Published in final edited form as:

Neuron. 2022 May 18; 110(10): 1728–1741.e7. doi:10.1016/j.neuron.2022.02.016.

Brain-wide electrical dynamics encode individual appetitive social behavior

Stephen D. Mague[1,2,*], **Austin Talbot**[10,*], **Cameron Blount**[2], **Kathryn K. Walder-Christensen**[1,2,6], **Lara J. Duffney**[7], **Elise Adamson**[1,2,11], **Alexandra L. Bey**[1,2], **Nkemdilim Ndubuizu**[2], **Gwenaëlle E. Thomas**[1,3], **Dalton N. Hughes**[1,3], **Yael Grossman**[1,2], **Rainbo Hultman**[9], **Saurabh Sinha**[8], **Alexandra M. Fink**[2], **Neil M. Gallagher**[3], **Rachel L Fisher**[2], **Yong-hui Jiang**[7], **David E. Carlson**[5,11,†], **Kafui Dzirasa**[1,2,3,4,12,13,†]

[1]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA

[2]Dept. of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, North Carolina 27710, USA

[3]Dept. of Neurobiology, Duke University Medical Center, Durham, North Carolina 27710, USA

[4]Dept. of Neurosurgery, Duke University Medical Center, Durham, North Carolina 27710, USA

[5]Dept. of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, North Carolina 27710, USA

[6]Dept. of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, USA

[7]Dept. of Pediatrics, Duke University Medical Center, Durham, North Carolina 27710, USA

[8]Dept. of Neurology, Duke University Medical Center, Durham, North Carolina 27710, USA

[9]Department of Molecular Physiology and Biophysics, Psychiatry, University of Iowa, Iowa City, IA, 52242 USA

[10]Department of Statistical Science, Duke University, Durham North Carolina 27708, USA

Correspondence should be sent to: Kafui Dzirasa, M.D. Ph.D., Dept. of Psychiatry and Behavioral Sciences, Duke University Medical Center Durham, NC 27710, USA, kafui.dzirasa@duke.edu, Twitter: @KafuiDzirasa or David E. Carlson, PhD, Department of Biostatistics and Bioinformatics, Department of Civil and Environmental Engineering, Duke University, Durham, NC 27708, USA, david.carlson@duke.edu.

*These authors contributed equally

†Senior authors; Contributed equally

Author contributions

Conceptualization and Methodology – S.D.M., A.T., L.J.D., E.A., D.E.C., and K.D.; Formal Analysis – S.D.M., A.T., C.B., K.K.W.C., E.A., G.E. T., D.N.H., S.S., A.M.F., N.M.G., D.E.C., and K.D.; Investigation – S.D.M., A.T., C.B., K.K.W.C., L.J.D., E.A., A.L.B., N.N., D.N.H., Y.G., R.H., S.S., A.M.F., R.L.F., D.E.C., and K.D.; Resources –D.E.C., K.D.; Writing – Original Draft, S.D.M, A.T., C.B., D.E.C., and K.D.; Writing – Review & Editing, S.D.M, A.T., K.K.W.C., A.L.B., E.A., G.E.T., D.E.C., and K.D.; Visualization – S.D.M., A.T., D.E.C., and K.D.; Supervision – S.D.M., Y.H.J., D.E.C., and K.D.; Project Administration and Funding Acquisition – S.D.M., Y.H.J, D.E.C., and K.D. See Supplemental materials for detailed author contributions.

Declaration of Interests

The authors declare no competing interests

[11]Department of Civil and Environmental Engineering, Duke University, Durham North Carolina 27708, USA

[12]Department of Biomedical Engineering, Duke University, Durham North Carolina 27708, USA

[13]Lead Contact

Summary

The architecture whereby activity across many brain regions integrates to encode individual appetitive social behavior remains unknown. Here we measure electrical activity from eight brain regions as mice engage in a social preference assay. We then use machine learning to discover a network that encodes the extent to which individual mice engage another mouse. This network is organized by theta oscillations leading from prelimbic cortex and amygdala that converge on ventral tegmental area. Network activity is synchronized with cellular firing, and frequency-specific activation of a circuit within this network increases social behavior. Finally, the network generalizes, on a mouse-by-mouse basis, to encode individual differences in social behavior in healthy animals but fails to encode individual behavior in a 'high confidence' genetic model of autism. Thus, our findings reveal the architecture whereby the brain integrates distributed activity across timescales to encode an appetitive brain state underlying individual differences in social behavior.

eTOC Blurb

Mague, Talbot, et al., applied machine learning to discover a brain-wide electrical network that encodes individual rewarding social experiences in mice. Stimulation of circuits within the network increases social behavior. Strikingly, the network fails to encode individual social experience in a genetic mouse model of Autism.

Keywords

Social behavior; brain networks; electome; machine learning; autism

Introduction

Social behaviors play a critical role in survival. To appropriately regulate social behavior, mammals must integrate external sensory cues with internally generated brain-states. Many mechanisms whereby the brain processes external senses such as vision and audition have been elucidated. Similarly, studies have also uncovered multiple regions that contribute to complex social behavior. For example, anterior cingulate cortex (Cg) signals empathy in humans (Morrison et al., 2004), prefrontal cortex regulates social hierarchy in rodents (Wang et al., 2011), and medial dorsal thalamus (MD) plays a critical role in social appetitive behavior (Ferguson and Gao, 2018). Circuit-level interactions between regions have also been shown to play a role in regulating social behavior. Recent rodent studies have demonstrated that ventral hippocampus (Hip)➔prefrontal cortex circuits mediate social memory (Phillips et al., 2019), ventral tegmental area (VTA)➔nucleus accumbens (NAc) circuits encode social reward (Gunaydin et al., 2014), and prefrontal cortex➔amygdala

(AMY) circuits are critical for social avoidance and socially aversive learning (Allsop et al., 2018; Kumar et al., 2014; Schaich Borg et al., 2017). Nevertheless, the biological processes whereby the brain synthesizes this otherwise distributed information to regulate how individual animals respond under social appetitive conditions remain to be clarified. This question is central to determining how individual animals organize themselves within broader social frameworks under healthy conditions, and to revealing the potential mechanisms whereby neural coding of social information is disrupted under psychiatric disease states which broadly disrupt social function.

Human electroencephalographic (EEG) studies have described the emergence of synchronized electrical oscillations between cortical regions at the milliseconds time scale during social perception (Fraiman et al., 2014; Rodriguez et al., 1999), and functional magnetic resonance imaging (fMRI) studies have revealed synchronized neural activity across the brain at the seconds timescale (Sokolov et al., 2018). Together, this suggests that the brain integrates neural activity across multiple brain regions and timescales to encode appetitive states that drive individual social behavior. Nevertheless, collecting spatially- and temporally-resolved *in vivo* measurements of brain activity in humans, who can provide self-reports of their social emotional state, remains a challenge.

Though brain states have classically been inferred in preclinical models (i.e., rodents) using behavioral measurements in isolation, rather than directly measuring the brain-wide activity that underlies those states, several studies have linked physiology with behavior. For example, a murine *in vivo* calcium imaging study identified synchronous activity across multiple cortical and limbic regions on the 100ms–seconds timescale during exposure to social novelty (Kim et al., 2016). Similarly, we have previously observed synchronous electrical oscillations on the 10ms–100ms timescale during aversive states in mice and rats (Carlson et al., 2017; Hultman et al., 2018; Schaich Borg *et al.*, 2017). We have also found that these oscillations exhibit synchronous activity with millisecond-timescale cellular firing in the brain (Carlson et al., 2014; Hultman et al., 2016). Thus, we hypothesized the existence of a network-level mechanism involving the synchronization of oscillations that integrates cellular firing across brain regions and timescales (milliseconds to seconds) to corporately encode an appetitive brain state underlying individual differences in social behavior.

To address these questions, we implanted C57BL/6J (C57) mice in regions with well-established roles in social behavior including Cg, infralimbic (IL) and prelimbic (PL) cortex (the anatomic subdivisions of prefrontal cortex), AMY (basolateral and central), NAc (core and shell), MD, Hip, and VTA. We then recorded electrical oscillations and cellular firing across these regions, concurrently, as mice performed a social exploration task. Using machine learning, we integrated electrical activity across these regions and across the milliseconds to seconds timescale into what we call an electrical functional connectome ("*electome*"). By analogy to the connectome, the *electome* describes the detailed pattern of electrical interactions across a group of brain areas. Moreover, by analogy to gene networks, *electome* networks describe a collection of brain circuits within the *electome* that together encode distinct brain states.

Learning *electome* networks that comprise the total *electome* is typically an unsupervised process, but we augmented a supervised approach to increase relevancy to complex social behavior. After confirming that an appetitive social *electome* network we discovered is synchronized with cellular firing and generalizes to new animals on a mouse-by-mouse basis, we also showed that causal frequency-specific manipulation of network components modulates not only the electome but also behavior. Additionally, we showed that the network is dysfunctional in a genetic mouse model of ASD. Thus, our findings reveal a new mechanism whereby the brain integrates activity across space and time to encode social behavior in health and disease.

Results

Machine learning to discovering an appetitive social behavioral state

To discover the network architecture underlying a putative appetitive social brain state, we performed multisite electrical recordings while mice of the C57 strain were subjected to a task modeled after a classic social preference assay (Moy et al., 2007). In this behavioral assay, mice freely explore a large arena that is divided into two chambers: a small container housing a novel age- and sex-matched mouse of another strain (C3H) is situated in one chamber, and a second container holding a novel object is situated in the other (Fig. 1A). The location of the experimental mouse is tracked throughout ten minutes of exploration; social preference is calculated based on the relative time spent proximal to each container (Fig. 1B). Importantly, by repeating this assay across days with new social partner mice and objects (Fig. 1C), we collected nearly 100 minutes of electrical recordings for each of our 28 implanted mice. Moreover, because social behavior can vary from day to day, pooling data across 10 sessions provided a more reliable assessment of each individual animal's innate social preference.

As expected, mice spent substantially more time interacting with a social stimulus than an object across the recording sessions (main effect of stimulus $F_{1, 594.8}=349.1$, $P<0.0001$ using a 3-way repeated measures ANOVA of unequal variance comparing stimulus, sex, and session; there were no other significant main effects or interactions; n=19 male mice; n=9 female mice, see also Fig. 1C). Finally, because there were no main effects of sessions or sex, we pooled data across sessions, and we included both male and female mice in the group of animals utilized to discover the network structure within the neural data.

We used discriminative cross spectral factor analysis non-negative matrix factorization (dCSFA-NMF) to discover the network structure within this neural data (Talbot et al., 2020). dCSFA-NMF is a supervised machine learning approach that we designed to be both *descriptive* (i.e., discovers brain activity measures that are integrated across seconds of time) and *predictive* (i.e., discovers networked patterns of brain activity that encode external behavioral variables) (Vu et al., 2018). Importantly, dCSFA-NMF is based on widely accepted measures of brain activity, such that the resultant *electome* networks are *interpretable* (Vu *et al.*, 2018). Specifically, each learned *electome* network integrates local field potential (LFP) power (measurement of oscillatory amplitudes across frequencies resolved from 1 to 56Hz; neural correlate of cellular population activity and synaptic activity within brain regions), LFP synchrony (quantification of how two LFPs correlate

across frequencies resolved from 1 to 56Hz over a millisecond timescale; neural correlate of brain circuit function between brain regions), and LFP Granger synchrony (statistical forecasting based on Granger causality testing; neural correlate of information transfer within a circuit). Finally, our dCSFA-NMF model yields an activity score for each *electome* network, which indicates the strength at which that network is represented during each one-second segment of LFP. A given brain area or circuit can belong to multiple *electome networks*, providing the opportunity for distinct *electome networks* to functionally interact to yield a global emotional brain state (Hultman *et al.*, 2018). Thus, dCSFA-NMF integrates spatially distributed neural activity across milliseconds to seconds of time in a manner that both models naturally occurring brain networks and predicts external behavioral conditions widely shown to induce and/or reflect emotional states in mice (see Fig. 1D, for dCSFA-NMF schematic).

Our goal was to discover a network that encoded individual differences in sociability. Thus, we reasoned that this network should perform three computations: 1) the network should encode whether an animal was socially engaged, 2) the network should encode appetitive information, and 3) the network should integrate the first and second computations to encode individual behavioral outcomes in a social appetitive task. To achieve the first two criteria, we trained a model to discriminate social vs. object interactions using the data collected/recorded during the appetitive social preference task. Here, we reasoned that the social interactions during the social preference assay were social in nature and that they were more appetitive than object interactions. To meet the third criteria, we weighted each mouse in our model relative to their individual social preference, biasing our model towards fitting the mice with the strongest social preference (Fig. 1E). As such, we set out to identify a network that encoded and integrated social and appetitive information in a manner that regulates individual behavioral outcomes.

Modeling our data with six *electome* networks optimally balanced complexity (i.e., explaining more variance) with parsimony (i.e., choosing fewer networks to represent the brain; see supplemental Fig. S1c). As expected, the supervised *electome* network showed the highest predictive performance (*electome* network #1, hereafter referred to as Social-*Electome* Network; *EN-Social*; Fig. 2A). We then probed the activity of *EN-Social* across all the timepoints while mice explored the two-chamber assay. Though our initial learning model only used data widows labeled as social and object classes, we found that *EN-Social* activity exhibited dynamics that reflected behaviorally relevant task variables. Specifically, we found that *EN-Social* activity increased at the onset of social interactions and sloped downward as epochs of social interactions concluded. We also found that *EN-Social* activity decreased during object interactions (Fig. 2B). Critically, the discriminatory strength of *EN-Social* was directly correlated with social preference across our population of mice ($P=0.002$, $R=0.56$ using spearman correlation; see Fig. 2C), suggesting that our weighting strategy successfully biased dCSFA-NMF to learn a network that integrated social and appetitive information to drive individual behavior.

Network dynamics reflect brain-wide cellular activity

We next verified that this putative socially-relevant *electome* network was a *bona fide* representation of biological activity and not simply an abstract mathematical construct (Hultman *et al.*, 2018). To achieve this, we determined whether *EN-Social* activity demonstrated a relationship with the activity of cells recorded simultaneously from the implanted brain regions, an undisputed reflection of biological function. Since we found that cellular firing was broadly related to social vs. object interactions in the two-chamber assay (112/326 cells, see Fig. 3A), we used a permutation test to rigorously test our findings. Here, we found that *EN-Social* exhibited a relationship to the activity of ~18% of the cells we recorded (Fig. 3B–C), much higher than would be expected by chance. Thus, we confirmed that *EN-Social* reflects a network-level neural process that emerges from cellular firing across the brain (Carlson *et al.*, 2014; Hultman *et al.*, 2018).

Individual brain network components fail to independently encode individual behavior

After finding that *EN-Social* reflects an appetitive social brain state in the group of mice used to train our model, we explored the extent to which activity within the nodes and circuits that composed *EN-Social* independently encoded social behavior. *EN-Social* mapped to LFP theta (4–11Hz) power within all the implanted regions. Additionally, *EN-Social* comprised prominent theta synchrony across all the implanted brain regions except Hip (Fig. 4A, see blue highlights). The network also mapped to oscillatory activity in two higher frequency bands: 30–40Hz and 50–56Hz. The 30–40Hz oscillations showed local activity in Hip and MD, as well as synchrony between all the implanted brain regions except Hip (Fig. 4A, see green highlights). The higher frequency gamma band (50–56Hz) showed local activity within all the brain regions we measured except AMY and Cg, and synchrony between all the implanted brain regions except Hip (Fig. 4A, see red highlights). Prominent circuit directionality, quantified as the difference in the Granger synchrony between each pair of brain regions (i.e., area A➔B versus area B➔A), was observed only in the theta frequency range (Fig. 4B–C). This activity emerged from PL, IL, and AMY, relayed through Cg and NAc to MD, and converged in VTA (see supplemental Fig. S2). Thus, *EN-Social* emerged from brain regions previously shown to play a prominent role in social behavior and converged on a brain region critical for reward regulation.

We then explored the extent to which the constituent circuit elements of *EN-Social* encoded social behavior. Specifically, we quantified the extent to which activity in several prominent network nodes/circuit elements (4–11Hz power and coherence measures) encoded social vs. object interactions. While many of these circuit elements contained social information, none of them encoded social engagement as strongly as *EN-Social* (Fig. 4D). In fact, some of the circuit elements showed higher activity during object interactions (i.e., MD). Strikingly, when we compared encoding in each of these network nodes/circuit elements to social preference across the training set of mice, we failed to discover any significant relationships (Fig. 4E). Thus, while many previously identified circuits contained information regarding ongoing social encounters, individual behavior was only successfully encoded at the network level.

EN-Social generalizes across mice to encode social context and valence

We next set out to establish *EN-Social* as a true measure of an appetitive social brain state by testing the generalizability of this network, a gold-standard machine-learning validation strategy (Vu *et al.*, 2018). Specifically, rather than simply testing whether *EN-Social* encoded object vs. social interactions in the same group of animals performing additional sessions of the two-chamber behavioral assay, we tested whether the *electome* network we learned in the initial group of mice generalized to a new cohort of C57 mice performing a different behavioral assay that also quantifies social behavior (Fig. 5A). Likewise, we also examined *EN-Social* activity in two orthogonal behavioral tasks to test whether this network was indeed encoding a brain state relevant to the valence of external stimuli. To emphasize, the machine learning electome model was completely blind to these tasks and data, so this represents a test of its generalizability. Thus, we tested whether 1) our network encoded social engagement 2) whether the network encoded appetitive states, and 3) whether the network effectively integrated social and appetitive information to regulate behavior.

First, we acquired neural activity in eight mice exposed to our Free object/social interaction test (FOSIT; Fig. 5B). In this assay, the C57 subject mice are repeatedly exposed to a novel object or a novel conspecific (age- and sex-matched) stimulus mouse during a single recording session. Encounters in FOSIT occur in the absence of sub chambers used for the two-chamber assay, such that the stimulus partner mouse can initiate social interactions with the implanted experimental mouse. When we projected neural data obtained during FOSIT into our initial *electome* model, we found that *EN-Social* activity was higher in reciprocal social interactions than it was in the object condition ($\chi^2_{3,39}$ =20.52 and P=1.3×10^{-4} using Freidman's test; P=0.002 using post-hoc two-tailed sign-rank test with false discover rate correction; n=10 new mice). *EN-Social* activity was also higher during non-reciprocated social interactions initiated by the stimulus partner mouse than when the two mice were not interacting (P<0.02 using two-tailed sign-rank test). Finally, *EN-Social* activity was highest when the experimental mouse was engaged in social interactions (P=0.03 using two-tailed sign-rank test). Together, our findings verified that *EN-Social* generalized to encode social information in new mice performing a different appetitive social task.

Having established that *EN-Social* contained social information, we next tested whether *EN-Social* encoded the valence of stimuli (rather than solely encoding sensory information or the salience of cues) by testing whether the network signaled valence in other orthogonal behavioral tasks. First, we subjected a new cohort of C57 mice to an intermittent sucrose access test designed to model an appetitive state associated with food reward. Here, implanted mice were individually housed in an arena fitted with two nose poke holes. A syringe placed in the back of each hole dispensed 10μL water in response to a nose poke. After several days of habituation, water vials were replaced with 2% sucrose for 1.5 hours during the dark cycle. Neural recordings were acquired during intermittent access to sucrose and the subsequent water consumption period and then projected into our initial *electome* model. *EN-Social* activity was higher following nose pokes for sucrose than for water (Fig. 5C; P=0.016 using sign-rank test; n=7 new mice), and *EN-Social* encoded the sucrose vs. water conditions to the same extent that it encoded the social vs. object conditions in the FOSIT (U=129; P=0.48 using rank-sum test; AUC = 0.59±0.02 and 0.61±0.02, for *EN-*

Social in the sucrose vs. water condition and the social vs. object condition in the FOSIT). Thus, activity in *EN-Social* encoded reward in a behavioral context that was unrelated to social behavior.

Second, we probed whether *EN-Social* activity encoded the location of mice on a classic elevated plus maze assay used to model avoidance behaviors. In this assay, mice are placed on a large plus-shaped platform that is elevated off the floor (Fig. 5D). Two of the arms of the maze are walled and the other two are open. The time that mice spend in the open arms of the maze is increased by myriad anxiolytic manipulations and decreased by anxiogenic manipulations (Krishnan et al., 2007; Marcinkiewcz et al., 2016; Rodgers et al., 1992). Furthermore, several anxiety-related neural signatures are observed as animals explore the open arms of the maze (Felix-Ortiz et al., 2016; Padilla-Coreano et al., 2016; Padilla-Coreano et al., 2019; Seidenbecher et al., 2003); thus, the open arms of this assay are widely accepted as an environmental context that induces an aversive anxiety-like state in C57 mice. When we projected neural data acquired from C57 mice subjected to the elevated plus maze into our *electome* model, we found that *EN-Social* encoded the open vs. closed arm location of mice [$\chi^2_{2,56}$ =14 and P<0.001 using Freidman's test; P=0.006 using two-tailed sign-rank test; n=19 mice (12 mice from the initial *EN-Social* training set in a new, untrained-on behavioral condition and 7 new mice); Fig. 5D]. Strikingly, *EN-Social* activity was lower in the open arm than the closed arm (AUC=0.42±0.02 for open arm vs. closed arm, with an AUC below 0.5 signifying a negative relationship but the same strength as an AUC=0.58 relationship; P=0.004 using two-tailed sign-rank test). Moreover, *EN-Social* activity was higher while mice were in their home cage (prior to the assay) than when they were in the open arm of the assay (P=0.049 using two-tailed sign-rank test). These results demonstrated that the network did not simply encode a brain state related an animal's arousal or the salience of sensory cues. Rather, since the strength of *EN-Social* encoding was the same for the FOSIT and elevated plus maze test assays (U=275; P=0.24 for comparisons of |AUC-0.5| for the two tasks using a rank-sum test), our findings showed that *EN-Social* encoded a state related to the valence of external stimuli. Thus, *EN-Social* encoded social, appetitive, and aversive information in a manner that generalized across mice, raising the question as to whether activity in *EN-Social* was selective.

EN-Social fails to encode individual responses to non-social stimuli or social aversion

To probe whether *EN-Social* showed selectivity in its encoding, we first tested whether *EN-Social* encoded individual differences in the sucrose drinking (non-social appetitive) and elevated plus maze (non-social aversive tasks). When we explored the sucrose vs. water condition, we found that *EN-social* decoding failed to correlate with the sucrose preference of individual mice (P=0.43 using spearman rank correlation, Fig. 6A, left). In fact, the mice which showed the highest sucrose preference tended to show the least network discrimination of sucrose vs. water. Similarly, *EN-Social* decoding of elevated plus maze location failed to correlate with the closed arm preference of individual mice (P=0.96 using spearman rank correlation, Fig. 6A, right). Thus, *EN-Social* activity did not regulate individual behavioral outcomes in the sucrose drinking or elevated plus maze assays as it had in the social preference assay.

Next, we tested whether *EN-Social* preferentially encoded social information or the valence of external stimuli. We reasoned that a social stimulus would be expected to increase *EN-Social* activity, while an aversive stimulus would be expected to decrease network activity. Thus, exposure to a socially aversive stimulus would allow us to determine whether the network showed encoding preferences. For this experiment, we exploited the widely validated chronic social defeat paradigm (Berton et al., 2006). Thirty-four implanted male mice were subjected to chronic social defeat stress (Fig. 6B) in which C57 mice were exposed to 10 consecutive days of subordination stress from a larger, more aggressive CD1 mouse. We then recorded five minutes of neural activity while these mice were in a small holding chamber. An aggressive CD1 mouse was then placed immediately outside the holding chamber, and we recorded five additional minutes of neural activity (Fig. 6C, top). Critically, during the latter period, the CD1 mice engaged in behaviors that were presumably aversive to the C57 mice, including biting them through the holding chamber openings. The following day, implanted C57 mice were subjected to a classic single chamber social preference test during which they could freely explore an arena that contained a CD1 mouse housed in a small chamber (Fig. 6C, bottom). Whereas control animals typically prefer spending time in the zone proximal to the CD1 mouse, following chronic social defeat stress, ~60% of C57 mice show behavioral avoidance (i.e., spend less time exploring) of the CD1 mouse (Berton *et al.*, 2006).

We identified twenty-one implanted mice (62%) that showed avoidance to the CD1 mouse and used these for our subsequent analysis. Thus, for this subset of mice, exposure to the CD1 was an aversive social experience. When we quantified network activity for these mice when they were in the small holding camber, we found that *EN-Social* exhibited strong activation to the aversive social stimulus ($P<6\times10^{-6}$ using a two-tailed sign-rank test; AUC=0.72±0.02, Fig. 6D). This showed that the network preferentially encoded social information (which we previously determined increased network activity) rather than negatively valanced stimuli (which we previously found decreased network activity).

Finally, we tested whether EN-*Social* also encoded individual behavioral outcomes in response to socially aversive information. When we compared individual network encoding during the socially aversive neurophysiological assay to behavioral outcomes measured during subsequent behavioral testing, we found no relationship between these two variables (P=1 using spearman rank correlation, Fig. 6D). Thus, *EN-Social* did not regulate differences in individual behavior under socially aversive conditions. Taken together, these findings showed that *EN-Social* selectively integrates social and appetitive information to encode individual behavioral outcomes.

Causal manipulation of EN-Social circuit impacts network activity and behavior

After demonstrating that *EN-Social* encoded a social appetitive brain state, we tested the impact of selective manipulation of neural activity within in a key node of the network during concurrent neurophysiological recordings and behavioral assessments. We targeted the PL➔NAc element, a component of *EN-Social* (4–11Hz, see Fig. 4A–C), because a prior causal optogenetic had study implicated this circuit in appetitive social behavior (Murugan et al., 2017). We implanted mice (n=10) with recording electrodes and bilateral stimulating

fibers in NAc following infection with channel rhodopsin-2 (AAV5-CamKII-Chr2, Fig. 7A) in PL, bilaterally. Animals were then subjected to our FOSIT assay during stimulations with blue light (473nm) to activate Chr2, or yellow light (589nm) as a negative control (10Hz, 1mW bilaterally, 5ms pulse-width; Fig. 7B). Critically, we confirmed activation of the prelimbic cortex➔nucleus accumbens terminals in all of the experimental animals (Fig. 7C, left), and we excluded mice that exhibited pronounced local oscillatory responses to blue light stimulation across all of the implanted brain regions (Fig. 7C, right; n=2) given our prior observations that supraphysiological optogenetic stimulation can suppress network level activity (Hultman *et al.*, 2018).

Causal activation of PL→NAc at 10Hz enhanced *EN-Social* activity and increased social behavior. Specifically, we projected LFP data into our initial *electome* model and quantified *EN-Social* activity during periods of social interaction (Fig. 7D). We found that blue light stimulation enhanced *EN-Social* activity compared to yellow light stimulation (P=0.016 using sign-rank test; n=7 mice; Fig. 7E). Next, we compared the amount of time mice spent socially interacting during periods of blue and yellow light stimulation. We found that blue light stimulation increased social interaction time in the FOSIT ($F_{1,13}$=5.76; P=0.03 for stimulation effect using two-way RMANOVA; Fig. 7F, left). No differences in object interaction time were observed for the two light stimulation conditions ($F_{1,13}$=1.67; P=0.22 for stimulation effect using two-way RMANOVA; Fig. 7F, right). Interestingly, we observed lower social interaction and network activity in response to stimulation of PL➔NAc at 20Hz (1mW bilaterally, see Supplemental Fig. S3). Thus, our findings showed that activation of PL➔NAc at the endogenous frequency of the network enhanced both *EN-Social* activity and increased social interaction. Causal activation of a different circuit element that was not strongly represented in *EN-Social* did not increase *EN-Social* activity (PrL➔AMY, 10Hz, 1mW bilaterally; n=7 mice; P = 0.94 using one-tailed sign-rank test, see supplemental Fig. S4) nor did it alter social or object interaction time ($F_{1,22}$=0.23, P=0.63; $F_{1,20}$=0.80, P=0.38; for stimulation effect on social and object interaction, respectively, using a two-way RMANOVA; n=11 mice, see Supplemental Fig. S4). Taken together with our other validation experiments, these results provided broad evidence that *EN-Social* was causally related to appetitive social behavior.

After establishing *EN-Social* as a generalized and putatively causal appetitive social-emotional brain state under healthy conditions, we wondered whether any aspects of *EN-Social* would be altered in a psychiatric disease state associated with social deficits. Autism spectrum disorder (ASD) is a pervasive neurodevelopmental disorder for which social deficits are a core feature. They include deficits in social attention and engagement and deficient processing of social information (Crawford et al., 2016; Dawson et al., 2012; Dawson et al., 2004; Klin et al., 2015). Genetic manipulations are implicated in ~52% of ASD cases (Gaugler et al., 2014), and one such high confidence gene is ANK2 which codes the Ankyrin-B protein (SFARI-GENE, 2020; Yang et al., 2019). Importantly, unlike many other genes that are implicated in syndromic ASD, ANK2 mutations yield social deficits without impacting executive cognitive dysfunction. We previously developed an *Ank2* mouse model based on a gene mutation identified in a patient with ASD. Heterozygous mice show decreased social behavior on multiple assays, decreased juvenile vocalizations, and increased cognitive flexibility (Yang *et al.*, 2019).

We implanted adult *Ank2* male mice and their wild-type littermate controls with recording electrodes and subjected them to ten sessions of neural recordings in the two-chamber social assay (Fig. 8A–B). *Ank2* mice exhibited normal social preference (U=64; P=0.86 using rank-sum test; Fig. 6C) and did not show seizure activity (Fig. 8D–E). When we projected their LFP activity into our initial *electome* model, *Ank2* mice and their littermate controls both exhibited *EN-Social* activity that was higher during social vs. object encounters ($F_{1,16}$=30.5; P=4.7×10^{-5} for social vs. object effect using a mixed-model ANOVA; n=11 and 7, for wild-type and *Ank2* mice, respectively; Fig. 8F). Furthermore, no differences in *EN-Social* activity were observed across genotype ($F_{1,6}$=0.58; P=0.46 for genotype effect; $F_{1,16}$=1.04; P=0.32 for interaction effect), demonstrating that *EN-Social* continued to encode socially relevant information in the mutants.

The discriminatory strength of *EN-Social* was directly correlated with social preference on a mouse-by-mouse basis across the group of wild-type littermate control mice ($F_{1,14}$=10.1; P=0.007 for interaction effect using Analysis of Covariance; P=0.004 and RHO=0.81 for wild-type mice using spearman rank test; see Fig. 8G), demonstrating that the socially appetitive brain state we discovered in our original group of training animals generalized to new mice. Strikingly, *EN-Social* activity was not correlated with social preference in the *Ank2* mutants (P=0.14 for *Ank2* mutants using spearman rank test; see Fig. 8H). To probe whether this network level dysfunction reflected a failure of *EN-Social* to encode appetitive information, or a select failure to integrate social and appetitive information, we subjected normal and mutant mice to a sucrose drinking assay. When we compared *EN-Social* activity in response to sucrose consumption, we observed similar activity between normal and mutant mice (U=64; P=0.48 using rank-sum test). Thus, only the *EN-Social* integration code was disrupted by *Ank2* dysfunction such that network activation failed to regulate individual behavior. Altogether, these results confirmed that *EN-Social* was altered in a disease state associated with social deficits.

Discussion

The manner whereby cells, segregated across multiple brain regions, integrate their activity over time to generate social brain states remains an unaddressed question. Human studies have sought to discover this network-level mechanism by probing changes in brain-wide hemodynamic responses using fMRI and/or fast electrical activity across the scalp using EEG. These studies have revealed multiple brain regions and several fast neural oscillatory features that putatively contribute to social processing (Fraiman *et al.*, 2014; Rodriguez *et al.*, 1999; Sokolov *et al.*, 2018). Nevertheless, fMRI is limited in its ability to resolve neural activity at the timescale of cellular activity in the brain (i.e., milliseconds), EEG does not quantify neural activity deep within the brain, and causality testing via direct manipulation of the human brain remains a challenge. Preclinical animal studies, on the other hand, readily facilitate causality testing of genetic and cellular/molecular mechanisms; however, approaches that monitor electrical activity across multiple regions have yet to be broadly applied to the study of appetitive social behavior. Given these limitations, network models that describe the causal mechanism whereby fast neural activity throughout the depth of the brain integrates across space and time to encode social-appetitive behavior remain elusive.

Here we implanted recording wires into eight cortical and limbic brain regions located through the depth of the brain, allowing us to record millisecond-timescale electrical fluctuations as mice engaged in behaviors used to model appetitive and aversive social brain states. Our neural recordings yielded 5152 features that quantified fast timescale (i.e., milliseconds to hundreds of milliseconds) region-specific activity and between-region circuit activity each second. Importantly, the behaviors we sought to model were multimodal as social processing undoubtably exploits visual, tactile, olfactory, and auditory information and regulates motor output. As such, we anticipated that animals would exhibit neural activity profiles in 'non-social' brain regions (e.g., motor cortex) that were different during encounters with another mouse than during an encounter with an object. Thus, we did not have a principled strategy for selecting additional 'negative control regions' to include in our data set; rather, we allowed circuit elements that did not differ between social and non-social encounters (as represented by a specific brain area pairs and frequencies) within our data set to function as the putative negative controls.

Using machine learning, we discovered the biophysiological patterns whereby these features integrated across seconds of time to encode a socially appetitive brain state. Not all circuits were captured by this network, providing evidence that our network was indeed selective for specific network elements. Not only did we discover that activity in the resulting *electome* network encoded the onset and termination of social interaction epochs (Fig. 1E), but we also confirmed that the activation strength of *EN-Social* was correlated with the social preference of individual mice (Fig. 1F). Both these properties generalized to new groups of mice that were not used to discover the initial network on a mouse-by-mouse basis (see Fig. 8G and supplemental Fig. S5), and the network generalized across sex (Block et al., 2020). Strikingly, we found that *EN-Social* generalized to encode active and passive social engagement in a different task that allows two freely behaving mice to interact with each other (i.e., FOSIT), confirming its validity. *EN-Social* also encoded food reward. This network exhibited some spectral overlap (based on brain regions, frequencies, and directionality composition) with another *electome* network we recently found to signal goal progress (Vu et al., 2019), suggesting that the *EN-Social* may exploit general brain circuits which encode reward. *EN-Social* also negatively encoded a brain state related to open-arm avoidance on the elevated plus maze, demonstrating that the network signaled the valence of external stimuli.

To determine if we could disentangle the social (positive) and valence (positive and negative) coding aspects of *EN-Social*, we tested mice in a socially aversive context. By exploiting the chronic social defeat paradigm, we found that the susceptible mice (which exhibit social avoidance) exhibit *EN-Social* activation during an aversive exposure to an aggressor mouse. Most importantly, *EN-Social* failed to encode individual behavior in the non-social appetitive (sucrose drinking), non-social aversive (elevated plus maze), and social-aversive conditions we tested. Thus, *EN-Social* selectively encoded individual differences in socially appetitive conditions, suggesting that the network integrates socially relevant information with appetitive circuits to drive individual behavioral outcomes.

When we tested many of the circuit elements that composed *EN-Social* independently, we found that most contained social information (AUC significantly higher or lower than

0.5). This expected finding enabled us to benchmark our findings against myriad studies establishing the role of these circumscribed circuits in social behavior. None of the circuit elements we tested encoded social behavior as strongly as *EN-Social*, demonstrating that *EN-Social* activity better explained social behavior than the previously established circuit elements (many of which have been the subject of causal manipulations). Even more striking, none of the circuit elements encoded individual difference in behavior. Thus, while these established circuit elements may indeed contain socially relevant information, they must be integrated together to determine individual social behavior.

Activity in the *electome* network correlated with cellular firing throughout the brain, confirming its biological significance. The network was composed of theta oscillations (4–11Hz) that synchronized across most of the regions we measured, showing directionality that emerged from AMY, PL, and IL, relayed through Cg and NAc to MD, and converged in VTA. Critically, directionality by no means implies that information flow is unidirectional (i.e., Granger A➔B exceeding Granger B➔A does not denote Granger A➔B but not Granger B➔A), nor does directional Granger coherence preclude other regions serving as anatomic relays (i.e., Granger A➔B does not exclude Granger A➔Z➔B). Nevertheless, it is notable that the activity pattern we discovered in *EN-Social* mirrored findings from other causal studies aimed at dissecting the individual anatomical circuits that contribute to social behavior. For example, while hyperactivation of PL projection neurons disrupts social preference (Yizhar et al., 2011), projection-specific studies revealed that PL ➔NAc, but not the PL➔AMY or PL➔VTA circuits, mediates this effect (Murugan *et al.*, 2017). This aligns with the directionality in *EN-Social* which exhibits activity in the PL➔NAc, but not the PL➔AMY or PL➔VTA circuits. Moreover, we confirmed that causal stimulation of PL➔AMY failed to increase *EN-Social* or social behavior.

In contrast to these previous findings, here we found that stimulation of PL➔NAc induced, rather than suppressed, social behavior, at least with stimulation at the 10Hz frequency. Our results comparing the network activity and behavioral outcomes induced by different stimulation frequencies suggest that this difference may in part be due to the higher stimulation frequency utilized in the prior work (Supplemental Figure S3), or other contributing factors such as the context in which social encounters occurred (novel area vs. habituated arena), or difference in experimental design (within-subjects vs. across-subjects) (Murugan *et al.*, 2017). Notably, several of these prior behavioral outcomes were observed using a 20Hz stimulation, which was not found as a component circuit of *EN-Social*, potentially explaining why stimulating this pathway at 10Hz enhanced *EN-Social* activity and promoted social behavior. Critically, together these findings also highlight an important role for using *electome networks* to interpret behavioral outcomes from causal manipulation studies, since causally activating a circuit at a frequency that does not match endogenous circuit activity may function to suppress behaviorally relevant network activity.

Prior work has also implicated Cg➔AMY in mediating aversive social-emotional states (Allsop *et al.*, 2018). While one of our early *electome* models learned in a socially aversive context identified this same circuit pathway (Schaich Borg *et al.*, 2017), Cg➔AMY was not prominently featured in our current appetitive social *electome* network. Thus, the network we discovered here clarifies how distinct circuits integrate in a normal physiological context

to encode an appetitive social-emotional brain state. Future analysis using adversarial machine learning models may clarify whether and/or disambiguate which specific aspects of *EN-Social* uniquely signal appetitive social behavior rather than generally signal all social or appetitive brain states.

Finally, we also employed an *Ank2*-based mouse model of ASD as a naturalistic strategy to causally manipulate *EN-Social*. Importantly, despite exhibiting social deficits across several behavioral assays, we previously found that *Ank2* mice demonstrate normal social preference in the two-chamber assay. Thus, we anticipated that this assay would enable us to probe *EN-Social* using the *Ank2* mice under conditions where they showed similar behavioral outputs as their wild type littermates. Like our causal behavioral (e.g., elevated plus maze open arm) and optogenetic manipulations in healthy mice, alteration of *Ank2* function also disrupted *EN-Social* function. *EN-Social* continued to encode the difference between social and object interactions in the ANK2 mutants, consistent with our clinical observations that most individuals with high functioning autism can discriminate other individuals from objects. The network also showed similar activity between mutants and healthy controls during an appetitive sucrose consumption task. Nevertheless, *EN-Social* failed to encode individual differences in social preference in the mutants. Thus, *Ank2* dysfunction selectively disrupted the ability of *EN-Social* to integrate social and appetitive information in a manner that regulated individual behavior. Critically, since the mutants continued to exhibit normal social preference, these findings raise the intriguing potential that a different set of brain circuits or networks may sub-optimally regulate socially appetitive behavior in the mutants. Future experiments in which *electome* networks are trained across larger groups of *Ank2* mutants may clarify this question.

Overall, we set out to identify the network level architecture that encodes a social appetitive brain state in mice. Using machine learning, we discovered a network that both encodes social vs. objection conditions and the individual social preference of mice across a population. Conceptually, we reasoned that this network should also capture appetitive information to successfully encode these two features. We then used behavior as a causal manipulation to validate the social vs. object and appetitive encoding functions of the network. In each case, we also found that the network generalized to new subjects. Moreover, we also showed the network only regulated individual behavioral differences related to social function, thereby establishing its selectivity. Finally, we employed a series of causal optogenetic manipulations, and a causal genetic manipulation related to ASD to validate the architecture of the network and its social encoding properties. Together, these findings provide compelling evidence that *EN-Social* encodes a generalized social appetitive brain state in healthy mice. Moreover, we also show that the network encodes individual difference in behavior.

Social dysfunction is at the heart of nearly all psychiatric disorders. Nevertheless, the neural substrates that play a determinative role in individual social behavioral outcomes remain elusive. This is in part due to the lack of socially relevant biomarkers in preclinical models that generalize on a subject-by-subject basis. Overall, this study establishes a collection of neural circuits that corporately encode the appetitive brain state underlying individual

differences in social behavior. Thus, this network can serve as a novel substate to objectively quantify social brain state function in broad range of mouse models of psychiatric disorders.

STAR Methods

RESOURCE AVAILABILITY

Lead Contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Kafui Dzirasa (kafui.dzirasa@duke.edu).

Materials availability—No unique reagents generated in this study

Data and code availability

- The code base for dCSFA-NMF analysis can be found at https://doi.org/10.5281/zenodo.6078304
- All data required to reanalyze the data reported in this paper will be shared by the lead contact upon request. Data requested for additional academic/non-commercial purposes will shared by the lead contact upon request.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animal Care and Use—Male and female C57BL/6J (C57) mice purchased from the Jackson Labs were used for two-chamber experiments used to train the initial Electome model, the subsequent studies using the free object social interaction test, the forced interaction test, the elevated plus maze, the sucrose consumption task, and the optogenetic manipulation studies. *Ank2* mutant mice were generated as previously described (Yang *et al.*, 2019). These male mutants and their WT littermate controls were bred within the Duke Vivarium. C3H strain mice used for the two-chamber social interaction test were purchased from Jackson Labs. CD1 male mice purchased from Charles River Laboratory. CD1 mice were individually housed. All other mice were housed 3–5 per cage. Mice were maintained on a 12-hour light/dark cycle, in a humidity- and temperature-controlled room with water and food available *ad libitum*.

Studies were conducted with approved protocols from the Duke University Institutional Animal Care and Use Committee and were in accordance with the NIH guidelines for the Care and Use of Laboratory Animals. Studies were conducted using mice that were 12–24 weeks old.

METHOD DETAILS

Electrode implantation surgery—Mice were anesthetized with 1% isoflurane, placed in a stereotaxic device, and metal ground screws were secured above the cerebellum and anterior cranium. The recording bundles designed to target basolateral and central amygdala (AMY), medial dorsal thalamus (MD), nucleus accumbens core and shell (NAc), VTA,

medial prefrontal cortex (mPFC), and VHip were centered based on stereotaxic coordinates measured from bregma (Amy: −1.4mm AP, 2.9 mm ML, −3.85 mm DV from the dura; MD: −1.58mm AP, 0.3 mm ML, −2.88 mm DV from the dura; VTA: −3.5mm AP, ±0.25 mm ML, −4.25 mm DV from the dura; VHip: −3.3mm AP, 3.0mm ML, 3.75mm DV from the dura; mPFC: 1.62mm AP, ±0.25mm ML, 2.25mm DV from the dura; NAc: 1.3mm AP, 2.25mm ML, −4.1 mm DV from the dura, implanted at an angle of 22.1°). We targeted cingulate cortex, prelimbic cortex, infralimbic cortex using the mPFC bundle by building a 0.5mm and 1.1mm DV stagger into our electrode bundle microwires. Animals were implanted bilaterally in mPFC and VTA. All other bundles were implanted in the left hemisphere. The NAc bundle included a 0.6mm DV stagger such that wires were distributed across NAc core and shell. We targeted BLA and CeA by building a 0.5mm ML stagger and 0.3mm DV stagger into our AMY electrode bundle. In order to mitigate pain and inflammation related to the procedure, all animals received carprofen (5 mg/kg, s.c.) injections once prior to surgery and then once every 24 hours for three days following electrode implantation.

Histological Confirmation—Histological analysis of implantation sites was performed at the conclusion of experiments to confirm recording sites used for neurophysiological analysis. Animals were perfused with 4% paraformaldehyde and brains were harvested and stored for 24 hrs in PFA. Brains were cryoprotected with sucrose and frozen in OCT compound and stored at −80°C. Brains were sliced at 35μm and stained using either DAPI (ab104139, AbCam, Cambridge, MA), NeuroTrace fluorescent Nissl Stain (N21480, ThermoFisher Scientific, Waltham, MA) or cresyl violet (C5042, Sigma-Aldrich, St. Louis, MO) using standard protocols. Images were obtained using a Nikon Eclipse fluorescence microscope at 4x and 10x magnifications. We took the following approaches to histological confirmation. When we performed complete histological analysis on 54 animals, we found 11/432 mistargeted implants (2.5% error rate). We observed a similar error rate (~3%) after complete histological analysis on an additional 56 mice. Since machine learning analysis benefits from larger data sets and can be more robust to data variance than classic frequentist statistics, we employed the following strategy. To learn our *EN-Social* model, we concluded that a training set containing 27/28 accurate data points per region was more desirable than a training set that contained 21/21 accurate data points per region. Thus, we used all 28 implanted animals to learn our initial model. We employed a similar strategy for our validation analysis. Specifically, presuming accurate targeting with 97% certainty, we included animals with missing or damaged histological slices in our analysis. However, if there was clear histological confirmation of mistargeting for any of the recorded regions, the animal was removed.

Neurophysiological data acquisition—Mice were connected to a headstage (Blackrock Microsystems, UT, USA) without anesthesia, and placed in each behavioral arena. Neuronal activity was sampled at 30kHz using the Cerebus acquisition system (Blackrock Microsystems Inc., UT). Local field potentials (LFPs) were bandpass filtered at 0.5–250Hz and stored at 1000Hz. An online noise cancellation algorithm was applied to reduce 60Hz artifact. Neuronal data were referenced online against a wire within the same brain area that did not exhibit a SNR > 3:1. At the end of the recording, cells were sorted again using an offline sorting algorithm (Plexon Inc., TX) to confirm the quality of the

recorded cells. Only cellular clusters well-isolated with respect to background noise, defined as a Mahanalobis distance greater than 3 compared to the null point, were used for our unit-Electome Factor correlation analysis. Clusters that exhibited more than 99% of their inter-spike-interval distribution above 2ms were defined as single units (93.5% of recorded neurons). Ultimately, we chose to use both single and multi-units for our analysis since our sole objective was to determine whether the Electome Network activity showed temporal dynamics that reflected cellular activity. This strategy mirrors our prior experiments probing the network level mechanisms underlying depression vulnerability (Hultman *et al.*, 2018). Neurophysiological recordings were referenced to a ground wire connected to both ground screws.

LFP preprocessing to remove signal artifact—We used a heuristic to remove recording segments with non-physiological signals. First, we estimated the envelope of the signal in each channel using the magnitude of the Hilbert transform. For any 1-second window where the envelope exceeds above a pre-selected low threshold, the entire segment is removed if the envelope exceeds a second, high threshold at any point within that window. The two thresholds were determined independently for each brain region. The high threshold was selected to be 5 times the median absolute deviation of the envelope value for that region. Five median absolute deviations was chosen as the high threshold because it is roughly equivalent to 3 standard deviations from the mean for normally distributed data, but is robust to outliers in the data. The low threshold was empirically chosen to be 3.33% of the high threshold. If more than half the window was removed for a channel, we removed the rest of that window for that channel as well. In addition, any windows where the standard deviation of the channel is less than 0.01 were also removed. Using this approach, 13±3.5% of the data/mouse (n=28 for our model training) were excluded from this analysis. This conservative strategy optimized the potential of our learning model to discover a network that was uniquely related to appetitive social emotional brain states.

Determination of LFP oscillatory power and cross-area synchrony and granger coherence—LFPs were averaged across wires within region to yield a composite LFP measure. Signal processing was performed using Matlab (The MathWorks, Inc., Natick, MA). For LFP power, a sliding Fourier transform with Hamming window was applied to the averaged LFP signal using a 1 second window and a 1 second step. Frequencies were analyzed with a resolution of 1Hz. LFP cross-structural coherence was calculated from the pairs of averaged LFPs using magnitude-squared coherence

$$C_{AB}(f) = \frac{|Psd_{AB}(f)|^2}{Psd_{AA}(f)Psd_{BB}(f)}$$

where coherence is a function of the power spectral densities of A and B, and their cross-spectral densities.

The spectral Granger causality (Geweke, 1982) features were calculated using the *Multivariate Granger Causality* (MVGC) MATLAB toolbox (Barnett and Seth, 2014). The non-stationary data required a highpass, so a highpass Butterworth filter with a stopband at

1Hz and a passband starting at 4Hz was applied to the data. Granger causality values for each window were calculated using a 20-order AR model via the *GCCA_tsdata_to_smvgc* function of the MVGC toolbox. Granger causality values were calculated for all integer frequency values within the desired range for all directed pairs of brain regions in the dataset.

For calculating electome network using, the exponential of all Granger causality values was used, which gives a ratio of total power to 'unexplained' power. Since the original formulation involves logarithms, it hinders the addibility of the features, so the exponential is suitable for inclusion in the electome model. Specifically,

$$\exp(f_{Y \to X}(\lambda)) = \frac{|S_{XX}(\lambda)|}{|S_{XX}(\lambda) - H_{XY}(\lambda)\Sigma_{Y|X}H_{XY}(\lambda)^*|}$$

where $f_{Y \to X}(\lambda)$ represents Granger causality at frequency λ from region Y to region X, $S_{XX}(\lambda)$ represent the spectral power in region X at frequency λ, and $H_{XY}(\lambda)\Sigma_{Y|X}H_{XY}(\lambda)^*$ represents the component of that power that is predicted by region Y. We capped values for this ratio at 10 to prevent any non-physiological signal from dominating the electome factors we learned using dCSFA-NMF.

Discriminative Cross-Spectral Factor Analysis – Nonnegative Matrix Factorization—To apply our Supervised Cross-Spectral Factor Analysis – Nonnegative Matrix Factorization (CSFA-NMF) model, which fully described elsewhere (Talbot *et al.*, 2020), we consider each window of data to be an independent stationary measurement. This implies that the relevant dynamics happens at the scale of windows, so the extracted electome scores are all that is needed for later analysis. In this work, we choose a 1 second window because this balanced fine-grained behavior with enough length of signal to estimate the relevant LFP features. Prior work has shown relative robustness to windows between .5s to 5s in similar methods (Ulrich et al., 2015), so we expect similar results for similar window lengths; however, 5s here would not be able to capture the short-term scale of behavior necessary for this analysis.

For each window of data, we have the generated features, consisting of spectral power features, coherence features, and exponential granger features, totaling P distinct features per window. Using the subscript n to denote window and state that there are N total windows. We describe the preprocessed data as $x_n \in \mathbb{R}_+^P$ (the P-dimensional non-negative domain) and the observed behavioral label as $y_n \in \{0,1\}$, where the binary indicates a social or non-social behavioral label. To briefly described this model, we set up an objective function to learn the K different electome factors,

$$\min_{W,d,\phi} \sum_{n=1}^{N} \|x_n - W f(x_n;\phi)\|_2^2 + \lambda \left\| y_n - d^T f(x_n;\phi) \right\|_2^2,$$

where each electome is described by a column in $W \in \mathbb{R}_+^{P \times K}$ (e.g., $W = [w_1, \ldots, w_K]$), the electome factor scores are given by the multi-output function $f(x_n; \phi): \mathbb{R}_+^p \rightarrow \mathbb{R}_+^K$, and the relationship between the electome factor scores and the behavioral labels is given by $d \in \mathbb{R}^p$. The relative importance of reconstructing the observed data and the importance of the predictive task were balanced by choosing the hyperparameter λ. This represents a novel method to fit an NMF model using supervised autoencoders and requires the user to choose a parametrization for $f(x_n; \phi)$. In our method, this is simply set to an affine function following by a non-linearity, $f(x_n; \phi) = softplus(Ax + b)$, where the parameters of the function are $\phi = \{A, b\}$ and the softplus means an element-wise operation of the operation $\text{softplus}(a) = \log(1 + \exp(a))$, which maps a real number to the non-negative space. This function can vary in complexity to allow greater model complexity, but we found that this function was sufficient in practice. Because this objective function follows a supervised autoencoder structure, a common deep learning structure, we are able to implement this technique in Tensorflow (Abadi et al., 2016) using the ADAM algorithm for learning (Diederik and Ba, 2014).

A benefit of using this structure for learning is that performing statistical inference from new data is fast and straightforward. In factor models, one typically has to set up an optimization algorithm to find the maximum a posteriori estimate. However, in our supervised CSFA-NMF framework, we can calculate the electome scores on new data simply by calling the function $f(x_n; \phi)$, allowing easy portability and facilitating future real time applications.

Hyper-parameter Selection—The proposed CSFA-NMF procedure requires us to choose several different settings in the algorithm, which was done with a cross-validation procedure where complete mice from the training set were left out. The hold-out mice, as described in the manuscript, were *not* used for hyperparameter selection and represent a true blind test set. Specifically, we must choose the number of electome factors K, the importance of the supervised task λ, the relative importance of the power features, coherence features, and exponential Granger features, and the parameterization of the mapping function $f(x_n; \phi)$.

We had dual goals in our analysis: reconstructing the original data well, which is to say that the learned *Electomes* actually describe the neural measurements well and predict the behavioral task well. The reconstruction error was evaluated by the Mean Squared Error on the validation mice, and the performance on the behavioral task was evaluated by the mean Area Under the Curve (mean AUC) on the validation mice. Greater emphasis was placed on the behavioral task, so for each candidate number of electome networks K, we used the cross-validation procedure to choose the settings that maximized the mean AUC. After that, an elbow analysis was used to choose the number of electome networks K, which is to mean we choose the K after which minimal gains in explaining the observed data was observed.

Two-chamber social Interaction test—Social preference was measured using a two-chamber assay in which animals explored a novel object or a novel mouse. The apparatus was a rectangular arena (61cm × 42.5cm × 22cm, L×W×H) constructed from clear plexiglass with a clear plexiglass wall dividing the arena into two equal chambers with

an opening in the middle allowing free access between both chambers. The floor of the arena was constructed using a one-way mirror that allowed for video recording from beneath in order to avoid obstruction from electrophysiological recording equipment. Plastic, circular holding cages (8.3cm diameter and 12cm tall) were centered in each of the two chambers and were used to house either a novel object or sex- and age-matched C3H target mouse. The arena was evenly lit with indirect white light (~125 lux). Test mice were handled and habituated to the social preference chambers and empty holding cages for a least three days prior to testing. Subsequently, mice underwent ten separate social preference test sessions, with at least one day off in between sessions, in which the test mice were allowed to freely explore the arena for ten minutes; the holding cages contained either a novel object or novel C3H target mouse. The side of the chamber holding the object/mouse was determined pseudorandomly, such that the object/mouse would not be placed in the same chamber on more than two consecutive sessions in order to prevent side biases and to distinguish target-specific effects from location-specific effects. Plastic toys and glass objects were used as novel objects with the object being between 3–5cm in all directions. Video data was tracked using Bonsai Visual Reactive Programming software and the time spent in the proximity (~5cm) of preference scores.

The social preference for each session was defined as:

$$\frac{\overline{InteractinTime_S} - \overline{InteractionTime_O}}{\overline{InteractionTime_S} + \overline{InteractionTime_O}}$$

where InteractionTime_S is the total time spent proximal to the other mouse, and InteractionTime_O is the total time spent proximal to the object.

Free Object/Social Interaction Test—The Free Object/Social Interaction Test (FOSIT) allowed for free exploration of either novel objects or novel sex-matched conspecific mice during a single session. Plastic and glass objects were used similarly to social preference testing. The test was run in a clear arena (35cm × 31cm) lit using indirect white light (125 lux). The test mouse was placed into the arena that contained either a novel object or a target mouse and allowed free exploration (i.e., the objects/mice were not kept in holding cages as in the social interaction test) for five minutes. Following this five-min trial, the test mouse was placed into a new, identical arena that contained either a novel object or novel mouse for another five minutes. The order of object/mouse trials was determined pseudorandomly, such that the test mouse would not see a novel object or mouse for more than two consecutive sessions in order to prevent habituation to the stimulus type. Additionally, in order to control for the location of the target mouse, the novel object was pseudorandomly placed in one of four quadrants of the area such that each subsequent object placement was in a different quadrant from the previous object trial and that each quadrant was used at least once. A total of ten trials were run so that each test mouse was able to interact with five novel objects and five novel mice over the course of the ~50-min session. The amount of time interacting with the objects and target mice was hand-scored by experienced raters. For social trials, interactions were distinguished based on physical

engagement (i.e., reciprocal interaction, test mouse investigating the target mouse, and target mouse investigating the test mouse).

Spike-*Electome Factor* activity correlation—Data acquired during the fifth session of the two-chamber social interaction test were used for this analysis. Cellular firing activity was averaged within one-sec non-overlapping windows for the ten-min recording period. The social firing preference of each cell was defined as:

$$\frac{\overline{FR_S} - \overline{FR_O}}{\overline{FR_S} + \overline{FR_O}}$$

where FR_S is the neuronal firing rates observed during social interactions and FR_O is the neuronal firing rates observed during interactions with an object. A rank-sum test of all one-sec observations was used to determine if a cell signaled social vs. object interactions. We used a spearman rank correlation to quantify the relationship between cellular firing during the ten-minute sessions and *Electome Factor* activity. We performed 1000 permutations for which Electome activity time bins were randomly shuffled within the social and object conditions. We then calculated the spearman rank correlation for each permutation. A cell was deemed to be positively correlated with the *Electome* network if it exhibited a spearman Rho above the 97.5% of the permutated distribution, and negatively correlated if it was below the 2.5%.

Sucrose consumption—Neural responses to sucrose or water delivery were measured in a rectangular chamber (30cm × 19cm × 28cm) constructed from black plastic Legos. Two nose poke holes, spaced 6.5cm apart along one of the long walls, detected nose pokes via IR beam breakage and delivered 10μL of fluid from a 27-gauge syringe situated within the hole; a five-sec timeout followed each fluid delivery in which subsequent nose pokes were not rewarded. Mice implanted with electrodes were habituated to the fluid drinking apparatus for two days prior to electrophysiological recordings. During habituation, singly housed mice had *ad libitum* access to food and used the nose poke holes for access to water. Subsequently, electrophysiological recordings were collected during two fluid-drinking sessions: one for sucrose and one for water. Specifically, two hours into the dark-cycle, mice were recorded for 1.5 hours while poking for administration of a 2% sucrose solution from both poke holes. Immediately following sucrose administration, the sucrose was switched out and water was delivered through the poke holes for an additional two hours. Timestamps for each nose poke were synchronized and stored alongside electrophysiological data.

Elevated plus maze test—The elevated plus maze (EPM) has been previously described. Briefly, the EPM consists of four cross-shaped arms (30.5cm length× 30.5cm width, at 91.4cm height from floor) and a 5cm×5cm central region. Two ‘closed’ arms are surrounded on three sides by walls of 16.5cm height and the other two ‘open’ arms are surrounded by a short piece of tape approximately 1 mm in height. Mice were habituated to the behavioral room for two hours, 24 hours before testing. Following a one-hr habituation period on the test day mice were placed in the center region of the elevated plus maze facing a closed arm.

Neural recordings were obtained for ten minutes, and the location of the mice was captured using video recordings. All EPM testing was performed at 50 lux.

Chronic social defeat stress—Male mice implanted with electrodes underwent 10 days of cSDS as previously described (Berton *et al.*, 2006; Hultman *et al.*, 2016; Hultman *et al.*, 2018; Krishnan *et al.*, 2007). Specifically, male retired-breeder CD1 (Charles River) mice were used as resident aggressors for the social defeat and were singly-housed prior to the experiments. C57 mice were then randomly assigned to control or defeat groups such that no entire cage was assigned to the same group. All C57 mice were singly housed prior to being subjected to cSDS. Particularly aggressive CD1s, as defined by demonstrating at least one successful act of aggression toward an intruder C57 male within 60 secs, were selected for use for cSDS. Intruder male C57 mice were introduced to the cage of a novel CD1 aggressor for 5 min daily, and then housed adjacent to the same aggressor for 24 hrs. During this time, mice were separated by a transparent and porous plexiglass barrier to enable constant sensory exposure.

During bouts of exposure to the CD1 mice, hallmark behavioral signs of subordination stress were observed including escape, submissive postures (i.e., defensive upright and supine) and freezing. Following the last 24-hr exposure to a CD1 aggressor mouse, all C57s were housed individually. Mice that exhibited significant injuries during social defeat stress were removed from post-stress analysis (Hultman *et al.*, 2016). These methods are identical to those previously described(Hultman *et al.*, 2018).

Neural recordings during averse social interaction.—We utilized the previously described forced interaction test (Hultman *et al.*, 2016; Hultman *et al.*, 2018; Kumar *et al.*, 2014). Briefly, C57 mice were placed in a 3.25” × 7” Plexiglas cylinder. Following a 5-min neurophysiological recording period, a CD1 aggressor mouse was introduced to the cage outside of the cylinder (18” high walls surround the outer cage to prevent escape and a lid is place over the inner chamber to prevent the aggressor from climbing in). Neurophysiological data were then recorded for five additional minutes. Neural data acquired during the full 10-minute recording period was utilized for subsequent analysis.

Choice social interaction test (Post Chronic stress)—Male mice were placed within a novel arena (46cm × 46cm) with a small cage located at one end, and each mouse’s movement was monitored for 150 seconds. Mice were then removed from the testing chamber, and reintroduced 30 seconds later after a non-aggressive male CD1 mouse was placed in the small cage. The time C57 mice spent in the interaction zone was quantified using Ethovision XT 7.1 software (Noldus Information Technology, Wageningen, Netherlands). The interaction ratio was calculated as (Interaction time when CD1 was present)/(Interaction time when CD1 was absent) (Hultman *et al.*, 2016; Kumar *et al.*, 2014). These methods are identical to those previously described(Hultman *et al.*, 2018). Notably, only mice that showed social avoidance were utilized for further network analysis.

Optogenetic manipulation of the prelimbic circuit elements—Ten-week only mice were anesthetized with 1.% isoflurane and placed in a stereotaxic device. A 33-gauge Hamilton syringe was used to infuse 0.5 μl of AAV2-CaMKlla-hChR2 (E123A)-EYFP

vector at a rate of 0.1 μl/min, bilaterally, into prelimbic cortex (1.8mm AP, ±0.5mm ML, 2.5mm DV from the skull) and the syringe was left in place for ten minutes following the injection. In order to mitigate pain and inflammation related to the procedure, all animals received carprofen (5 mg/kg, s.c.) injections once prior to surgery and then once every 24 hours for three days following viral injections.

Eight weeks after viral surgeries, mice were anesthetized again, and recording electrodes were implanted as described above. A fiberoptic cannula was built into the nucleus accumbens or amygdala bundle (Hultman *et al.*, 2016; Kumar et al., 2013). The tip of the 100μm diameter fiberoptic (Doric Lenses) was situated 400μm above the tip of the recording microwires in the core of the accumbens or amygdala. *In vivo* recordings and stimulations were conducted after 5–6 weeks of recovery. For Nac stimulations, we delivered light stimulation at 1–7mW bilaterally (473nm wavelength, LaserGlow, LRS-0473-GFM-00100-05; 589nm wavelength, LaserGlow, LRS-0589-GFF-00100-05), and the laser output was verified using a Power meter (Thorlabs, PM100D). We first stimulated a group of mice (n=3) at 7mW to mirror a prior study (Murugan *et al.*, 2017). All these animals exhibited seizures (observed behaviorally and confirmed using LFP recording). Repeat stimulation several days later at 1.5mW also induced seizures in these mice. We tested two additional stimulation naïve mice at 1.5mW, and one of these mice exhibited seizures as well. Thus, all experiments presented in the main manuscript were performed in mice with no prior stimulation. Mice were stimulated at 1mW bilaterally. Several mice showed ictal activity in the cortical channels at the onset of stimulation. Ictal activity was accompanied by immobility, backwards walking, and grooming in these mice. Ictal activity (restricted to the cortical channels) and behavioral responses subsided spontaneously usually within 5–10 seconds of stimulation onset, at which time mice became behaviorally activated and showed increase social investigation. For amygdala stimulation experiments, we used 1mW at 10Hz.

Recordings were performed in six blocks during a single session. In each block mice were exposed to a new object and a new mouse. Each block lasted 7.5 minutes, during which mice were stimulated with blue light for 2.5 minutes and yellow light for 2.5 minutes, with 1.25 minutes between each stimulation period. The order of blue vs. yellow light stimulation and social vs. object exposure was pseudorandomized for each mouse, such that animals never experienced the same color light stimulation first for all three social or object blocks. Additionally, the order of the social vs. object exposures were pseudorandomized for each block such that mice never experienced all three social or object exposures first.

For nucleus accumbens stimulation studies, two mice exhibited global LFP responses to stimulation and were removed from further analysis. One mouse did not show any LFP responses to stimulation in nucleus accumbens and was thus removed from further analysis. For amygdala studies, none of the mice exhibited local LFP responses in amygdala during stimulation. This was consistent with the weaker PFC projections observed in amygdala C57BL6/J mice compared to nucleus accumbens (Oh et al., 2014). Viral expression was confirmed histologically following experiments (see supplemental Figures S4G and S4H).

For analysis, behavior was hand-scored from video recordings as described above by experienced raters blind to laser conditions (blue vs. yellow). We combined behavioral and neurophysiological measures for all periods in which the subject mouse was engaged with the other mouse (unilateral or bilateral). *EN-Social* activity observed during these social encounters was normalized to the activity observed during object encounters and compared across blue and yellow light stimulation trials. Periods in which the experimental partner mouse was unilateral engaged with the subject mouse were excluded.

QUANTIFICATION AND STATISTICAL ANALYSIS

Electome Model Fitting—The statistical analyses for the Electome model were performed using Python 3.6 and Tensorflow version 1.09. We used machine learning to define a single relevant electome. The total number of Electomes was chosen to minimize the reconstruction loss with the minimal number of factors as defined previously. The reconstruction loss was weighted such that each mouse and each condition were weighted equally. The supervision loss weighting was determined as the amount of entropy contained in the binary variable of social vs object score. The supervision strength was started at a low value and gradually annealed to the final value.

While multiple Electomes networks were learned from the training data, only one electome network's activity was predictive of the social activity. Therefore, when applied to the test set, only this electome was evaluated statistically, so no multiple comparisons corrections were required (a major advantage of such a factor model formulation). The predictive ability for each mouse was quantified using the area under the curve using the network strengths.

Validation Testing—For validation testing, we projected LFP data recorded from new mice and/or new paradigms into our initial learned *Electome* network model. We then performed direct comparison across conditions (e.g., behavioral conditions, genotypes, etc.) using the median *Electome* network activity score for each condition per mouse. Activity scores were compared using non-parametric statistics, or parametric statistics a Box–Cox transformation was applied the raw data. To further enable evaluation of the robustness of our findings, the decoding strength (area under the curve of the receiver operating characteristic, which takes into account the activity scores for all of the transformed time windows) was also provided in the main text alongside the statistical results obtained through direct comparisons of the median activity scores. For optogenetic validation studies we normalized network activity observed social interactions to the network activity observed during object interactions for each stimulation type (blue vs. yellow). We utilized this strategy because the 10 Hz signal induced in cortex and striatum had the potential to diminish our detection of *EN-Social* (which reflected organized LFP patterns at frequencies including 10Hz). A longer-term solution will require the implementation of NMF-based statistical approaches that reconstruct *EN-Social* activity from data sets in which a subset of areas are excluded (see our prior model based on gaussian processes as an example) (Gallagher et al., 2017).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to thank Vann Bennett for contributing the *Ank2* mutant mice; Staci Bilbo, Cagla Eroglu, and Carina Block for comments on this work; Ilana Witten and Malavika Murugan for technical guidance, Derek Southwell and Timothy Nyangacha for technical support. This work was supported by WM Keck Foundation grant to KD and Fan Wang; NIH grants R01MH120158 to KD, R21MH104316 to KD and YHJ, R01ES025549 to KD, SB, and CE, 1R01EB026937 to DEC and KD, and 1R01MH125430 to DEC and KD. A special thanks to Freeman Hrabowski, Robert and Jane Meyerhoff, and the Meyerhoff Scholarship Program.

References

Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, et al. (2016). TensorFlow: A system for large-scale machine learning. arXiv.

Allsop SA, Wichmann R, Mills F, Burgos-Robles A, Chang CJ, Felix-Ortiz AC, Vienne A, Beyeler A, Izadmehr EM, Glober G, et al. (2018). Corticoamygdala Transfer of Socially Derived Information Gates Observational Learning. Cell 173, 1329–1342 e1318. 10.1016/j.cell.2018.04.004. [PubMed: 29731170]

Barnett L, and Seth AK (2014). The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. Journal of neuroscience methods 223, 50–68. 10.1016/j.jneumeth.2013.10.018. [PubMed: 24200508]

Berton O, McClung CA, Dileone RJ, Krishnan V, Renthal W, Russo SJ, Graham D, Tsankova NM, Bolanos CA, Rios M, et al. (2006). Essential role of BDNF in the mesolimbic dopamine pathway in social defeat stress. Science 311, 864–868. [PubMed: 16469931]

Block CL, Eroglu O, Mague SD, C S, Blount C, Malacon KE, Beben KA, Ndubuizu N, Talbot A, Gallagher N, et al. (2020). Prenatal Environmental Stressors Impair Postnatal Microglia Function and Adult Behavior in Males. BioRxiv. 10.1101/2020.10.15.336669

Carlson D, David LK, Gallagher NM, Vu MT, Shirley M, Hultman R, Wang J, Burrus C, McClung CA, Kumar S, et al. (2017). Dynamically Timed Stimulation of Corticolimbic Circuitry Activates a Stress-Compensatory Pathway. Biol Psychiatry 82, 904–913. 10.1016/j.biopsych.2017.06.008. [PubMed: 28728677]

Carlson D, Schaich Borg J, Dzirasa K, and Carin L (2014). On the relations of LFPs & Neural Spike Trains Advances in Neural Information Processing Systems 27.

Crawford H, Moss J, Oliver C, Elliott N, Anderson GM, and McCleery JP (2016). Visual preference for social stimuli in individuals with autism or neurodevelopmental disorders: An eye-tracking study. Molecular autism 7, 24. 10.1186/s13229-016-0084-x. [PubMed: 27054022]

Dawson G, Jones EJ, Merkle K, Venema K, Lowy R, Faja S, Kamara D, Murias M, Greenson J, Winter J, et al. (2012). Early behavioral intervention is associated with normalized brain activity in young children with autism. J Am Acad Child Adolesc Psychiatry 51, 1150–1159. 10.1016/j.jaac.2012.08.018. [PubMed: 23101741]

Dawson G, Toth K, Abbott R, Osterling J, Munson J, Estes A, and Liaw J (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. Developmental psychology 40, 271–283. 10.1037/0012-1649.40.2.271. [PubMed: 14979766]

Diederik KP, and Ba J (2014). Adam: A Method for Stochastic Optimization. arXiV.

Felix-Ortiz AC, Burgos-Robles A, Bhagat ND, Leppla CA, and Tye KM (2016). Bidirectional modulation of anxiety-related and social behaviors by amygdala projections to the medial prefrontal cortex. Neuroscience 321, 197–209. 10.1016/j.neuroscience.2015.07.041. [PubMed: 26204817]

Ferguson BR, and Gao WJ (2018). Thalamic Control of Cognition and Social Behavior Via Regulation of Gamma-Aminobutyric Acidergic Signaling and Excitation/Inhibition Balance in the Medial Prefrontal Cortex. Biol Psychiatry 83, 657–669. 10.1016/j.biopsych.2017.11.033. [PubMed: 29373121]

Fraiman D, Saunier G, Martins EF, and Vargas CD (2014). Biological motion coding in the brain: analysis of visually driven EEG functional networks. PLoS One 9, e84612. 10.1371/journal.pone.0084612. [PubMed: 24454734]

Gallagher NM, Ulrich K, Dzirasa K, Carin L, and Carlson DE (2017). Cross-Spectral Factor Analysis Advances in Neural Information Processing Systems 30.

Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, Mahajan M, Manaa D, Pawitan Y, Reichert J, et al. (2014). Most genetic risk for autism resides with common variation. Nat Genet 46, 881–885. 10.1038/ng.3039. [PubMed: 25038753]

Geweke J (1982). Measurement of Linear Dependence and Feedback between Multiple Time Series. Journal of the American Statistical Association 77, 304–313.

Gunaydin LA, Grosenick L, Finkelstein JC, Kauvar IV, Fenno LE, Adhikari A, Lammel S, Mirzabekov JJ, Airan RD, Zalocusky KA, et al. (2014). Natural neural projection dynamics underlying social behavior. Cell 157, 1535–1551. 10.1016/j.cell.2014.05.017. [PubMed: 24949967]

Hultman R, Mague SD, Li Q, Katz BM, Michel N, Lin L, Wang J, David LK, Blount C, Chandy R, et al. (2016). Dysregulation of Prefrontal Cortex-Mediated Slow-Evolving Limbic Dynamics Drives Stress-Induced Emotional Pathology. Neuron 91, 439–452. 10.1016/j.neuron.2016.05.038. [PubMed: 27346529]

Hultman R, Ulrich K, Sachs BD, Blount C, Carlson DE, Ndubuizu N, Bagot RC, Parise EM, Vu MT, Gallagher NM, et al. (2018). Brain-wide Electrical Spatiotemporal Dynamics Encode Depression Vulnerability. Cell 173, 166–180 e114. 10.1016/j.cell.2018.02.012. [PubMed: 29502969]

Kim CK, Yang SJ, Pichamoorthy N, Young NP, Kauvar I, Jennings JH, Lerner TN, Berndt A, Lee SY, Ramakrishnan C, et al. (2016). Simultaneous fast measurement of circuit dynamics at multiple sites across the mammalian brain. Nature methods 13, 325–328. 10.1038/nmeth.3770. [PubMed: 26878381]

Klin A, Shultz S, and Jones W (2015). Social visual engagement in infants and toddlers with autism: Early developmental transitions and a model of pathogenesis. Neurosci Biobehav Rev 50, 189–203. 10.1016/j.neubiorev.2014.10.006. [PubMed: 25445180]

Krishnan V, Han MH, Graham DL, Berton O, Renthal W, Russo SJ, Laplant Q, Graham A, Lutter M, Lagace DC, et al. (2007). Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions. Cell 131, 391–404. [PubMed: 17956738]

Kumar S, Black SJ, Hultman R, Szabo ST, DeMaio KD, Du J, Katz BM, Feng G, Covington HE 3rd, and Dzirasa K (2013). Cortical Control of Affective Networks. J Neurosci 33, 1116–1129. [PubMed: 23325249]

Kumar S, Hultman R, Hughes D, Michel N, Katz BM, and Dzirasa K (2014). Prefrontal cortex reactivity underlies trait vulnerability to chronic social defeat stress. Nat Commun 5, 4537. 10.1038/ncomms5537. [PubMed: 25072279]

Marcinkiewcz CA, Mazzone CM, D'Agostino G, Halladay LR, Hardaway JA, DiBerto JF, Navarro M, Burnham N, Cristiano C, Dorrier CE, et al. (2016). Serotonin engages an anxiety and fear-promoting circuit in the extended amygdala. Nature 537, 97–101. 10.1038/nature19318. [PubMed: 27556938]

Morrison I, Lloyd D, di Pellegrino G, and Roberts N (2004). Vicarious responses to pain in anterior cingulate cortex: is empathy a multisensory issue? Cogn Affect Behav Neurosci 4, 270–278. 10.3758/cabn.4.2.270. [PubMed: 15460933]

Moy SS, Nadler JJ, Young NB, Perez A, Holloway LP, Barbaro RP, Barbaro JR, Wilson LM, Threadgill DW, Lauder JM, et al. (2007). Mouse behavioral tasks relevant to autism: phenotypes of 10 inbred strains. Behav Brain Res 176, 4–20. [PubMed: 16971002]

Murugan M, Jang HJ, Park M, Miller EM, Cox J, Taliaferro JP, Parker NF, Bhave V, Hur H, Liang Y, et al. (2017). Combined Social and Spatial Coding in a Descending Projection from the Prefrontal Cortex. Cell 171, 1663–1677 e1616. 10.1016/j.cell.2017.11.002. [PubMed: 29224779]

Oh SW, Harris JA, Ng L, Winslow B, Cain N, Mihalas S, Wang Q, Lau C, Kuan L, Henry AM, et al. (2014). A mesoscale connectome of the mouse brain. Nature 508, 207–214. 10.1038/nature13186. [PubMed: 24695228]

Padilla-Coreano N, Bolkan SS, Pierce GM, Blackman DR, Hardin WD, Garcia-Garcia AL, Spellman TJ, and Gordon JA (2016). Direct Ventral Hippocampal-Prefrontal Input Is Required for Anxiety-Related Neural Activity and Behavior. Neuron 89, 857–866. 10.1016/j.neuron.2016.01.011. [PubMed: 26853301]

Padilla-Coreano N, Canetta S, Mikofsky RM, Alway E, Passecker J, Myroshnychenko MV, Garcia-Garcia AL, Warren R, Teboul E, Blackman DR, et al. (2019). Hippocampal-Prefrontal Theta Transmission Regulates Avoidance Behavior. Neuron 104, 601–610 e604. 10.1016/j.neuron.2019.08.006. [PubMed: 31521441]

Phillips ML, Robinson HA, and Pozzo-Miller L (2019). Ventral hippocampal projections to the medial prefrontal cortex regulate social memory. Elife 8. 10.7554/eLife.44182.

Rodgers RJ, Lee C, and Shepherd JK (1992). Effects of diazepam on behavioural and antinociceptive responses to the elevated plus-maze in male mice depend upon treatment regimen and prior maze experience. Psychopharmacology (Berl) 106, 102–110. 10.1007/bf02253596. [PubMed: 1738787]

Rodriguez E, George N, Lachaux JP, Martinerie J, Renault B, and Varela FJ (1999). Perception's shadow: long-distance synchronization of human brain activity. Nature 397, 430–433. [PubMed: 9989408]

Schaich Borg J, Srivastava S, Lin L, Heffner J, Dunson D, Dzirasa K, and de Lecea L (2017). Rat intersubjective decisions are encoded by frequency-specific oscillatory contexts. Brain Behav 7, e00710. 10.1002/brb3.710. [PubMed: 28638715]

Seidenbecher T, Laxmi TR, Stork O, and Pape HC (2003). Amygdalar and hippocampal theta rhythm synchronization during fear memory retrieval. Science 301, 846–850. [PubMed: 12907806]

SFARI-GENE (2020). https://gene.sfari.org/about-gene-scoring/.

Sokolov AA, Zeidman P, Erb M, Ryvlin P, Friston KJ, and Pavlova MA (2018). Structural and effective brain connectivity underlying biological motion detection. Proc Natl Acad Sci U S A 115, E12034–E12042. 10.1073/pnas.1812859115. [PubMed: 30514816]

Talbot A, Dunson D, Dzirasa K, and Carlson D (2020). Supervised Autoencoders Learn Robust Joint Factor Models ofNeural Activity. arxiv.

Ulrich K, Carlson D, Dzirasa K, and Carin L (2015). GP Kernels for Cross-Spectrum Analysis. Advances in Neural Information Processing Systems 28.

Vu MT, Adali T, Ba D, Buzsaki G, Carlson D, Heller K, Liston C, Rudin C, Sohal V, Widge AS, et al. (2018). A Shared Vision for Machine Learning in Neuroscience. J Neurosci. 10.1523/JNEUROSCI.0508-17.2018.

Vu MT, David LK, Thomas G, Vagwala M, Burrus C, Gallagher NM, Wang J, Blount C, hughes DN, Adamson E, et al. (2019). Brain-wide electrical spatiotemporal dynamics encode reward anticipation. BioRxiv.

Wang F, Zhu J, Zhu H, Zhang Q, Lin Z, and Hu H (2011). Bidirectional control of social hierarchy by synaptic efficacy in medial prefrontal cortex. Science 334, 693–697. 10.1126/science.1209951. [PubMed: 21960531]

Yang R, Walder-Christensen KK, Kim N, Wu D, Lorenzo DN, Badea A, Jiang YH, Yin HH, Wetsel WC, and Bennett V (2019). ANK2 autism mutation targeting giant ankyrin-B promotes axon branching and ectopic connectivity. Proc Natl Acad Sci U S A 116, 15262–15271. 10.1073/pnas.1904348116. [PubMed: 31285321]

Yizhar O, Fenno LE, Prigge M, Schneider F, Davidson TJ, O'Shea DJ, Sohal VS, Goshen I, Finkelstein J, Paz JT, et al. (2011). Neocortical excitation/inhibition balance in information processing and social dysfunction. Nature 477, 171–178. 10.1038/nature10360. [PubMed: 21796121]

Highlights

- Machine learning model discovers and integrates circuits into affective brain network
- Brain-wide network encodes rewarding social experience of individual mice
- Causal activation of network sub-circuits selectively induces social behavior
- Social brain network fails to encode individual behavior in a mouse model of autism

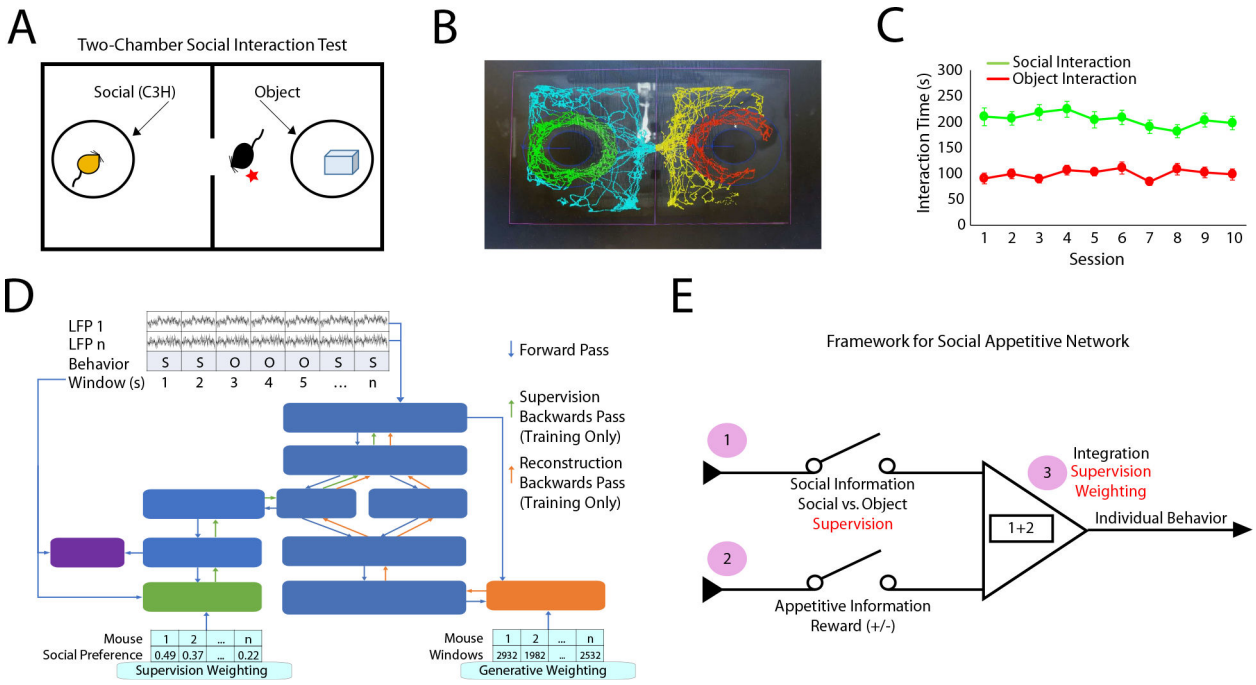

Figure 1: Approach to decode social appetitive behavior.

A) Schematic of the two-chamber social assay and **B)** automated scoring approach used to quantify social and object interaction. **C)** Mice exhibit stable interaction times across repeated sessions at the population level (n=36 mice). **D)** Schematic of machine learning model used to discover social-emotional brain state. **E)** Coding features that conceptually underlie a social-emotional brain state.

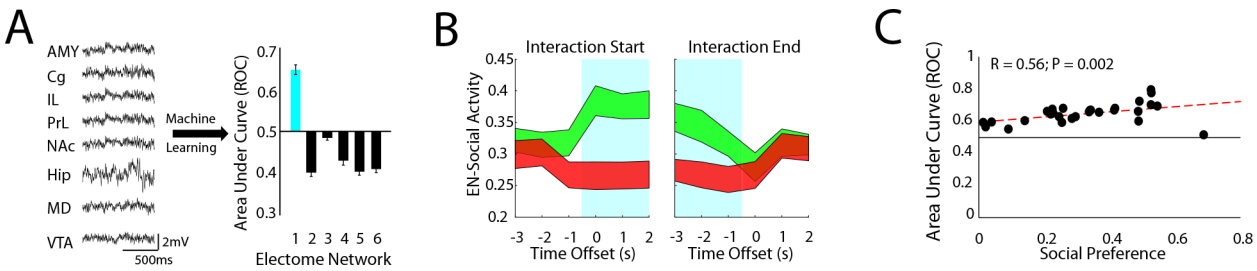

Figure 2: An *electome* network encodes a social-appetitive brain state.

A) Machine learning was used to discover six networks composed of multi-regional LFP activity (n=28 mice; AMY, Cg, IL, PrL, NAc, Hip, MD, and VTA). The supervised *electome* network (blue; *EN-Social*) showed the strongest classification of social vs. object interactions. **B)** *EN-Social* event-related activity. Blue highlights identify time windows subjected to supervision by class (social vs. object). Data shown as mean±95% C.I. **C)** Decoding accuracy of *EN-Social* activity within animal vs. social preference (P=0.002 using spearman correlation).

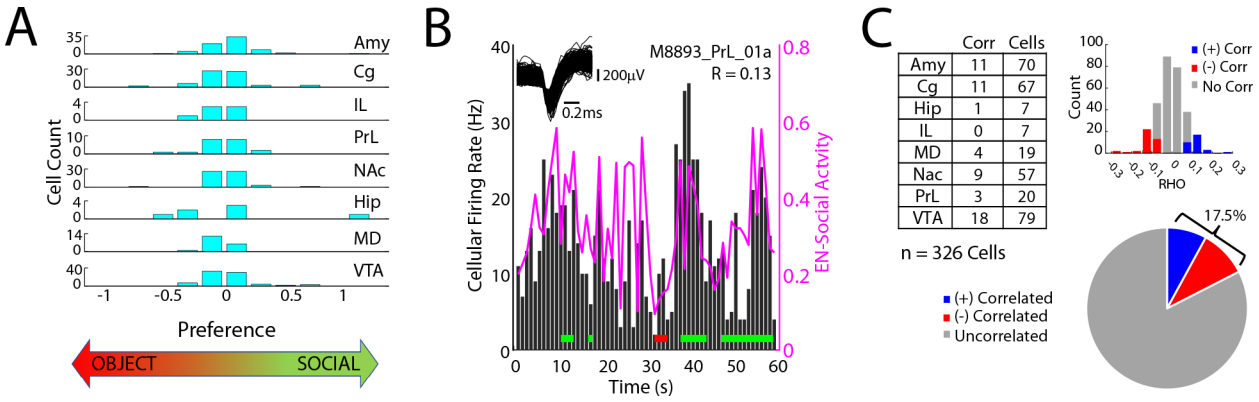

Figure 3: Social-appetitive *electome* network maps to cellular activity.

A) Cellular firing preference for object vs. social interactions during two-chamber assay (cellular activity analyzed from session #5). Significant differences were observed between the two conditions for 112/326 cells (P<0.05 using rank-sum test). **B)** Representative example of cell that showed activity correlated with *EN-Social*. Horizontal red and green lines signify object and social interactions, respectively**. C)** Cellular firing vs. *EN-Social* activity across the multi-regional population of cells (P<0.05 using permutation test; recorded from session #5 of two-chamber assay).

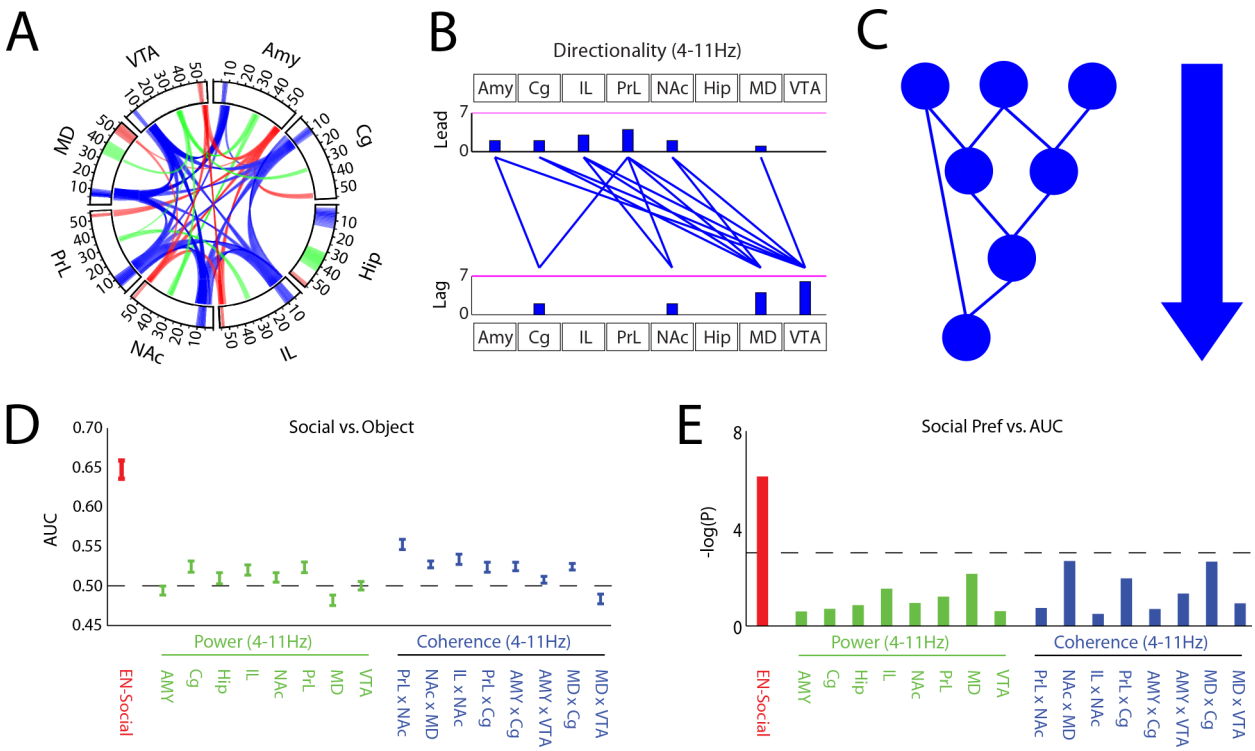

Figure 4: Circuit elements within *EN-Social* fail to encode individual behavior.

A) Power and synchrony measures that compose *EN-Social*. Brain areas and oscillatory frequency bands ranging from 1 to 56Hz are shown around the rim of the circle plot. Spectral power measures that contribute to the *electome* are depicted by the highlights around the rim, and cross spectral (i.e., synchrony) measures are depicted by the lines connecting the brain regions through the center of the circle (*electome* activity is shown at a relative spectral density threshold of 0.33, signifying the 85th percentile of retained features). **B)** Granger offset measures were used to quantify directionality within the *electome* network. Prominent directionality was observed across the theta (4–11Hz) frequency band (shown at a spectral density threshold of 0.33). Histograms quantify the number of lead and lagging circuit interactions for each brain region. **C)** Schematic of signal directionality within *EN-Social*. **D)** Decoding accuracy of *EN-Social* circuit elements for social vs. object interactions. **E)** Decoding accuracy of *EN-Social* circuit element activity within animal vs. social preference using spearman correlation. Threshold corresponds with P<0.05.

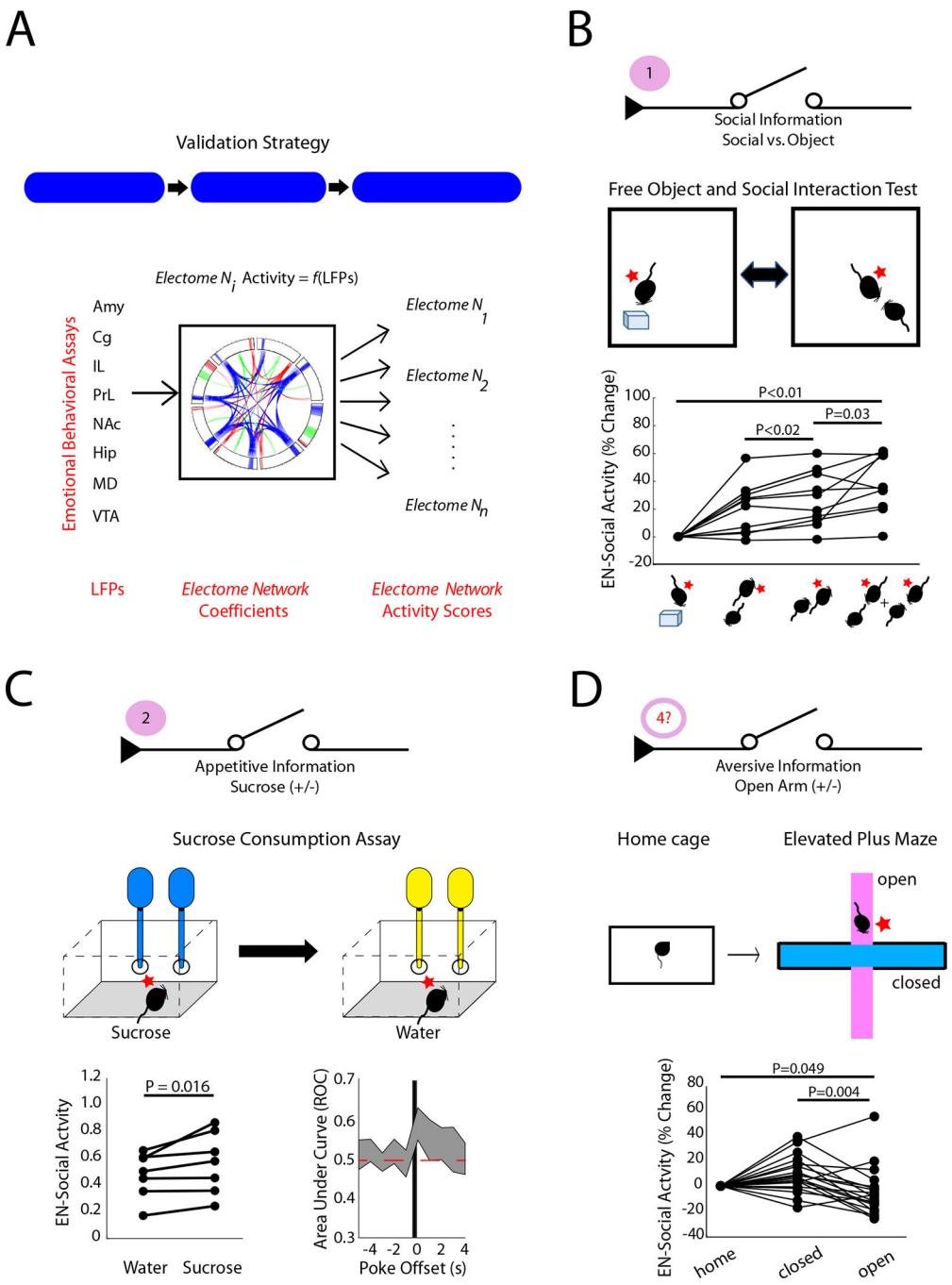

Figure 5: *Electome* network generalizes to encode social brain state and valence.

A) Strategy for validating *EN-Social*. **B)** Activity in the *EN-Social* network increased during distinct social appetitive brain states (n = 10 new mice; P<0.05 using Friedman's test, and post-hoc testing using sign-rank test with false discover rate correction). The position of the subject mouse is shown relative to an object or another experimental mouse on the bottom. **C)** *EN-Social* activity during water vs. sucrose consumption (left) and decoding accuracy vs. nose poke onset (right; n=7 new mice;). **D)** *EN-Social* activity during home cage and elevated plus maze recordings (n=19 mice, 7 of which were new to the study).

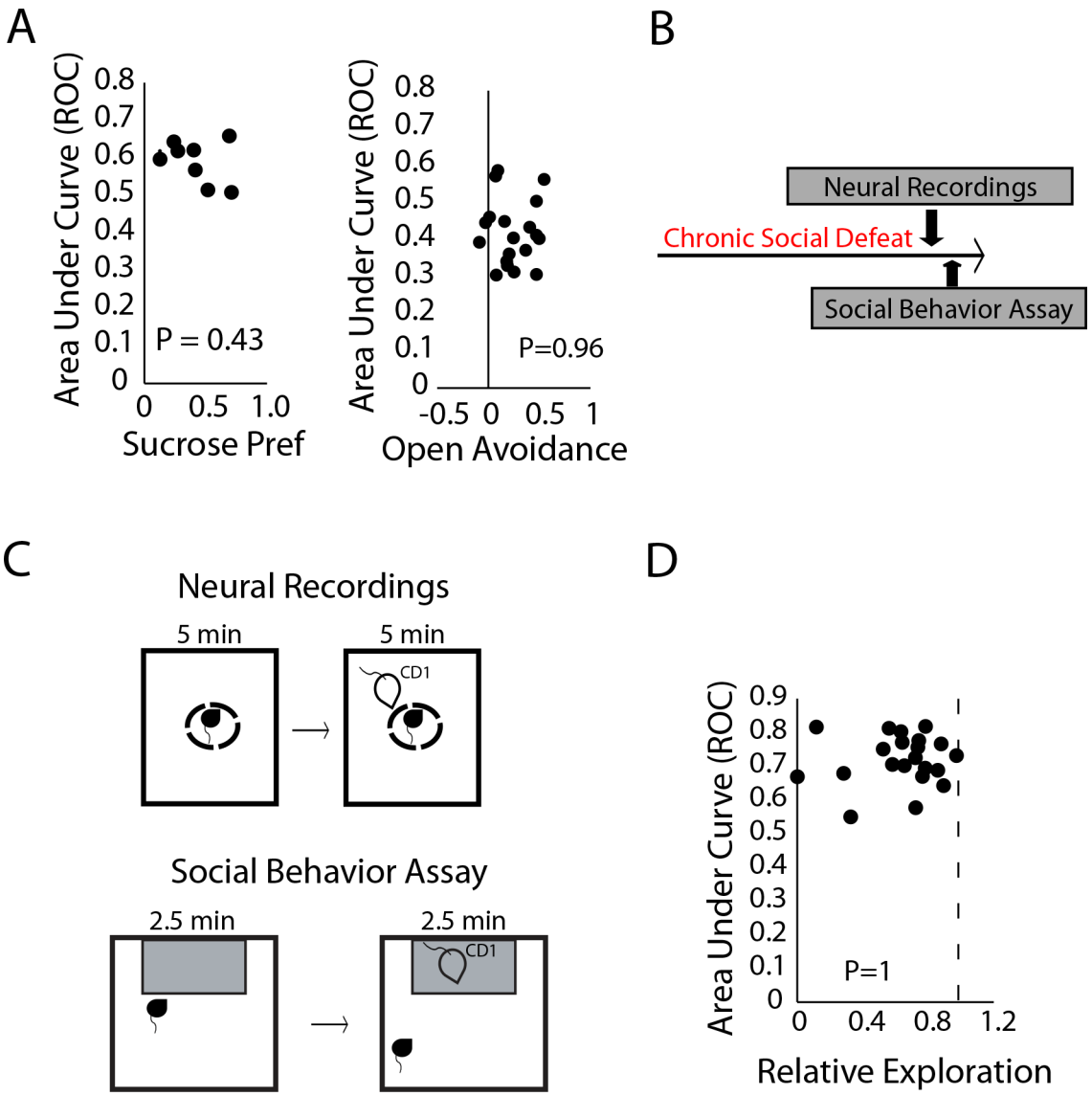

Figure 6: *EN-Social* fails to encode individual responses to non-social stimuli or social aversion.

A) *EN-Social* decoding for sucrose/water verses the sucrose preference of individual mice (left). *EN-Social* decoding for open arm/closed arm verses the open arm avoidance (i.e., closed and center arm preference) of individual mice (right). **B)** Timeline for chronic social defeat stress experiment. **C)** Protocol utilized to induce and assess neural activity (top) and behavior (bottom) during socially aversive conditions. **D)** Decoding accuracy of CD1 vs. empty area compared to social avoidance of CD1 mice in susceptible animals.

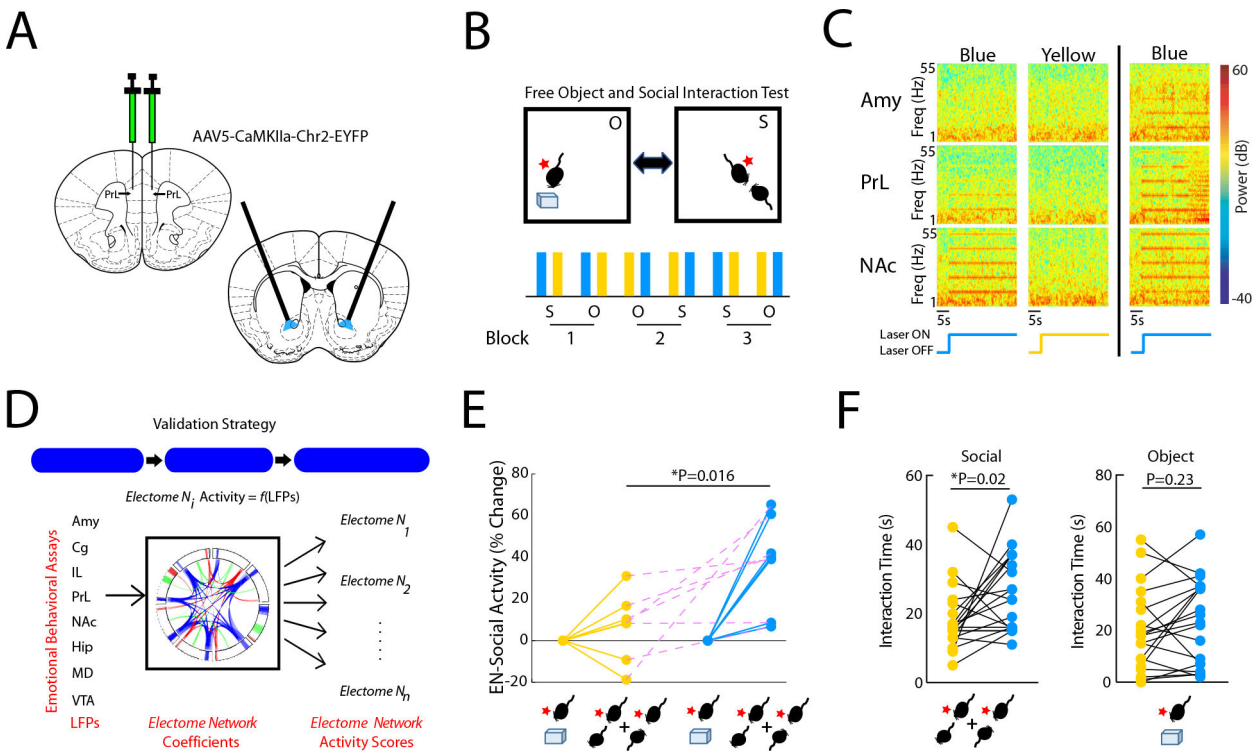

Figure 7: Causal activation of the prefrontal cortex to nucleus accumbens circuit element enhances *EN-Social* activity.

A) Strategy used to activate PL terminals in NAc. **B)** Experimental paradigm for FOSIT. **C)** Power spectral plots showing increased 10Hz oscillatory activity during blue light stimulation. Plots show representative spectral patterns from a mouse during blue (left) and yellow (middle) light stimulation trials included in analysis. Representative plots from mouse that showed increased 10Hz activity across all brain regions during blue light stimulation (right). **D)** Strategy used for *EN-Social* validation. **E)** *EN-Social* activity during blue light stimulation. Network activity was pooled across periods of social interaction by the subject mice and compared between the blue and yellow light stimulation periods. **F)** Social (left; $P<0.05$) and object interaction time (right; $P>0.05$) during blue and yellow light stimulation (all technical replicates are shown).

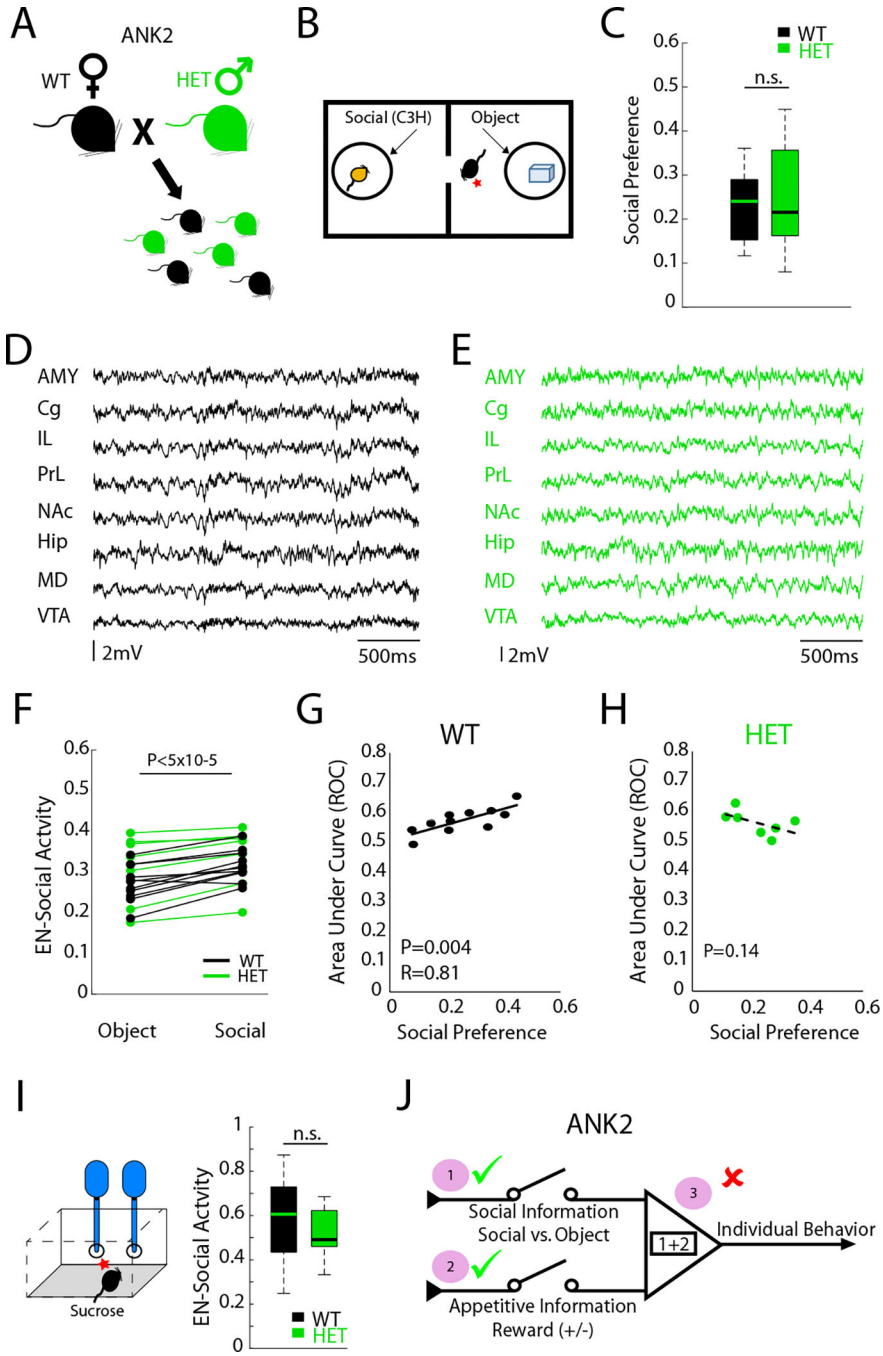

Figure 8: *Electome* network fails to encode individual social preference in a genetic model of autism spectrum disorder.

A-B) *Ank2* mice and their littermate controls were subjected to two-chamber social assay. **C)** Both *groups* showed preference for social interactions (P>0.05). **D-E)** Representative LFP activity in d) wild type and e) *Ank2* mice showing no seizure activity. **F)** *EN-Social* activity during social and object interactions (P<0.05 for conditions; P>0.05 for genotype effects). **G-H)** *EN-Social* activity vs. appetitive social behavior in G) wild-type mice (P<0.05) and H) *Ank2* mutants (P>0.05). **I)** Network activity during sucrose consumption

in wild-type mice and *Ank2* mutants (P>0.05). **J)** Summary of *EN-Social* function in *Ank2* mutants.

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-GFP	Thermofisher Scientific	Cat#A-11122; RRID: AB_221569
Bacterial and virus strains		
pAAV-CaMKIIa-hChr2(E123A)-EYFP (AAV5)	Lee et al., Nature (2010)	Addgene Plasmid: 26969-AAV5
Chemicals, peptides, and recombinant proteins		
Mounting Medium with DAPI	Abcam	CAT#ab104139
NeuroTrace 500/525 Gren Fluorescent Nissl Stain	ThermoFisher Scientific	CAT#N21480
Cresyl Violet acetate	Sigma-Aldrich	CAT#C5042
Deposited data		
Custom MATLAB scripts for dCSFA-NMF analysis	This paper	https://doi.org/10.5281/zenodo.6078304
Experimental models: Organisms/strains		
Mouse: C57BL6/J	The Jackson Laboratory	Jax: 000664
Mouse: C3H/HeJ	The Jackson Laboratory	Jax: 000659
Mouse: Ank2-P2580fs	Yang et al., 2019	
Mouse: CD1 IGS	Charles River Laboratories	Crl: 022
Software and algorithms		
Central Suite 7.0.5	Blackrock Neurotech	Version 7.0.5
Offline Sorter v3	Plexon	Version 3
MATLAB v 2013a	The Mathworks, Inc	Versions 2013a, 2015b, 2018b
Ethovision XT 7.1	Noldus	Version XT 7.1
Bonsai Visual Reactive Programming	Bonsai	Version 2.2
NMF Cross Frequency Spectral Analysis	This paper	https://zenodo.org/record/6078304#.Yg_IeS-B3fY
Other		
Omnetics Connectors	Omnetics Connector Corp	CAT#A8987-001
Optic Fibers	Doric Lenses	CAT#MFC_100/1.5-0.22_12mm_MF2.5_FLT; CAT#MFC_100/1.5-0.22_8mm_MF2.5_FLT
474nm Laser	LaserGlow Technologies	LRS-0473-GFM-00100-05
589nm Laser	LaserGlow Technologies	LRS-0589-GFF-00100-05