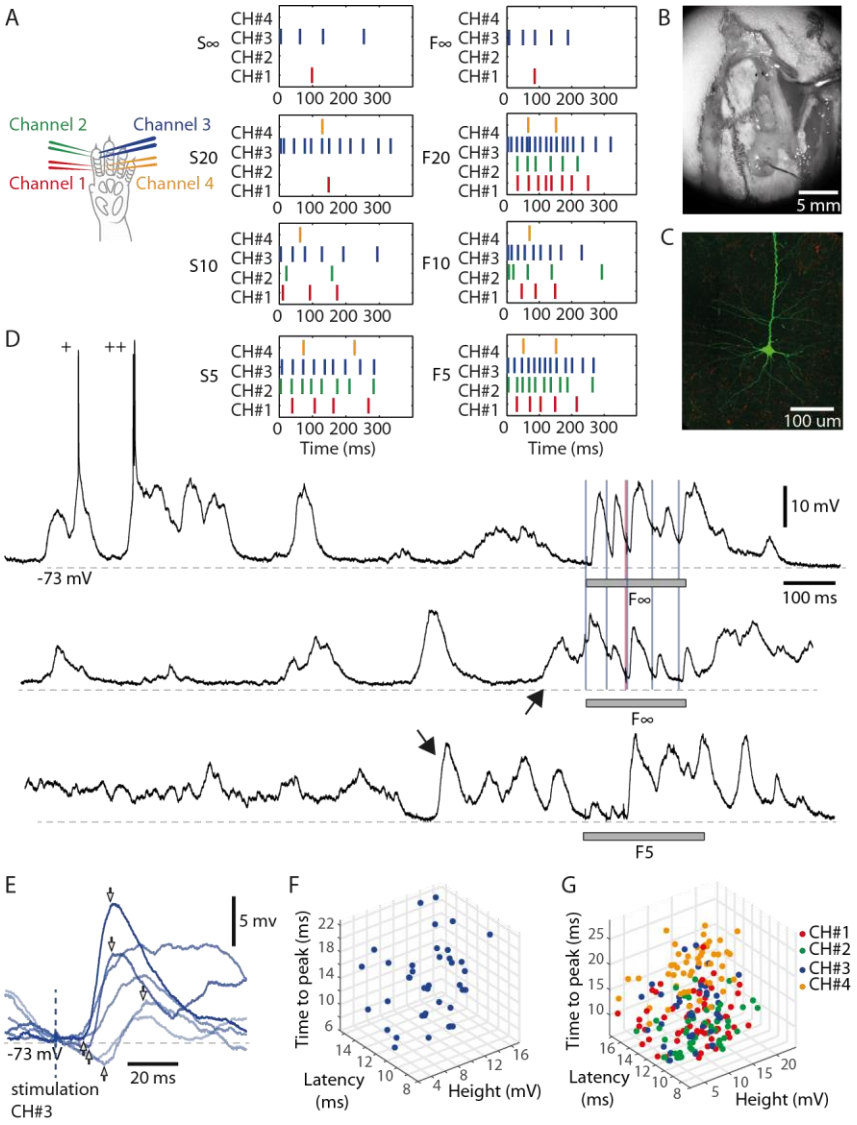

Fig. 1. Electrical skin stimulation and general properties of intracellular responses. (**A**) Locations of the four pairs of intracutaneous needle electrodes ('channels') inserted in digit 2 to stimulate the tactile primary afferents. The eight spatiotemporal patterns were re-used from a previous publication (*9*). (**B**) Neuronal recordings were made in an exposed cortical area of 4 by 2 mm, located in the center of the photo. An electrocorticography (ECoG) surface electrode was placed on a separate, smaller exposed area in the lower half of the photo. (**C**) A stained layer III pyramidal neuron from one of the recordings. (**D**) Intracellular traces illustrating the spontaneous activity and responses evoked by the stimulation patterns (grey horizontal bars). '+' indicates occasional spikes. Arrows indicate specific spontaneous activity patterns. Colored lines indicate the times of individual stimulation pulses in pattern F(inf). (**E**) Six superimposed raw intracellular traces evoked by isolated single pulse stimulations to channel #3. Pairs of arrows indicate the response onset latency time, the time-to-peak and the peak amplitude for three sample responses. (**F**) 3D plot of the measured parameters for all responses evoked by isolated single pulse stimulation of channel #3 in the same neuron. (**G**) Similar display for all four channels used.

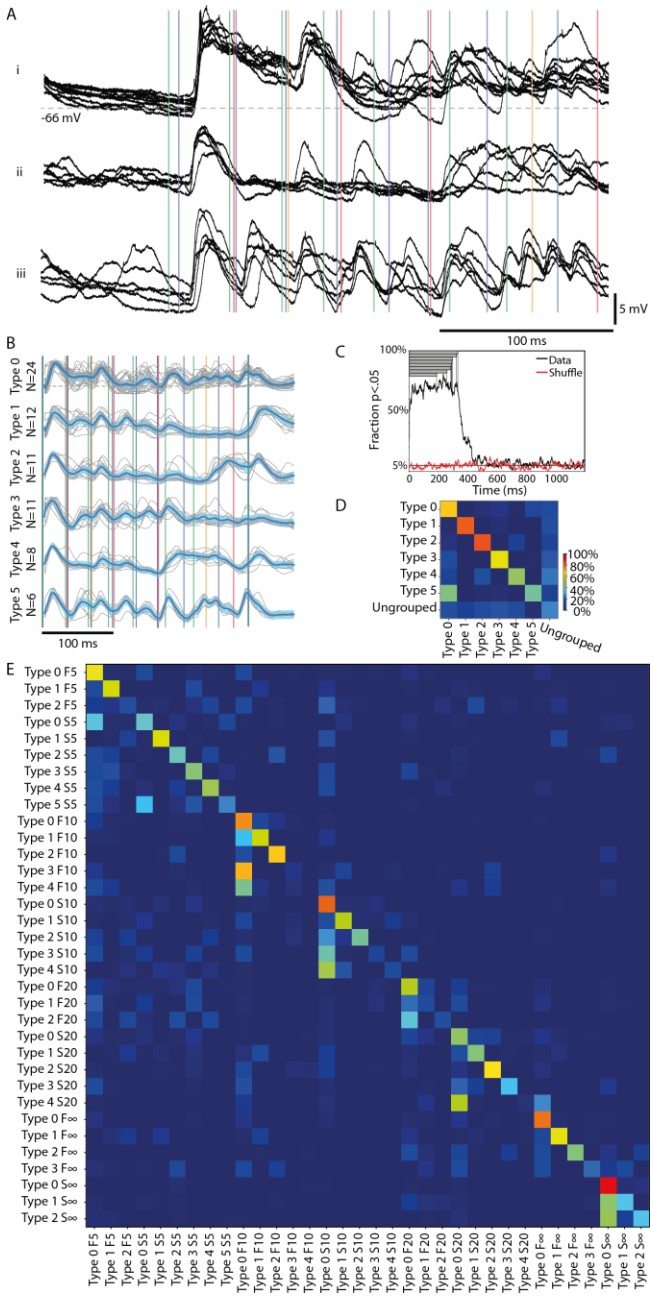

Fig. 2. Different response types evoked by the same stimulation pattern. (**A**) Each panel i-iii shows, for a sample cell, examples of qualitatively different superimposed raw data traces, evoked by the same stimulation pattern (S5). The stimulation pattern (vertical colored lines) outlasted the traces. (**B**) The six identified response types for all responses evoked by the 100 repetitions of the same stimulation pattern in this cell are indicated by their mean time-voltage curve (dark blue), Threshold difference (light blue; see Fig. S2), raw trace members (grey) alongside the stimulation pattern. (**C**) Time evolution of the specificity of the responses of each response type. The black curve is the fraction of all response types where specificity could be detected (at $p<0.05$, Kruskal-Wallis test; H-statistics in Fig. S3) across each sample time point. The red curve shows the corresponding fraction for responses with shuffled response type labels (within each stimulation pattern, for each cell). Grey bars show the durations of the eight

stimulation patterns. (**D**) Confusion matrix for the members of each identified response type for the sample stimulation pattern and neuron illustrated in **A-B**. (**E**) Confusion matrix for all response types identified for all stimulation patterns for this sample neuron.

Naturalistic tactile inputs, as opposed to single stimulation pulses, are composed of spatiotemporal patterns of skin sensor activation, which might allow the input to impact the internal state of the cortical networks as the input pattern still unfolds. Indeed, repeated delivery of a specific stimulation pattern generated responses that sorted into a few different categories (Fig. 2A). A response classification method indicated that the majority of the responses in the illustrated case were divisible into six response types (Fig. 2B). This sample neuron displayed 3-6 types for each of the other seven stimulation patterns tested (Fig. 1A). Across all neurons, the responses evoked by each stimulation pattern were on average divisible into 3.8+/-1.3 response types (Table 1). Using Kruskal-Wallis, we tested whether the responses classified into one response type were separable (at $p<0.05$) from all other responses evoked by the same stimulation pattern in the same cell. Across the grand total of 494 response types all stimulation patterns and all neurons, this specificity occurred for more than 60% of the response types across each time point for the duration of the stimulation patterns, whereas it rapidly dropped to chance levels after the stimulation patterns ceased (Fig. 2C, Fig. S3).

We next used Principal Component Analysis (PCA) as verification that the response types were distinctly different from each other. This method provided a measure of the accuracy, or the distinctness of separation of the individual responses, as summarized in a confusion matrix for the sample stimulation pattern in this neuron (Fig. 2D). The mean accuracy across the different response types is indicated in the diagonal, and was on average 59.7% for this stimulation pattern (including ungrouped responses; chance level was = 1/7 = 14.3%). Across the population of recorded cells, the accuracy of the separation of the responses into the identified responses types was generally above 60% for each of the eight stimulation patterns (Table 1).

Whereas the majority of the evoked responses could be classified as belonging to one of the response types (Table 1), 'ungrouped' responses (Fig. 2D) were by definition a much broader class and consequently had a much greater risk of confusion (i.e. a low value in the diagonal and a higher prevalence of well-above-zero values outside the diagonal, Fig. 2D). To evaluate the 'ungrouped' responses we used the F1 score, where a high value indicates a small risk of confusion with any of the specific classes of responses. In Fig. 2D, the F1 score was 0.24, indicating a relatively high degree of confusion. But across the dataset, the F1 score was 0.48 (Table 1), which indicates that the ungrouped responses were quite well separated from the defined response types, despite being a heterogeneous group.

We next used PCA to look at the response types evoked by all the eight stimulation patterns in the illustrated cell (Fig. 2E). In this case, the accuracy of the response types (ungrouped response were excluded here) was on average 57.6% (chance level = 1/34 = 2.9%), whereas across all cells recorded, this accuracy was 38.2+/-21.0%. Hence, a majority of the evoked responses for a single stimulation pattern could be divided into distinct types, which with a high accuracy were separable from responses not belonging to that type. Within each response type, though, a certain degree of variability occurred (Fig. 2B) suggesting that the number of resolvable response types is likely to be higher than what we could identify from 100 repetitions. But this analysis demonstrated a discontinuity of responses generated by identical inputs.

Table 1. Performance measures for the response type separation using PCA and kNN

MEAN (STD)	Grand Average	F_5	S_5	F_{10}	S_{10}	F_{20}	S_{20}	F_{INF}	S_{inf}
Identified response types (N)	3.8 (1.3)	3.5 (1.0)	3.9 (1.4)	4.0 (1.4)	4.0 (1.6)	4.1 (1.5)	3.5 (1.8)	3.7 (2.2)	3.6 (1.8)
Class separation accuracy (%)	60.9 (8.1)	59.5 (7.5)	62.6 (14.2)	61.8 (9.0)	61.5 (13.2)	56.7 (14.2)	61.9 (13.0)	60.7 (10.3)	62.7 (12.2)
Fraction 'Ungrouped' (%)	37.5 (14.6)	39.5 (12.9)	34.8 (12.8)	33.1 (18.2)	34.1 (17.6)	43.4 (15.9)	38.2 (18.0)	38.6 (15.3)	37.9 (20.4)
F1-score, 'Ungrouped' (0-1)	.48 (.18)	.51 (.21)	.48 (.19)	.45 (.21)	.45 (.21)	.54 (.18)	.50 (.21)	.48 (.22)	.47 (.26)

We previously found the responses evoked by a given stimulation pattern to be overall different between cells (*9*), which on average was the case also in the present set of recordings (Fig. 3A-C). Here, the issue was instead if the identified response types evoked by the same stimulation pattern across cells were distinct from each other. Fig. 3D illustrates a confusion matrix of a sample stimulation pattern, for which the accuracy was 42.7% (chance = 1/48 =2.1%). Across all eight stimulation patterns, the accuracy was 39.5+/-3.5%. Moreover, in the sample illustration (Fig. 3D), we found that 45 out of 48, or 93.8%, of the response types were separable from the responses of all other types (i.e. decoding accuracy higher than chance). Across all stimulation patterns 44.1+/-2.9 response types, or 89.6+/-3.0% of the total number of response types identified, were similarly separable. A separate analysis of the responses evoked by the individual pulses that composed each stimulation pattern further indicated the responses for each neuron to be statistically different from the responses evoked in other cells (Figs S4-S6).

We next compared the distribution of the relative probability of the desynchronized ECoG state for each stimulation patterns (a total of 8 times 13 = 104 data points across all cells) with the corresponding distribution for each of the 494 response types identified. The distributions did not differ significantly (p=0.68, t-test; Fig. S7). Hence, this aspect of the initial brain state was not predictive of the response type, suggesting that the response types were not an effect merely of the initial brain state but rather the result of an interplay between the input and the dynamics of internal state evolution as the stimulus presentation unfolded.

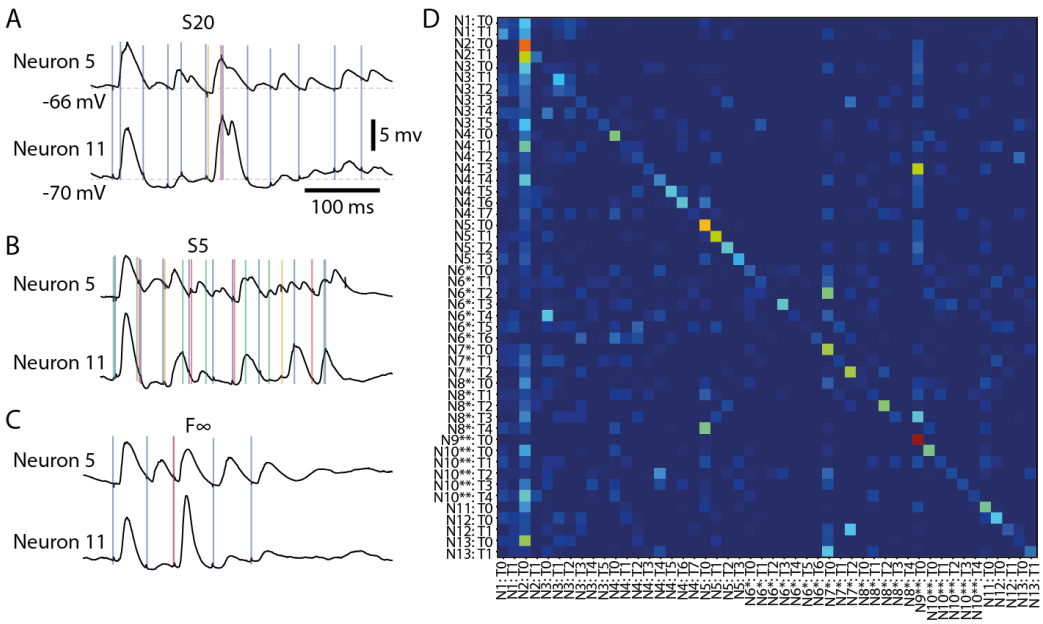

Fig. 3. Responses and response types to the same input patterns in different neurons. (**A**) Averages of intracellular responses of neuron #5 (N=100 responses) and neuron #11 (N=80 responses) to stimulation pattern S20. Vertical colored lines indicate the individual stimulation pulses of the stimulation pattern. (**B-C**) Similar display for the responses to two other input patterns. Neuron #5 pattern S5 is also illustrated in Fig. 2A-B,D. (**D**) Confusion matrix for all response types identified for all neurons for pattern F∞. Neurons (N#) recorded in the same experiments are indicated by * and **, respectively. T# indicates the response type.

Our results show that in combination with the current state of internal cortical activity, responses evoked by tactile sensory input patterns tend to fall into a limited subset of preferred response states, which are specific to each cortical neuron. The character of these recurring responses indicate that the networks defining the internal brain states, which may encompass the entire neocortex (*12, 13*), and which are known to be essential for forming the percept of each given external input (*14, 15*), have a non-uniform landscape of solutions that are adaptable to the time-evolving match between the internally set expectations and the actual sensory activation pattern. The findings suggest that the multidimensional latent state defined by large populations of neurons across the neocortex (*8*) work according to attractor like dynamics (*16*) but with multiple metastable states for each given input. The response types observed in individual neurons would thus be local subnetwork-instantiations of the input-updated brain-wide state estimations of the world and the own body, which could correspond to the fundamental mechanisms for perception and illusion phenomena. We expect these principles to reflect a general computational strategy used by the neocortex across all sensory systems.

References and Notes:

1. A. Spanne, H. Jorntell, Questioning the role of sparse coding in the brain. *Trends in neurosciences* **38**, 417 (Jul, 2015).

2. A. Luczak, P. Bartho, K. D. Harris, Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* **62**, 413 (May 14, 2009).
3. M. D. Golub *et al.*, Learning by neural reassociation. *Nature neuroscience* **21**, 607 (Apr, 2018).
4. A. Arieli, A. Sterkin, A. Grinvald, A. Aertsen, Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* **273**, 1868 (Sep 27, 1996).
5. P. Berkes, G. Orban, M. Lengyel, J. Fiser, Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* **331**, 83 (Jan 7, 2011).
6. J. Fiser, C. Chiu, M. Weliky, Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature* **431**, 573 (Sep 30, 2004).
7. C. Curto, S. Sakata, S. Marguet, V. Itskov, K. D. Harris, A simple model of cortical dynamics explains variability and state dependence of sensory responses in urethane-anesthetized auditory cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **29**, 10600 (Aug 26, 2009).
8. C. Stringer *et al.*, Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364**, 255 (Apr 19, 2019).
9. C. M. Oddo *et al.*, Artificial spatiotemporal touch inputs reveal complementary decoding in neocortical neurons. *Scientific reports* **8**, 45898 (Apr 04, 2017).
10. F. Bandeira, R. Lent, S. Herculano-Houzel, Changing numbers of neuronal and non-neuronal cells underlie postnatal brain growth in the rat. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 14108 (Aug 18, 2009).
11. F. Bengtsson, R. Brasselet, R. S. Johansson, A. Arleo, H. Jorntell, Integration of sensory quanta in cuneate nucleus neurons in vivo. *PloS one* **8**, e56630 (2013).
12. A. Wahlbom, J. M. D. Enander, F. Bengtsson, H. Jorntell, Focal neocortical lesions impair distant neuronal information processing. *The Journal of physiology* **597**, 4357 (Aug, 2019).
13. J. M. D. Enander *et al.*, Ubiquitous Neocortical Decoding of Tactile Input Patterns. *Frontiers in cellular neuroscience* **13**, 140 (2019).
14. F. A. Geldard, C. E. Sherrick, The cutaneous "rabbit": a perceptual illusion. *Science* **178**, 178 (Oct 13, 1972).
15. G. Robles-De-La-Torre, V. Hayward, Force can overcome object geometry in the perception of shape through active touch. *Nature* **412**, 445 (Jul 26, 2001).
16. D. L. Ringach, Spontaneous and driven cortical activity: implications for computation. *Current opinion in neurobiology* **19**, 439 (Aug, 2009).

Acknowledgments: The authors thank Jerry Loeb (USC Los Angeles) and Matthias Kohler (TUM Munich) for reviewing our **Funding**: This work was supported by the EU Grant FET 829186 ph-coding (Predictive Haptic COding Devices In Next Generation interfaces), the Swedish Research Council (project grant no. K2014-63X-14780-12-3). **Author contributions:** H.J. and J.N. designed the study. J.N. and H.J. performed the experiments. H.J. and J.N. wrote the article. J.M.D.E., J.N., H.M. and H.J. made the analysis. **Authors declare no competing interests. Data and materials availability:** All data are made available on FigShare (on acceptance).

Supplementary Materials:

Materials and Methods, Figures S1-S7, Table S1
References (##-##)