

1 **Amygdala Reward Neurons Form and Store Fear Extinction Memory**

2 Xiangyu Zhang¹, Joshua Kim¹ & Susumu Tonegawa^{1,2*}

3

4 ¹RIKEN–MIT Laboratory for Neural Circuit Genetics at the Picower Institute for Learning and
5 Memory, Department of Biology and Department of Brain and Cognitive Sciences,
6 Massachusetts Institute of Technology, Cambridge, MA 02139, USA

7 ²Howard Hughes Medical Institute at Massachusetts Institute of Technology, Cambridge, MA
8 02139, USA

9

10 *Correspondence: tonegawa@mit.edu

1 **Summary**

2 The ability to extinguish conditioned fear memory is critical for adaptive control of fear response,
3 and its impairment is a hallmark of emotional disorders like post-traumatic stress disorder (PTSD).
4 Fear extinction is thought to take place when animals form a new memory that suppresses the
5 original fear memory. However, little is known about the nature and the site of formation and
6 storage of the new extinction memory. Here, we demonstrate that a fear extinction memory engram
7 is formed and stored in a genetically distinct basolateral amygdala (BLA) neuronal population that
8 drive reward behaviors and antagonize the BLA's original fear neurons. The activation of the fear
9 extinction engram neurons and natural reward-responsive neurons overlap extensively in the BLA.
10 Furthermore, these two neuron subsets are mutually interchangeable in driving reward behaviors
11 and fear extinction behaviors. Thus, fear extinction memory is a newly formed reward memory.

12

13 **Keywords:** Fear extinction, memory, engram, reward, amygdala

1 **Introduction**

2 The inability to extinguish fear is a hallmark of many psychiatric disorders, such as PTSD
3 and generalized anxiety disorder (Shalev et al., 2017; Stein and Sareen, 2015). Following
4 Pavlovian fear conditioning (Pavlov and Anrep, 1927), repeated or prolonged presentations of the
5 conditioned stimulus (CS) without an expected aversive unconditioned stimulus (US) diminishes
6 the conditioned fear response, a phenomenon called fear extinction (Herry et al., 2010; Myers and
7 Davis, 2007; Quirk and Mueller, 2008). Fear extinction has been proposed to involve the formation
8 of a new memory in competition with the original fear memory (Bouton, 2004; Quirk and Mueller,
9 2008; Quirk et al., 2010). The amygdala is a key structure for fear memory (Davis, 1992; Duvarci
10 and Pare, 2014; Ehrlich et al., 2009; Maren and Fanselow, 1996), and is also involved in the fear
11 extinction memory (Amano et al., 2010; Grewe et al., 2017; Herry et al., 2008). However, it is
12 unknown whether the amygdala is the site for the storage of fear extinction memory, and if so,
13 which subset of amygdala neuron stores this memory.

14 Excitatory neurons in the mouse basolateral amygdala (BLA) respond to both positive and
15 negative valence stimuli (Beyeler et al., 2016; Davis and Whalen, 2001; Kim et al., 2016; Namburi
16 et al., 2015; Redondo et al., 2014), and more than 90% of these neurons are composed of two
17 genetically, functionally and anatomically distinct neuronal populations (Kim et al., 2016; Kim et
18 al., 2017). R-spondin-2-expressing (*Rspo2*⁺) neurons located in the anterior BLA respond to
19 negative valence stimuli and control negative behaviors and memories, whereas protein-
20 phosphatase-1-regulatory-inhibitor-subunit-1B-expressing (*Ppp1r1b*⁺) neurons located in the
21 posterior BLA (pBLA) respond to positive valence stimuli and control appetitive behaviors and

1 memories (Kim et al., 2016). Furthermore, these two neuronal populations antagonize each other
2 through feed-forward inhibitions mediated by local inhibitory interneurons (Kim et al., 2016).

3 In this study, we investigated a potential role of BLA *Ppp1r1b*⁺ neurons in fear extinction
4 using a contextual fear extinction paradigm. We found that fear extinction memory engram cells
5 are formed and stored within the BLA *Ppp1r1b*⁺ neuronal population, and these engram cells are
6 necessary for suppressing the original fear memory. Furthermore, fear extinction engram cells and
7 natural reward-responsive cells in the pBLA *Ppp1r1b*⁺ neurons are mutually interchangeable in
8 driving reward functions and fear extinction behaviors.

9

10 **Results**

11 **BLA *Ppp1r1b*⁺ neurons are activated during contextual fear extinction**

12 Fear extinction phenomena have been observed and studied in both cue-dependent fear
13 conditioning and context-dependent fear conditioning paradigms (Amano et al., 2010; Baldi and
14 Bucherelli, 2014; Herry et al., 2008; Trouche et al., 2013; Zushida et al., 2007). Since the BLA
15 plays a critical role in contextual fear conditioning (Calandreau et al., 2005; Goosens and Maren,
16 2001; Huff and Rudy, 2004; Redondo et al., 2014) and a substantial amount of information is
17 available regarding the excitatory neuronal subsets in BLA (see Introduction) (Kim et al., 2016),
18 we employed a contextual fear extinction paradigm in this study. On Day 1, the Extinction group
19 received contextual fear conditioning (CFC) in a box where the context served as the CS and three
20 rounds of footshocks served as the US. On Day 2, mice were returned to the conditioned box for
21 45 min in the absence of footshocks to receive contextual fear extinction training (FET). On Day
22 3, the mice were tested for 5 min extinction memory retrieval in the conditioned box and then

1 sacrificed (Figure 1A). Two control groups were set up as follows. The Non-Extinction group went
2 through the same protocol as the Extinction group, except that on Day 2 the mice stayed in their
3 home cage (HC) and didn't receive any extinction training (Figure 1A). The Non-Shock group
4 received the same protocol as the Extinction group, but received no footshocks on Day 1 (Figure
5 1A). The Extinction group and Non-Extinction group, but not the Non-Shock group, displayed
6 robust freezing behavior on Day 1 that increased as more footshocks were presented (Figure 1B).
7 On Day 3, the Extinction group froze much less than the Non-Extinction group (Figure 1C).

8 Double smFISH (single molecular Fluorescence *in situ* Hybridization) was performed to
9 detect the activity-dependent expression of the immediate early gene, *Fos*, in *Rspo2*⁺ and
10 *Ppp1r1b*⁺ cells (see Methods). The proportions of *Fos*⁺ cells of *Rspo2*⁺ cells and *Fos*⁺ cells of
11 *Ppp1r1b*⁺ cells were quantified across the anterior-posterior (A/P) axis for the aBLA and pBLA,
12 respectively (Kim et al., 2016). The proportion of *Fos*⁺/*Rspo2*⁺ was lower in the Extinction group
13 compared to the Non-Extinction group (Figures 1D, 1E and 1H). In contrast, the proportion of
14 *Fos*⁺/*Ppp1r1b*⁺ was higher in the Extinction group compared to the Non-Extinction group while
15 no difference was observed between the Non-Extinction and Non-Shock groups (Figures 1F, 1G
16 and 1I). When BLA *Ppp1r1b*⁺ neurons were optogenetically inhibited during extinction retrieval,
17 the neuronal activity of *Rspo2*⁺ cells (*Fos*⁺/*Rspo2*⁺) as well as the freezing level increased during
18 extinction retrieval (Figures S1). This result suggests that *Ppp1r1b*⁺ neurons suppress *Rspo2*⁺
19 neurons during fear extinction retrieval, which is consistent with the previously reported feed-
20 forward inhibition of BLA *Rspo2*⁺ cells by BLA *Ppp1r1b*⁺ cells (Kim et al., 2016).

21

22 **Dynamics of individual BLA neurons throughout CFC and fear extinction**

1 Next, we performed *in vivo* calcium imaging with a microendoscope to directly track
2 individual BLA neuronal activity during CFC followed by contextual fear extinction. The
3 genetically encoded calcium indicator GCaMP6f was expressed in BLA *Rspo2*⁺ and *Ppp1r1b*⁺
4 neurons by injecting adeno-associated virus 5 (AAV₅)-human synapsin (hsyn):double-floxed
5 inverse open reading frame (DIO)-*GCaMP6f* into the aBLA of *Rspo2*-Cre mice and the pBLA of
6 *Ppp1r1b*-Cre mice, respectively (Figures 2A-2C). The efficiency of Cre-loxP recombination in
7 *Rspo2*-Cre and *Ppp1r1b*-Cre mice lines are around 90% and 75%, respectively. (Kim et al., 2016).
8 We used an automated sorting algorithm to identify individual neurons and tracked their
9 longitudinal activity across days (Figures 2D, S2A, and S2B; see Methods) (Kitamura et al., 2017;
10 Mukamel et al., 2009; Okuyama et al., 2016). Neuronal activity was analyzed under five
11 conditions: during habituation to the conditioning chamber before CFC (Hab), after footshocks
12 during CFC (Shock), fear retrieval (FR), extinction training (FET) and extinction retrieval (ER)
13 (Figures 2D; see Methods). In order to quantify how neuronal activity changed across these
14 conditions, the activity in each condition was explicitly compared to activity in another condition
15 that served as a baseline, i.e. shock compared to habituation, fear retrieval compared to habituation,
16 FET compared to fear retrieval, and extinction retrieval compared to fear retrieval (Figures 2E and
17 2F; see Methods). The cells with increased activity were referred to as Up cells, those with
18 decreased activity were referred to as Down cells, and cells whose activity did not change were
19 referred to as No change cells.

20 BLA *Rspo2*⁺ cells were predominantly activated by footshocks during CFC and fear
21 retrieval, and inhibited during extinction training and extinction retrieval (Figures 2E, 2G, and
22 S2C-S2E). Conversely, BLA *Ppp1r1b*⁺ neurons were predominantly activated during extinction
23 training and extinction retrieval, and inhibited by footshocks and fear retrieval (Figures 2F, 2G,

1 and S2F-S2H). A larger percentage of BLA *Rspo2*⁺ cells were activated during fear retrieval
2 compared to extinction retrieval, whereas a larger percentage of BLA *Ppp1r1b*⁺ cells were
3 activated during extinction retrieval compared to fear retrieval (Figure 2G). Compared to BLA
4 *Rspo2*⁺ cells whose activity did not change in response to footshocks (Shock-No Change), shock-
5 activated *Rspo2*⁺ cells (Shock-Up) were preferentially inhibited during extinction training and
6 extinction retrieval (Figures 2H, 2J and 2K). Specifically, 30% of the shock-activated *Rspo2*⁺ cells
7 were also activated during fear retrieval and inhibited during both extinction training and
8 extinction retrieval (Figures 2J and 2K). Compared to BLA *Ppp1r1b*⁺ neurons whose activity did
9 not change by extinction training (E.T.-No Change), *Ppp1r1b*⁺ neurons that were activated during
10 extinction training (E.T.-Up) were preferentially reactivated during extinction retrieval (Figures
11 2I, 2L and 2M). Combined with the c-Fos quantification data (Figure 1), these results show that
12 pBLA *Ppp1r1b*⁺ neurons are recruited during fear extinction training and reactivated during
13 extinction memory retrieval whereas shock-activated aBLA *Rspo2*⁺ neurons are suppressed during
14 these processes.

15

16 **BLA *Ppp1r1b*⁺ neurons drive fear extinction learning**

17 We next investigated the roles of *Ppp1r1b*⁺ and *Rspo2*⁺ neurons in fear extinction at the
18 behavioral level using optogenetic manipulations. We injected a Cre-dependent AAV carrying
19 ChR2 (channelrhodopsin-2) or ArchT (archaerhodopsin) into the aBLA of the *Rspo2-Cre* mice
20 (Figure 3A and S3A) or the pBLA of the *Ppp1r1b-Cre* mice (Figure 3B and S3B). Mice underwent
21 the three-day contextual fear extinction protocol as before (Figure 1A). Starting 5 min after the
22 onset of extinction training on Day 2, 3 min of blue or green laser light was delivered during FET

1 repeatedly at 2 min intervals (Figure 3C). Inhibition of *Rspo2*⁺ neurons and activation of *Ppp1r1b*⁺
2 neurons both facilitated fear extinction learning (Figures 3E and 3F). In contrast, activation of
3 *Rspo2*⁺ neurons and inhibition of *Ppp1r1b*⁺ neurons both impaired fear extinction and extinction
4 memory retrieval (Figures 3D and 3G). These results show that pBLA *Ppp1r1b*⁺ neurons are
5 necessary and sufficient to form fear extinction memory, whereas aBLA *Rspo2*⁺ neurons
6 antagonize it.

8 **Fear extinction memory engram cells are formed in BLA *Ppp1r1b*⁺ neuronal population**

9 As memories are stored in specific engram cell circuits (Tonegawa et al., 2015), we then
10 investigated whether the engram for fear extinction memory can be identified in the pBLA
11 *Ppp1r1b*⁺ neurons. By using our previously established engram technology (Liu et al., 2012;
12 Ramirez et al., 2013; Redondo et al., 2014; Ryan et al., 2015), we injected AAV₉-*c-Fos-tTA*
13 (tetracycline transactivator) together with AAV₉-*TRE* (tetracycline response element)-*Chr2-*
14 *EYFP* (Ext-ChR2 group) or AAV₉-*TRE-EYFP* (Ext-EYFP group) into the pBLA of C57BL/6 mice
15 (Figures 4A and 4B; see Methods), and subjected them to a series of engram labeling and
16 behavioral steps (Figure 4D; see Methods). This genetic manipulation permits labeling of engram
17 neurons during memory formation or memory retrieval (Figures S4A-S4C) (Khalaf et al., 2018;
18 Liu et al., 2012; Ramirez et al., 2013). During fear extinction training, formation of the putative
19 fear extinction memory engram cells occurs in parallel with retrieval of the original fear memory.
20 In order to isolate the putative fear extinction memory engram neurons, we used the retrieval stage
21 of the fear extinction memory for their labelling (Day3) as fear memory expression will have
22 subsided to the background level by then (Figure 4D).

1 A large fraction ($79 \pm 2\%$) of pBLA neurons activated by fear extinction retrieval were
2 *Ppp1r1b*⁺ (Figure 4C). In order to access the contribution of food and water supplied in the home
3 cage (HC) to the baseline activity of *Ppp1r1b*⁺ neurons, we set up another negative control group,
4 HC-ChR2 group (Figure 4D). As expected, ChR2-EYFP⁺ cells in the HC-ChR2 group was very
5 sparse compared to those in the ChR2-EYFP group (Figure 4B, S4C-S4D). Three groups of mice,
6 Ext-ChR2 group, Ext-EYFP group and HC-ChR2 group, exhibited indistinguishable freezing
7 levels during the first round of contextual fear extinction protocol from Day 1 to Day 3 (Figure
8 4E). To examine the roles of these labeled neurons in contextual fear extinction, we subjected the
9 mice to another round of CFC in the same box for fear reinstatement (Day 5). When these neurons
10 were optogenetically activated during the second round of fear extinction training on Day 6 (see
11 Methods), the Ext-ChR2 group showed accelerated extinction compared to the Ext-EYFP and HC-
12 ChR2 groups (Figure 4F). This reactivation-induced accelerated extinction was not observed when
13 the laser was off (Figures S4E-S4F). These results show that the extinction memory engram is
14 formed in the BLA *Ppp1r1b*⁺ neuron population.

16 **BLA *Ppp1r1b*⁺ extinction engram cells are crucial for suppressing *Rspo2*⁺ fear cells**

17 We then investigated whether pBLA *Ppp1r1b*⁺ extinction engram neurons are not only
18 capable of driving fear extinction, but also necessary for maintaining extinction memory and for
19 suppressing *Rspo2*⁺ fear cells. We labeled the pBLA extinction engram cells with ArchT-mCherry
20 during fear extinction memory retrieval (Figures 5A-5C and S5A). As expected, a large proportion
21 ($75 \pm 2\%$) of the pBLA neurons activated by fear extinction retrieval were PPP1R1B⁺ (Figures
22 S5B-S5C). Optogenetic inhibition of these ArchT-labeled neurons specifically impaired the

1 retrieval of fear extinction memory on Day 4 (Figure 5D) and failed to suppress BLA *Rspo2*⁺ fear
2 neurons (Figures 5E-5G). The later is consistent with the feed-forward inhibition of *Rspo2*⁺
3 neurons by *Ppp1r1b*⁺ extinction neurons within the BLA (Kim et al., 2016). Together, these results
4 demonstrate a causal role of pBLA *Ppp1r1b*⁺ engram cells in extinction memory. These fear
5 extinction engram cells: i) are reactivated during extinction memory retrieval (Figure 2M), ii) have
6 the capacity to drive fear extinction (Figure 4F), and iii) are necessary for suppressing the original
7 fear memory (Figures 5D-5G).

8

9 **Reward neurons and fear extinction engram neurons overlap extensively in the amygdala**

10 Our previous study demonstrated that pBLA *Ppp1r1b*⁺ neurons responded to a variety of
11 appetitive stimuli (Kim et al., 2016), whereas our present study shows that fear extinction memory
12 is stored in the pBLA *Ppp1r1b*⁺ neuronal population (Figures 1-5). Therefore, we investigated the
13 relationship between these two subsets of *Ppp1r1b*⁺ neurons. To label reward-responding pBLA
14 neurons, AAV₉-*c-Fos-tTA* and AVV₉-*TRE-EYFP* virus were injected into pBLA of C57BL/6 mice
15 on Dox-ON diet (Figure 6A). One week post-surgery, all mice were partially deprived of water for
16 24 hrs on Dox-ON diet (Day 1) and water was given as a reward on Dox-OFF diet (Day 2) (Figure
17 6B). Mice were then divided into W/W (Water/Water), W/F (Water/Food), W/Ext
18 (Water/Extinction), and W/NonExt (Water/NonExtinction) groups, and kept on Dox-ON diet
19 throughout the remainder of the experiment (Figure 6B; see Methods).

20 Neurons activated by water reward on Day 2 were labeled with EYFP, and their proportions
21 of all neurons (EYFP⁺/DAPI⁺) were similar across all four groups (Figures 6C, 6F and S6A).
22 Neurons activated by the stimulus delivered on Day 5, namely the second delivery of water reward

1 (W/W group), food reward (W/F group), extinction memory retrieval (W/Ext group), and fear
2 memory retrieval (W/NonExt group), were detected by immunohistochemistry for endogenous c-
3 Fos. The proportions of c-Fos⁺ cells of all neurons (c-Fos⁺/DAPI⁺) were similar between W/W
4 group, W/F group and W/Ext group, and were substantially higher compared to the W/NonExt
5 group (Figures 6D, 6F, and S6A). The proportion of the c-Fos and EYFP double-positive cells of
6 EYFP⁺ cells (c-Fos⁺EYFP⁺/EYFP⁺) was used to measure the degree of similarity between the cells
7 responding to the pair of stimuli used in each mouse group. This proportion reached as high as
8 40% in the W/W group, which represented the ceiling effect of a repeated presentation of the same
9 reward. The proportion of W/Ext group was lower compared to W/W group but it did not differ
10 significantly from the proportion of W/F group, and both of these groups and W/W group gave
11 greater proportions than W/NonExt group (Figures 6E, 6F, and S6A). These results show that the
12 overlap of pBLA *Ppp1r1b*⁺ neurons activated by fear extinction and water reward is as high as the
13 overlap of pBLA *Ppp1r1b*⁺ neurons activated by two *bona fide* rewards (water and food).

14

15 ***Ppp1r1b*⁺ fear extinction engram cells and natural reward cells are interchangeable**

16 We then examined whether this observed cellular overlap could also be demonstrated at
17 the behavioral level. To test whether optogenetic activation of fear extinction engram cells can
18 drive appetitive behavior, four groups of C57BL/6 mice were set up (Figures 7A and 7B). In the
19 Ext-ChR2 and Ext-EYFP groups, extinction engram cells were labeled with ChR2-EYFP and
20 EYFP during extinction retrieval on Day 3, respectively (Figures 7B, S7A, and S7B). As controls,
21 neurons activated by water reward and fear memory retrieval were labeled with ChR2 in Water-
22 ChR2 group and CFC-ChR2 group, respectively (Figures 7B, S7C, and S7D). On Day 4, all mice

1 were subjected to an optogenetic self-stimulation behavior test or optogenetic place preference test
2 (Figures 7B, S7E and S7F; see Methods). 50% of Ext-ChR2 group mice and 67% of Water-ChR2
3 group mice displayed self-stimulation behavior, but not the Ext-EYFP group or CFC-ChR2 group
4 (Figure 7C). In the optogenetic real-time place preference test, mice in the Ext-ChR2 group and
5 Water-ChR2 group spent significantly more time on the stimulated side, whereas Ext-EYFP and
6 CFC-ChR2 groups spent equal times on both sides of the chamber (Figures 7D, 7E, and S7G).
7 Reciprocally, optogenetic activation of ChR2-expressing water reward cells accelerated fear
8 extinction learning (Figures 7F and 7G). Thus, activation of fear extinction engram cells generates
9 appetitive behaviors and activation of reward neurons facilitates fear extinction.

11 Discussion

12 In this study, we have shown that fear extinction requires new memory engram cells that
13 are formed and stored within BLA *Ppp1r1b*⁺ reward neuronal population. The fear extinction
14 engram cells are defined by their reactivation during extinction retrieval, and their sufficiency and
15 necessity to drive fear extinction. The overlap of neuronal activation of extinction memory engram
16 neurons and natural reward-responsive neurons is extensive and these two types of neurons are
17 mutually interchangeable in driving reward functions and fear extinction behaviors. Thus, the shift
18 from fear to safety in the fear extinction process is mediated by newly formed *Ppp1r1b*⁺ reward
19 memory neurons and their inhibition of the original *Rspo2*⁺ fear memory neurons, both present
20 within the BLA.

21 The inhibition of *Rspo2*⁺ fear neurons by *Ppp1r1b*⁺ extinction neurons is likely to be
22 mediated by the local inhibitory interneurons in BLA (Kim et al., 2016), supporting the emerging

1 notion that the switch between fear expression and fear extinction takes place within the BLA
2 (Ehrlich et al., 2009; Grewe et al., 2017; Herry et al., 2008; Pare and Duvarci, 2012). Consistent
3 with previous recording studies (Grewe et al., 2017; Herry et al., 2008), our study shows fear
4 extinction is due to new learning rather than a loss of the original fear memory (Herry et al., 2010;
5 Myers and Davis, 2007; Pavlov, 1927). The *Ppp1r1b*⁺ extinction neurons may correspond to the
6 previously reported BLA neurons identified by *in vivo* recording, which were activated throughout
7 extinction learning after cue fear conditioning (Herry et al., 2008). However, our findings extend
8 beyond that study as we further demonstrated that the newly formed extinction memory is an
9 appetitive memory and inhibits the original fear memory.

11 **Fear extinction recruits the brain's reward circuit**

12 The finding that the fear extinction memory is a rewarding memory indicates that the
13 emotional valence associated with CS is switched from negative to positive during fear extinction.
14 Notably, the finding that activation of fear extinction engram cells can elicit a reward effect (Figure
15 7) supports the notion that the absence of expected punishments is in itself a reward. Therefore,
16 fear extinction is positively reinforced by recruiting the brain's reward circuitry. Previous studies
17 correlated the reward system to fear extinction (Correia et al., 2016; Felsenberg et al., 2018), but
18 their causal relationship was never demonstrated. In contrast, our results show a causal role of the
19 reward system in fear extinction: the pBLA reward neuron population is a necessary component
20 of the fear extinction memory. Furthermore, these findings provide strong evidence of the
21 antagonistic interactions between aversive and appetitive systems that have been observed in
22 Pavlovian fear conditioning (Nasser and McNally, 2012).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21

Multiple parallel circuits in fear extinction

In addition to direct feed-forward inhibition within BLA, BLA *Ppp1r1b*⁺ extinction neurons could also drive fear extinction through their interactions with other brain regions. As downstream targets of BLA, both intercalated cells (ITC) and central amygdala (CeA) have been shown to play important roles in fear extinction (Likhtik et al., 2008). Our previous study showed that BLA *Ppp1r1b*⁺ neurons provide monosynaptic projections to the *Prkcd*⁺ neurons located in the lateral subnuclei of CeA (CeL), which are also activated during fear extinction retrieval (Kim et al., 2017). Therefore, the fear extinction behavior can also be modulated via BLA *Ppp1r1b*⁺-CeA monosynaptic projection or BLA *Ppp1r1b*⁺-ITC-CeA di-synaptic projection. The infralimbic region (IL) of the prefrontal cortex is another critical cite for fear extinction (Burgos-Robles et al., 2007; Laurent and Westbrook, 2009; Vidal-Gonzalez et al., 2006) and its connections with BLA *Ppp1r1b*⁺ neurons may also be involved in this process (Cho et al., 2013; Kim et al., 2016; Likhtik et al., 2005; Quirk et al., 2006; Sotres-Bayon et al., 2012).

Intrinsic reinforcement signals during fear extinction learning

How *Ppp1r1b*⁺ cells are activated for the acquisition of fear extinction memory is another question that remains to be elucidated. Reinforcement learning theories would predict that acquisition of extinction memory is instructed by negative prediction error, in which the predictive value of the CS with respect to the occurrence of the US is compromised (McNally et al., 2011; Schultz, 2016). Given there is no obvious external US in the case of fear extinction, this stimulus

1 may be provided internally and preferentially to *Ppp1r1b* neurons in the BLA. The identification
2 of the agent and the circuit that provide this stimulus would be of fundamental importance.

3 Accumulating evidence suggests that dopamine reinforces fear extinction learning by
4 signaling the omission of expected aversive US in both animals and humans (de Jong et al., 2019;
5 Kalisch et al., 2019; Matsumoto and Hikosaka, 2009; Raczka et al., 2011; Ray Luo, 2018; Salinas-
6 Hernandez et al., 2018). In the mesolimbic dopaminergic system, amygdala receives dopaminergic
7 inputs from the ventral tegmental area (VTA) and substantia nigra (SNc) in the midbrain (Brinley-
8 Reed and McDonald, 1999; Pezze and Feldon, 2004). Future studies are needed to investigate how
9 dopamine functions as an instructive signal onto BLA neurons for fear extinction learning.

10

11 In conclusion, our study provides the empirical evidence for the key concept derived from
12 learning theory that an omission of the expected punishment is rewarding (Felsenberg et al., 2018;
13 Kim et al., 2006; Nasser and McNally, 2012), pinpointing potential targets for effective
14 interventions of fear-related disorders (Shalev et al., 2017; Stein and Sareen, 2015).

1 **References**

- 2 Amano, T., Unal, C.T., and Pare, D. (2010). Synaptic correlates of fear extinction in the amygdala.
3 *Nat Neurosci* *13*, 489-494.
- 4 Baldi, E., and Bucherelli, C. (2014). Entorhinal cortex contribution to contextual fear conditioning
5 extinction and reconsolidation in rats. *Neurobiol Learn Mem* *110*, 64-71.
- 6 Beyeler, A., Namburi, P., Glover, G.F., Simonnet, C., Calhoun, G.G., Conyers, G.F., Luck, R.,
7 Wildes, C.P., and Tye, K.M. (2016). Divergent Routing of Positive and Negative Information from
8 the Amygdala during Memory Retrieval. *Neuron* *90*, 348-361.
- 9 Bouton, M.E. (2004). Context and behavioral processes in extinction. *Learn Mem* *11*, 485-494.
- 10 Brinley-Reed, M., and McDonald, A.J. (1999). Evidence that dopaminergic axons provide a dense
11 innervation of specific neuronal subpopulations in the rat basolateral amygdala. *Brain Res* *850*,
12 127-135.
- 13 Burgos-Robles, A., Vidal-Gonzalez, I., Santini, E., and Quirk, G.J. (2007). Consolidation of fear
14 extinction requires NMDA receptor-dependent bursting in the ventromedial prefrontal cortex.
15 *Neuron* *53*, 871-880.
- 16 Calandrea, L., Desmedt, A., Decorte, L., and Jaffard, R. (2005). A different recruitment of the
17 lateral and basolateral amygdala promotes contextual or elemental conditioned association in
18 Pavlovian fear conditioning. *Learn Mem* *12*, 383-388.
- 19 Cho, J.H., Deisseroth, K., and Bolshakov, V.Y. (2013). Synaptic encoding of fear extinction in
20 mPFC-amygdala circuits. *Neuron* *80*, 1491-1507.
- 21 Correia, S.S., McGrath, A.G., Lee, A., Graybiel, A.M., and Goosens, K.A. (2016). Amygdala-
22 ventral striatum circuit activation decreases long-term fear. *Elife* *5*.
- 23 Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annu Rev Neurosci* *15*, 353-375.
- 24 Davis, M., and Whalen, P.J. (2001). The amygdala: vigilance and emotion. *Mol Psychiatry* *6*, 13-
25 34.
- 26 de Jong, J.W., Afjei, S.A., Pollak Dorocic, I., Peck, J.R., Liu, C., Kim, C.K., Tian, L., Deisseroth,
27 K., and Lammel, S. (2019). A Neural Circuit Mechanism for Encoding Aversive Stimuli in the
28 Mesolimbic Dopamine System. *Neuron* *101*, 133-151 e137.
- 29 Duvarci, S., and Pare, D. (2014). Amygdala microcircuits controlling learned fear. *Neuron* *82*,
30 966-980.
- 31 Ehrlich, I., Humeau, Y., Grenier, F., Ciochi, S., Herry, C., and Luthi, A. (2009). Amygdala
32 inhibitory circuits and the control of fear memory. *Neuron* *62*, 757-771.
- 33 Felsenberg, J., Jacob, P.F., Walker, T., Barnstedt, O., Edmondson-Stait, A.J., Pleijzier, M.W.,
34 Otto, N., Schlegel, P., Sharifi, N., Perisse, E., *et al.* (2018). Integration of Parallel Opposing
35 Memories Underlies Memory Extinction. *Cell* *175*, 709-722 e715.
- 36 Goosens, K.A., and Maren, S. (2001). Contextual and auditory fear conditioning are mediated by
37 the lateral, basal, and central amygdaloid nuclei in rats. *Learn Mem* *8*, 148-155.
- 38 Grewe, B.F., Grundemann, J., Kitch, L.J., Lecoq, J.A., Parker, J.G., Marshall, J.D., Larkin, M.C.,
39 Jercog, P.E., Grenier, F., Li, J.Z., *et al.* (2017). Neural ensemble dynamics underlying a long-term
40 associative memory. *Nature*.
- 41 Herry, C., Ciochi, S., Senn, V., Demmou, L., Muller, C., and Luthi, A. (2008). Switching on and
42 off fear by distinct neuronal circuits. *Nature* *454*, 600-606.
- 43 Herry, C., Ferraguti, F., Singewald, N., Letzkus, J.J., Ehrlich, I., and Luthi, A. (2010). Neuronal
44 circuits of fear extinction. *Eur J Neurosci* *31*, 599-612.

- 1 Huff, N.C., and Rudy, J.W. (2004). The amygdala modulates hippocampus-dependent context
2 memory formation and stores cue-shock associations. *Behav Neurosci* *118*, 53-62.
- 3 Kalisch, R., Gerlicher, A.M.V., and Duvarci, S. (2019). A Dopaminergic Basis for Fear Extinction.
4 *Trends Cogn Sci*.
- 5 Khalaf, O., Resch, S., Dixsaut, L., Gorden, V., Glauser, L., and Graff, J. (2018). Reactivation of
6 recall-induced neurons contributes to remote fear memory attenuation. *Science* *360*, 1239-1242.
- 7 Kim, H., Shimojo, S., and O'Doherty, J.P. (2006). Is avoiding an aversive outcome rewarding?
8 Neural substrates of avoidance learning in the human brain. *PLoS Biol* *4*, e233.
- 9 Kim, J., Pignatelli, M., Xu, S., Itohara, S., and Tonegawa, S. (2016). Antagonistic negative and
10 positive neurons of the basolateral amygdala. *Nat Neurosci* *19*, 1636-1646.
- 11 Kim, J., Zhang, X., Muralidhar, S., LeBlanc, S.A., and Tonegawa, S. (2017). Basolateral to Central
12 Amygdala Neural Circuits for Appetitive Behaviors. *Neuron* *93*, 1464-1479 e1465.
- 13 Kitamura, T., Ogawa, S.K., Roy, D.S., Okuyama, T., Morrissey, M.D., Smith, L.M., Redondo,
14 R.L., and Tonegawa, S. (2017). Engrams and circuits crucial for systems consolidation of a
15 memory. *Science* *356*, 73-78.
- 16 Laurent, V., and Westbrook, R.F. (2009). Inactivation of the infralimbic but not the prelimbic
17 cortex impairs consolidation and retrieval of fear extinction. *Learn Mem* *16*, 520-529.
- 18 Likhtik, E., Pelletier, J.G., Paz, R., and Pare, D. (2005). Prefrontal control of the amygdala. *J*
19 *Neurosci* *25*, 7429-7437.
- 20 Likhtik, E., Popa, D., Apergis-Schoute, J., Fidacaro, G.A., and Pare, D. (2008). Amygdala
21 intercalated neurons are required for expression of fear extinction. *Nature* *454*, 642-645.
- 22 Liu, X., Ramirez, S., Pang, P.T., Puryear, C.B., Govindarajan, A., Deisseroth, K., and Tonegawa,
23 S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature*
24 *484*, 381-385.
- 25 Maren, S., and Fanselow, M.S. (1996). The amygdala and fear conditioning: has the nut been
26 cracked? *Neuron* *16*, 237-240.
- 27 Matsumoto, M., and Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey
28 positive and negative motivational signals. *Nature* *459*, 837-841.
- 29 McNally, G.P., Johansen, J.P., and Blair, H.T. (2011). Placing prediction into the fear circuit.
30 *Trends Neurosci* *34*, 283-292.
- 31 Mukamel, E.A., Nimmerjahn, A., and Schnitzer, M.J. (2009). Automated analysis of cellular
32 signals from large-scale calcium imaging data. *Neuron* *63*, 747-760.
- 33 Myers, K.M., and Davis, M. (2007). Mechanisms of fear extinction. *Mol Psychiatry* *12*, 120-150.
- 34 Namburi, P., Beyeler, A., Yorozu, S., Calhoun, G.G., Halbert, S.A., Wichmann, R., Holden, S.S.,
35 Mertens, K.L., Anahtar, M., Felix-Ortiz, A.C., *et al.* (2015). A circuit mechanism for
36 differentiating positive and negative associations. *Nature* *520*, 675-678.
- 37 Nasser, H.M., and McNally, G.P. (2012). Appetitive-aversive interactions in Pavlovian fear
38 conditioning. *Behav Neurosci* *126*, 404-422.
- 39 Okuyama, T., Kitamura, T., Roy, D.S., Itohara, S., and Tonegawa, S. (2016). Ventral CA1 neurons
40 store social memory. *Science* *353*, 1536-1541.
- 41 Pare, D., and Duvarci, S. (2012). Amygdala microcircuits mediating fear expression and
42 extinction. *Curr Opin Neurobiol* *22*, 717-723.
- 43 Pavlov, I.P., and Anrep, G.V. (1927). *Conditioned reflexes; an investigation of the physiological*
44 *activity of the cerebral cortex* (London: Oxford University Press: Humphrey Milford).

- 1 Pavlov, P.I. (1927). Conditioned reflexes: An investigation of the physiological activity of the
2 cerebral cortex. *Ann Neurosci* *17*, 136-141.
- 3 Pezze, M.A., and Feldon, J. (2004). Mesolimbic dopaminergic pathways in fear conditioning. *Prog*
4 *Neurobiol* *74*, 301-320.
- 5 Quirk, G.J., Garcia, R., and Gonzalez-Lima, F. (2006). Prefrontal mechanisms in extinction of
6 conditioned fear. *Biol Psychiatry* *60*, 337-343.
- 7 Quirk, G.J., and Mueller, D. (2008). Neural mechanisms of extinction learning and retrieval.
8 *Neuropsychopharmacology* *33*, 56-72.
- 9 Quirk, G.J., Pare, D., Richardson, R., Herry, C., Monfils, M.H., Schiller, D., and Vicentic, A.
10 (2010). Erasing fear memories with extinction training. *J Neurosci* *30*, 14993-14997.
- 11 Raczka, K.A., Mechias, M.L., Gartmann, N., Reif, A., Deckert, J., Pessiglione, M., and Kalisch,
12 R. (2011). Empirical support for an involvement of the mesostriatal dopamine system in human
13 fear extinction. *Transl Psychiatry* *1*, e12.
- 14 Ramirez, S., Liu, X., Lin, P.A., Suh, J., Pignatelli, M., Redondo, R.L., Ryan, T.J., and Tonegawa,
15 S. (2013). Creating a false memory in the hippocampus. *Science* *341*, 387-391.
- 16 Ray Luo, A.U., Adam Weitemier, Luca Aquili, Jenny Koivumaa, Thomas J. McHugh and Joshua
17 P. Johansen (2018). A dopaminergic switch for fear to safety transitions. *Nature Communication*.
- 18 Redondo, R.L., Kim, J., Arons, A.L., Ramirez, S., Liu, X., and Tonegawa, S. (2014). Bidirectional
19 switch of the valence associated with a hippocampal contextual memory engram. *Nature* *513*, 426-
20 430.
- 21 Ryan, T.J., Roy, D.S., Pignatelli, M., Arons, A., and Tonegawa, S. (2015). Memory. Engram cells
22 retain memory under retrograde amnesia. *Science* *348*, 1007-1013.
- 23 Salinas-Hernandez, X.I., Vogel, P., Betz, S., Kalisch, R., Sigurdsson, T., and Duvarci, S. (2018).
24 Dopamine neurons drive fear extinction learning by signaling the omission of expected aversive
25 outcomes. *Elife* *7*.
- 26 Schultz, W. (2016). Dopamine reward prediction error coding. *Dialogues Clin Neurosci* *18*, 23-
27 32.
- 28 Shalev, A., Liberzon, I., and Marmar, C. (2017). Post-Traumatic Stress Disorder. *N Engl J Med*
29 *376*, 2459-2469.
- 30 Sotres-Bayon, F., Sierra-Mercado, D., Pardilla-Delgado, E., and Quirk, G.J. (2012). Gating of fear
31 in prelimbic cortex by hippocampal and amygdala inputs. *Neuron* *76*, 804-812.
- 32 Stein, M.B., and Sareen, J. (2015). CLINICAL PRACTICE. Generalized Anxiety Disorder. *N Engl*
33 *J Med* *373*, 2059-2068.
- 34 Tonegawa, S., Liu, X., Ramirez, S., and Redondo, R. (2015). Memory Engram Cells Have Come
35 of Age. *Neuron* *87*, 918-931.
- 36 Trouche, S., Sasaki, J.M., Tu, T., and Reijmers, L.G. (2013). Fear extinction causes target-specific
37 remodeling of perisomatic inhibitory synapses. *Neuron* *80*, 1054-1065.
- 38 Vidal-Gonzalez, I., Vidal-Gonzalez, B., Rauch, S.L., and Quirk, G.J. (2006). Microstimulation
39 reveals opposing influences of prelimbic and infralimbic cortex on the expression of conditioned
40 fear. *Learn Mem* *13*, 728-733.
- 41 Zushida, K., Sakurai, M., Wada, K., and Sekiguchi, M. (2007). Facilitation of extinction learning
42 for contextual fear memory by PEPA: a potentiator of AMPA receptors. *J Neurosci* *27*, 158-166.

1 **Acknowledgments**

2 We thank X. Zhou, K. Marshall, B. Bane, F.J. Bushard and W. Yu for technical assistance; C. Sun
3 and C.J. MacDonald for discussion on the calcium imaging data analysis; M. Pignatelli, S.
4 Muralidhar, C. Sun and C.J. MacDonald for comments and discussion on the manuscript; and all
5 members of the Tonegawa laboratory for their support.

6 **Funding**

7 This work was supported by the RIKEN Brain Science Institute, the Howard Hughes Medical
8 Institute, and the JPB Foundation.

9 **Author contributions:**

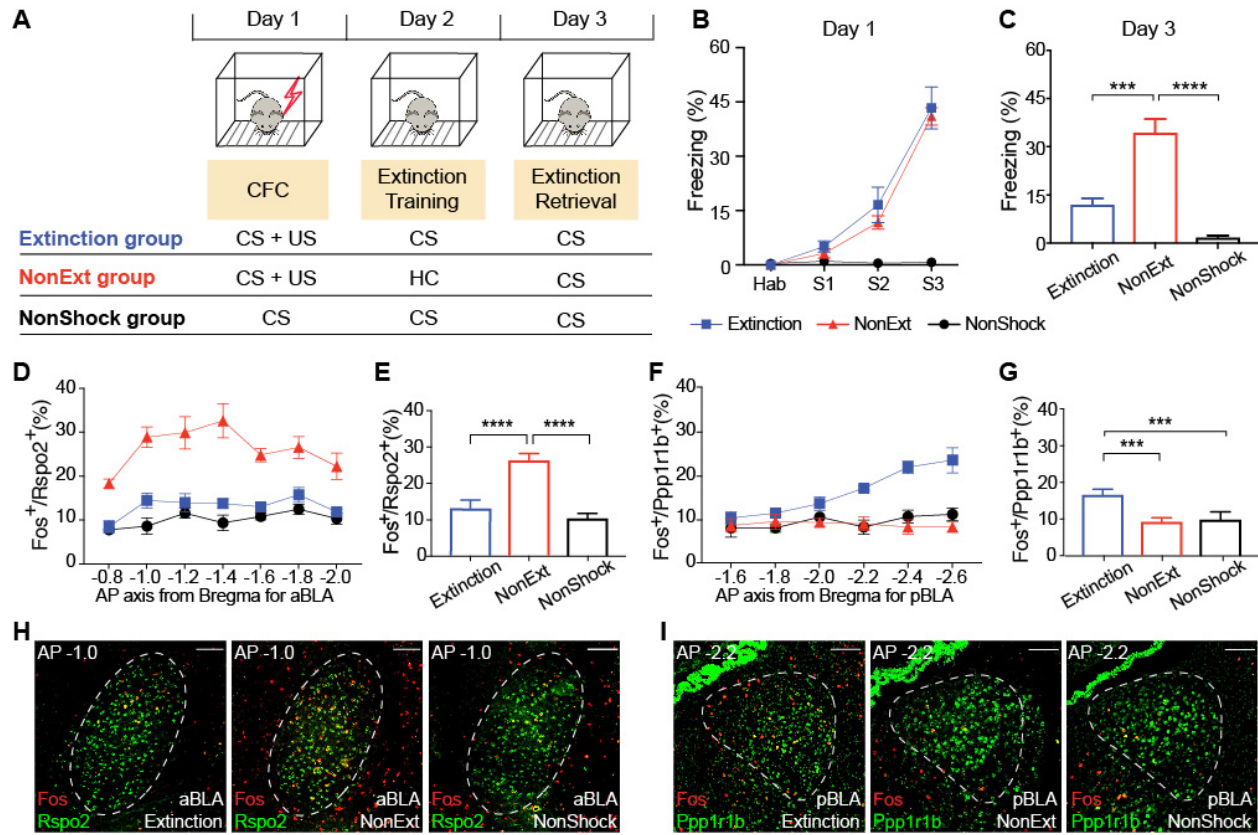
10 X.Z., J.K. and S.T. contributed to the study design. X.Z. contributed to the data collection and
11 interpretation. X.Z. conducted the surgeries, behavior experiments and histological analyses. X.Z.
12 and S.T. wrote the paper. All authors discussed and commented on the manuscript.

13 **Competing interests**

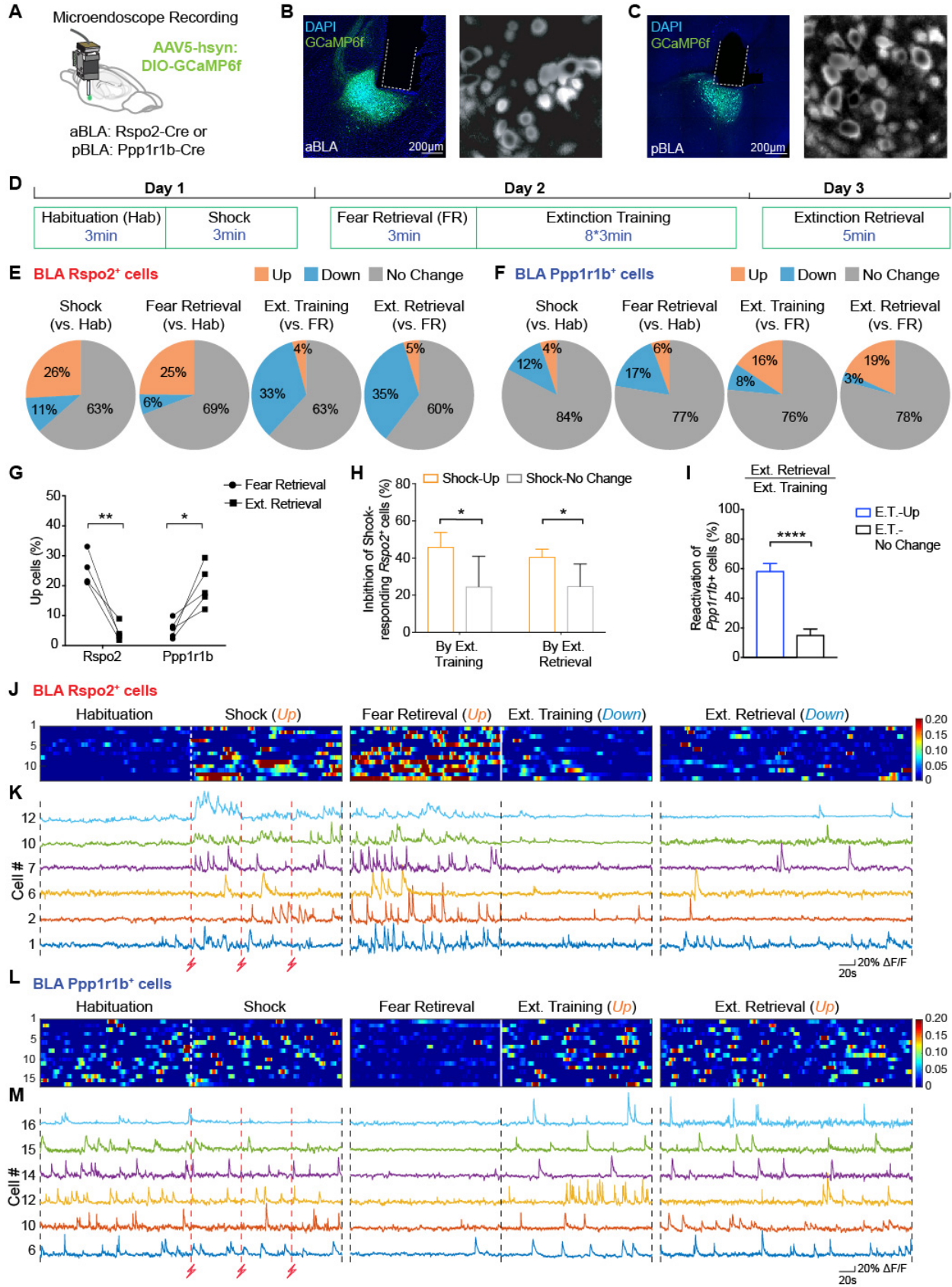
14 The authors declare no competing financial interests. Correspondence and requests for materials
15 should be addressed to Susumu Tonegawa.

16 **Data and material availability**

17 AAV₉-c-Fos-tTA, AAV₉-TRE-ChR2-EYFP, AAV₉-TRE-eArchT-mCherry, and AAV₉-TRE-
18 EYFP were developed at the Massachusetts Institute of Technology by the group of S.T.; virus
19 plasmids are available through a material transfer agreement. The Rspo2-Cre transgenic mouse
20 line was developed at RIKEN BSI by the group of Shigeyoshi Itohara and is available through a
21 material transfer agreement. All data is available in the main text or the supplementary materials.



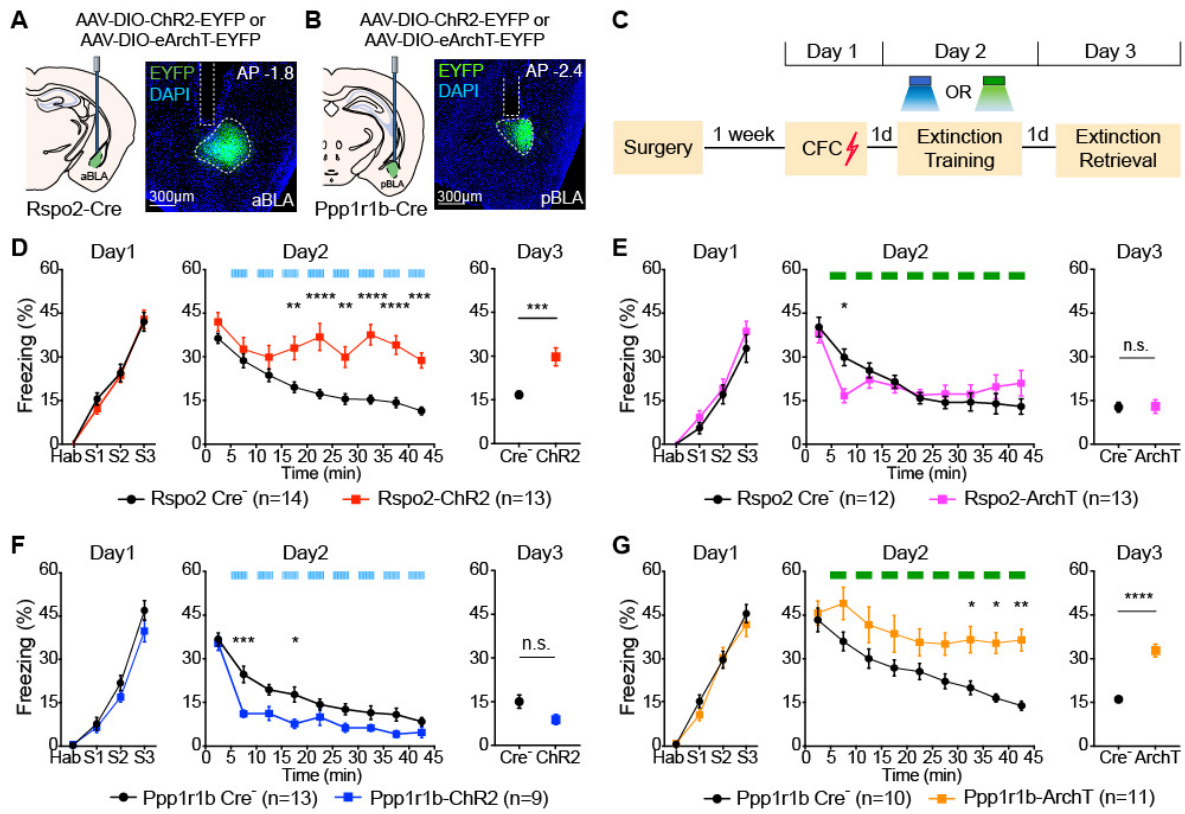
1 **Figure 1. BLA *Ppp1r1b*⁺ neurons are activated during contextual fear extinction**
2 (A) Experimental design of the Extinction group, NonExtinction group and NonShock group. (B
3 and C) Freezing levels of all three groups on Day 1 (B) and Day 3 (C). One-way ANOVA.
4 Extinction $n = 4$; NonExt $n = 4$; NonShock $n = 4$. (D) Percentages of *Fos*⁺ neurons within BLA
5 *Rspo2*⁺ neurons across A/P axis, -0.8 mm to -2.0 mm. (E) Average of the percentage of
6 *Fos*⁺/*Rspo2*⁺ shown in (D). One-way ANOVA. Extinction $n = 4$; NonExt $n = 4$; NonShock $n = 4$.
7 (F) Percentages of *Fos*⁺ neurons within BLA *Ppp1r1b*⁺ neurons across A/P axis, -1.6 mm to -2.6
8 mm. (G) Average of the percentage of *Fos*⁺/*Ppp1r1b*⁺ shown in (F). One-way ANOVA.
9 Extinction $n = 4$; NonExt $n = 4$; NonShock $n = 4$. (H and I) Double smFISH of Fos (red) and
10 *Rspo2* (green, H) in aBLA or *Ppp1r1b* (green, I) in pBLA. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$,
11 **** $P < 0.0001$. Data are presented as mean \pm SEM. Data are presented as mean \pm SEM. Scale
12 bars: 200 μ m (H and I)



1 **Figure 2. Longitudinal Ca²⁺ imaging of BLA *Rspo2*⁺ and *Ppp1r1b*⁺ neurons during fear**
2 **extinction**

3 (A) Cre-dependent expression of GCamp6f and implantation of microendoscope above the aBLA
4 of *Rspo2*-Cre and pBLA of *Ppp1r1b*-Cre mice. (B) aBLA of *Rspo2*-Cre mice. Left: Representative
5 image showing GCamp6f expression and GRIN lens implant; Right: Stacked FOV (Field of View)
6 image. (C) pBLA of *Ppp1r1b*-Cre mice. Left: Representative image showing GCamp6f expression
7 and GRIN lens implant; Right: Stacked FOV image. (D) Behavior protocol for Ca²⁺ recording. (E)
8 From left to right: percentages of BLA *Rspo2*⁺ neurons with increased (Up, orange), decreased
9 (Down, blue) or No change (grey) responses to shock (vs. habituation), fear retrieval (vs.
10 habituation), extinction training (vs. fear retrieval) and extinction retrieval (vs. fear retrieval). 169
11 cells total; 4 *Rspo2*-Cre mice. F.R. represents fear retrieval. (F) From left to right: percentages of
12 BLA *Ppp1r1b*⁺ neurons with increased (Up, orange), decreased (Down, blue) or No change (grey)
13 responses to shock (vs. habituation), fear retrieval (vs. habituation), extinction training (vs. fear
14 retrieval) and extinction retrieval (vs. fear retrieval). 179 cells total; 5 *Ppp1r1b*-Cre mice. F.R.
15 represents fear retrieval. (G) Percentages of activated BLA *Rspo2*⁺ and *Ppp1r1b*⁺ neurons during
16 fear retrieval on Day 2 and extinction retrieval on Day 3 per animal. (H) Compared to BLA *Rspo2*⁺
17 neurons with stable responses to footshocks (Shock-No change, grey), BLA *Rspo2*⁺ neurons with
18 increased response to shock (Shock-Up, orange) were preferentially inhibited during extinction
19 training ($45.9 \pm 3.9\%$ vs. $24.6 \pm 8.2\%$) and extinction retrieval ($40.5 \pm 2.1\%$ vs. $24.8 \pm 12.0\%$).
20 Paired t-test. (I) Compared to BLA *Ppp1r1b*⁺ neurons with stable activity during extinction
21 training (E.T.-No Change), BLA *Ppp1r1b*⁺ neurons with increased activity during extinction
22 training (E.T.-Up) were strongly reactivated during extinction retrieval ($58.2 \pm 5.3\%$ vs. $15.1 \pm$

1 4.2%). Paired t-test. **(J)** Relative fluorescence change ($\Delta F/F$) of BLA *Rspo2*⁺ neurons that showed
2 increased activity during shock, increased activity during fear retrieval, decreased activity during
3 extinction training and extinction retrieval (n=13). Each row represents the same cell that was
4 tracked across three days. **(K)** Ca²⁺ traces of six representative BLA *Rspo2*⁺ neurons that are shown
5 in (J) with corresponding cell numbers. **(L)** Relative fluorescence change ($\Delta F/F$) of BLA *Ppp1r1b*⁺
6 neurons that showed increased activity during both extinction training and extinction retrieval
7 (n=16). Each row represents the same cell that was tracked across three days. **(M)** Ca²⁺ traces of
8 six representative BLA *Ppp1r1b*⁺ neurons that are shown in (L) with corresponding cell numbers.
9 **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001. Data are presented as mean ± SEM.



1

Figure 3. Activation of BLA *Ppp1r1b*⁺ neurons and inhibition of BLA *Rspo2*⁺ neurons facilitate contextual fear extinction

(A) Left: diagram of bilateral injection of AAV-DIO-ChR2-EYFP or AAV-DIO-eArchT-EYFP virus and optical fiber implant in the aBLA of *Rspo2*-Cre mice. Right: representative histology.

(B) Left: diagram of bilateral injection of AAV-DIO-ChR2-EYFP or AAV-DIO-eArchT-EYFP virus and optical fiber implant in the pBLA of *Ppp1r1b*-Cre mice. Right: representative histology.

(C) Experimental protocol of optogenetic manipulation during fear extinction. (D) Optogenetic

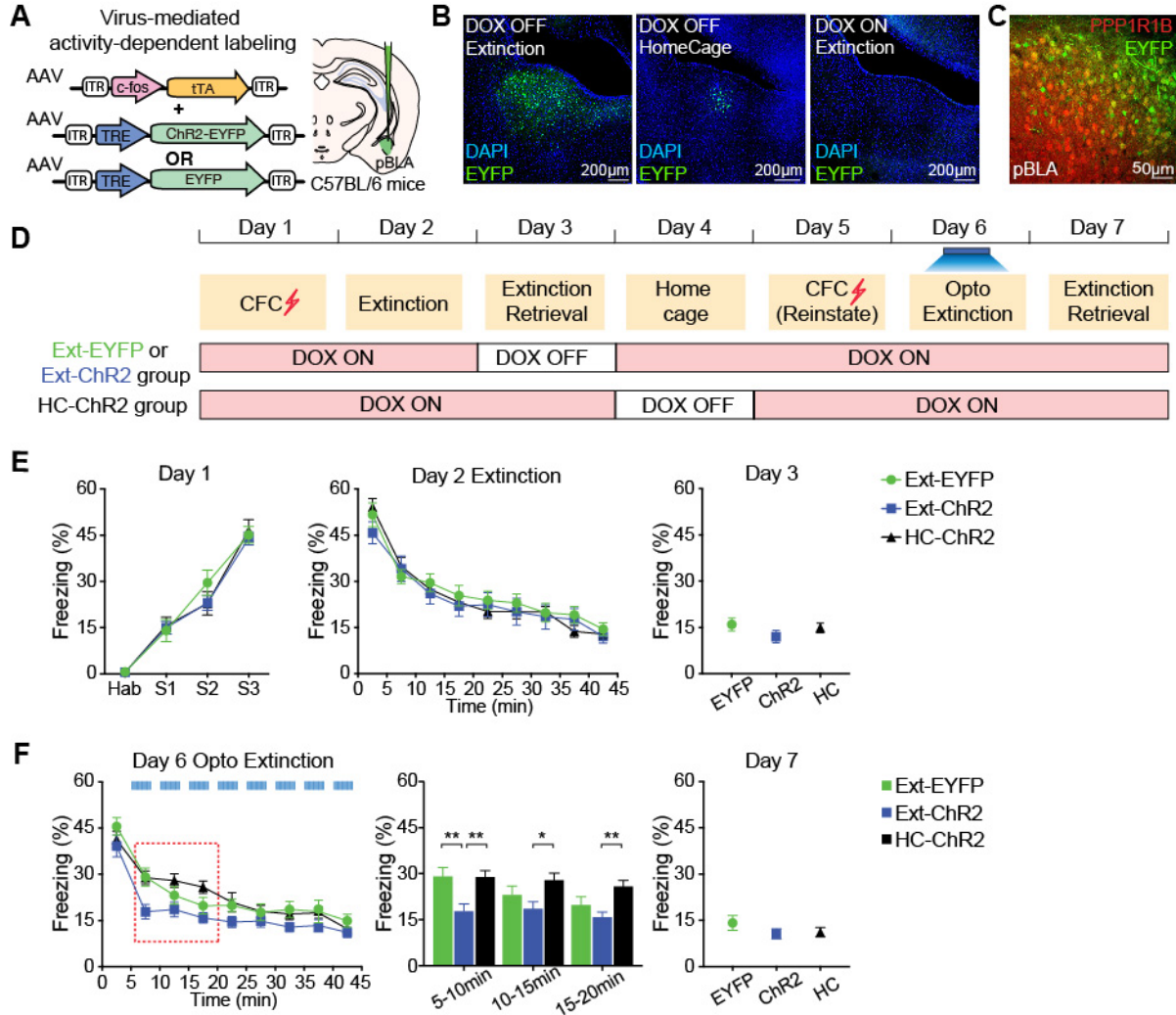
activation of BLA *Rspo2*⁺ neurons impaired fear extinction training and memory. *Rspo2*-Cre-: $n = 14$, *Rspo2*-ChR2: $n = 13$. Two-way RM (Repeated Measures) ANOVA. Single-phase exponential decay fit: $\tau_{Rspo2-Cre-} = 14.06$, $\tau_{Rspo2-ChR2}$ NA. (E) Optogenetic inhibition of BLA *Rspo2*⁺

neurons facilitated early stage of fear extinction training. *Rspo2* Cre-: $n = 12$, *Rspo2*-ArchT: $n = 13$. Two-way RM ANOVA. Single-phase exponential decay fit: $\tau_{Rspo2-Cre-} = 12.43$, $\tau_{Rspo2-ArchT} = 1.65$. (F) Optogenetic activation of BLA *Ppp1r1b*⁺ neurons facilitated fear extinction training.

Ppp1r1b Cre-: $n = 13$, *Ppp1r1b*-ChR2: $n = 9$. Two-way RM ANOVA. Single-phase exponential decay fit: $\tau_{Ppp1r1b-Cre-} = 11.13$, $\tau_{Ppp1r1b-ChR2} = 3.18$. (G) Optogenetic inhibition of BLA *Ppp1r1b*⁺

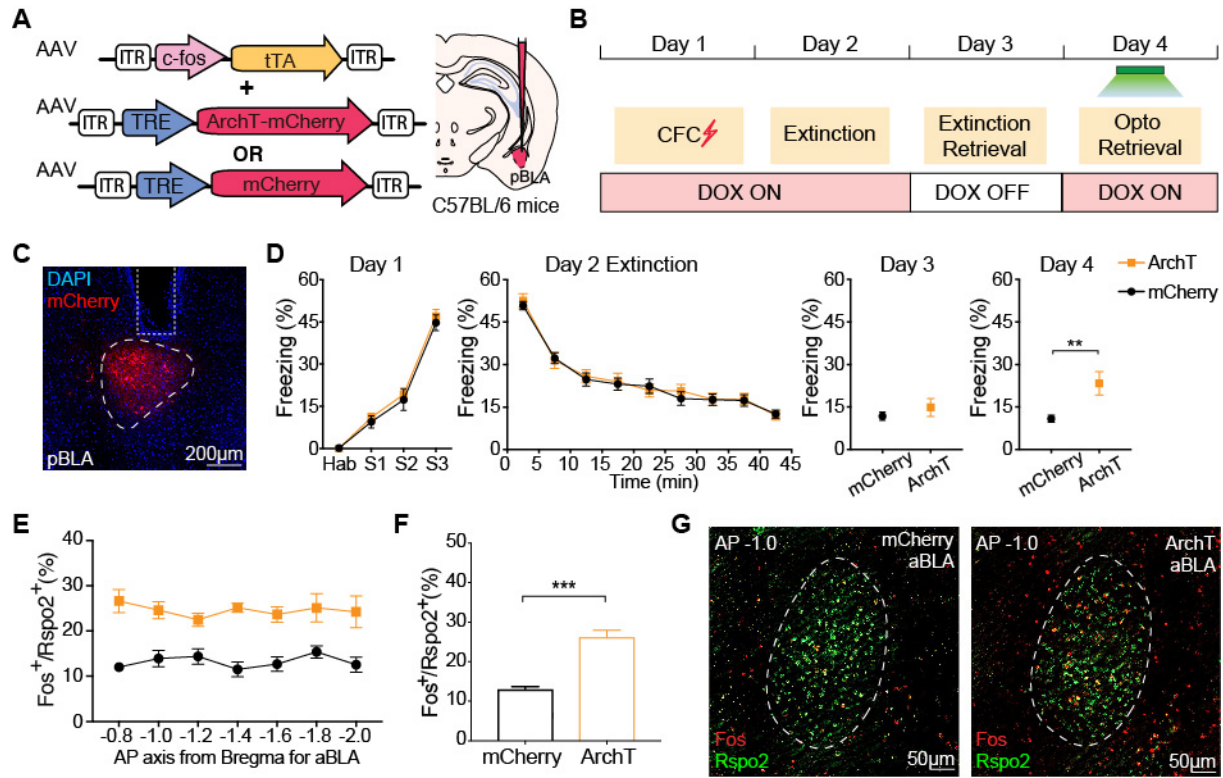
neurons impaired fear extinction training and memory. *Ppp1r1b* Cre-: $n = 10$, *Ppp1r1b*-ArchT: $n = 11$. Two-way RM ANOVA. Single-phase exponential decay fit: $\tau_{Ppp1r1b-Cre-} = 11.34$, $\tau_{Ppp1r1b-ArchT} = 20.36$. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. Data are presented as mean \pm

SEM.



1

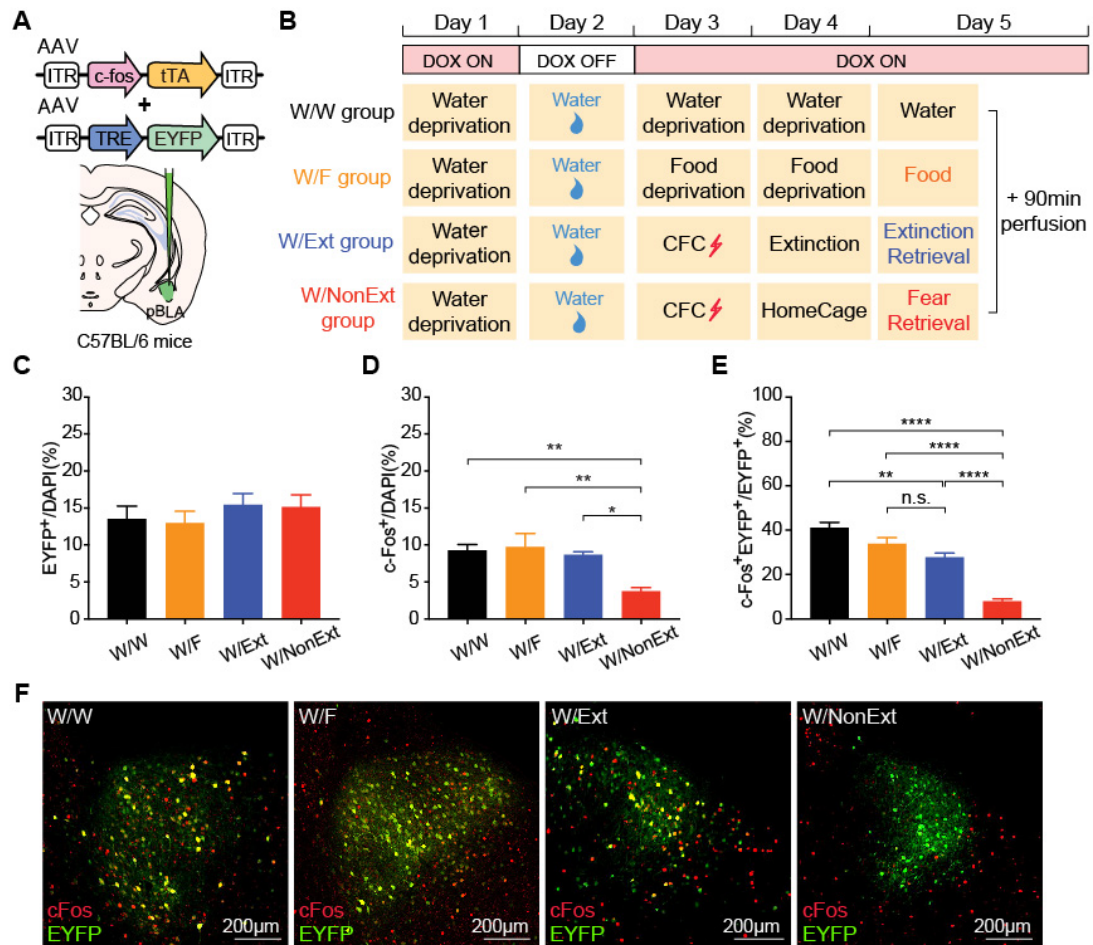
1 **Figure 4. Activation of BLA fear extinction memory engram cells promotes fear extinction**
2 (A) Virus-based activity-dependent labeling scheme. C56BL/6 mice were injected with AAV₉-c-
3 fos-tTA and AAV₉-TRE-ChR2-EYFP or AAV₉-TRE-EYFP virus and implanted with optical
4 fibers that bilaterally target pBLA. (B) Representative images of EYFP expression under three
5 different conditions: Dox OFF + Extinction (left), Dox OFF + Home Cage (middle) and Dox ON
6 + Extinction (right). (C) Image showing 79% of EYFP-expressing cells (green) were PPP1R1B⁺
7 (red) in pBLA. (D) Experimental protocol to label fear extinction engram cells during extinction
8 retrieval in Ext-ChR2 and Ext-EYFP groups. Fear extinction engram neurons were reactivated on
9 Day 6 after reinstatement. In HC-ChR2 control group, ChR2-expressing neurons were labeled
10 when mice were in home cage on Day 4. (E) From Day 1 to Day 3, Ext-EYFP, Ext-ChR2 and HC-
11 ChR2 group showed similar freezing level. Two way RM ANOVA test for Day 1 and Day 2. One-
12 way ANOVA test for Day 3. (F) On Day 6, reactivation of fear extinction engram cells facilitated
13 fear extinction in Ext-ChR2 group. Bar graph showing freezing levels at 5-10 min, 10-15min and
14 15-20 min time slots on Day 6 (marked with red dashed line). Two-way RM ANOVA. Ext-EYFP
15 group $n = 12$; Ext-ChR2 group $n = 16$; HC-ChR2 group $n = 12$. Interaction: $F(16, 296) = 1.566$;
16 $P = 0.077$. Time: $F(8, 296) = 69.65$; $P < 0.001$. Column factor: $F(2, 37) = 4.02$; $P = 0.0263$.
17 Subject (matching): $F(37, 296) = 7.869$; $P < 0.001$. $P(\text{EYFP vs. ChR2}) = 0.0020$; $P(\text{EYFP vs.}$
18 $\text{HC}) = 0.9977$; $P(\text{ChR2 vs. HC}) = 0.0025$.



1

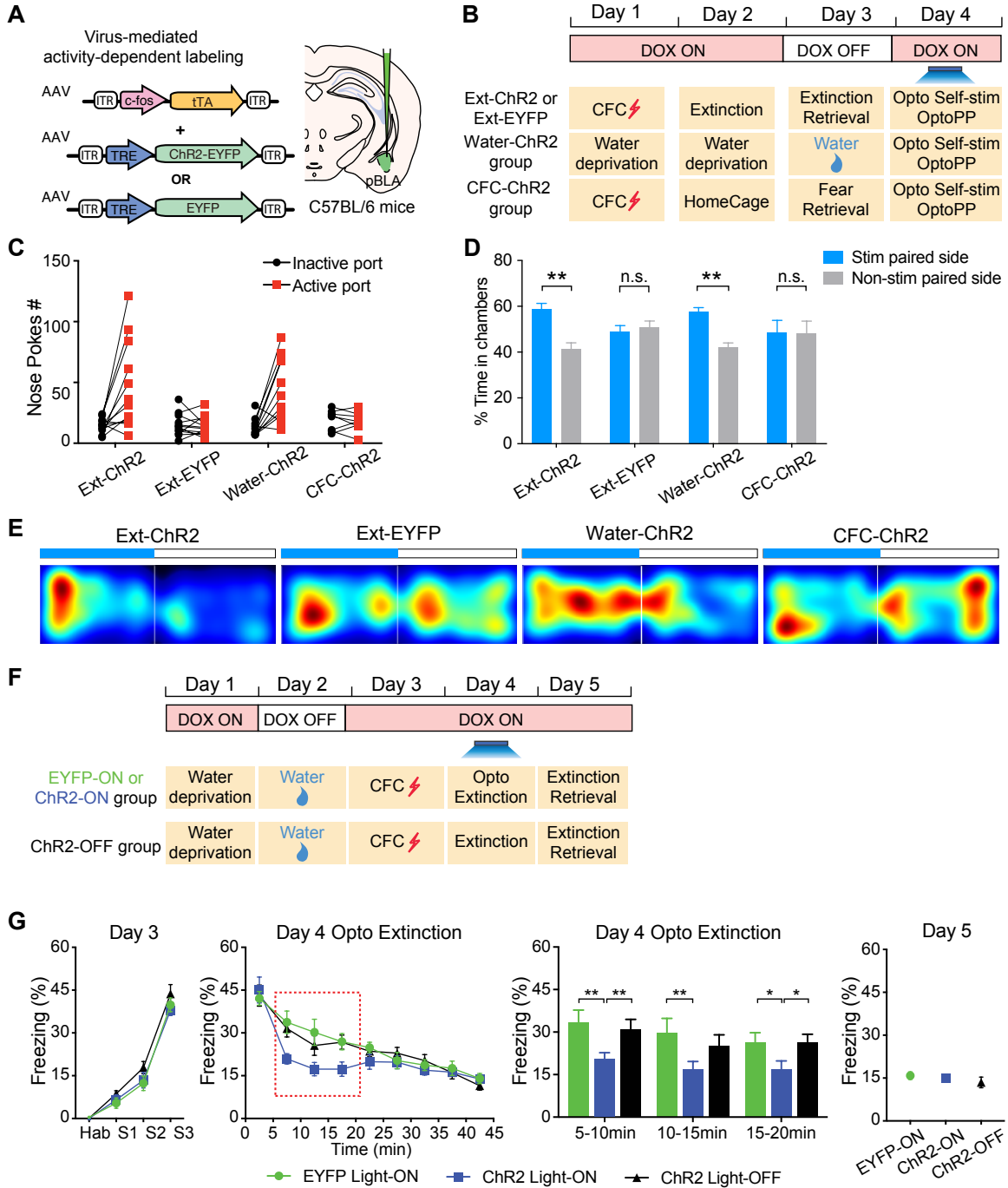
1 **Figure 5. BLA *Ppp1r1b*⁺ fear extinction memory engram cells suppress *Rspo2*⁺ fear cells**

2 (A) Virus-based activity-dependent labeling scheme. C56BL/6 mice were injected with AAV₉-c-
3 fos-tTA and AAV₉-TRE-ArchT-mCherry or AAV₉-TRE-mCherry virus and implanted with
4 optical fibers bilaterally targeting pBLA. (B) Experimental protocol. Fear extinction engram was
5 labeled on Day 3 during extinction retrieval and inhibited on the following day. (C) Representative
6 image showing activity-dependent expression of ArchT-mCherry and optical fiber implant in
7 pBLA. (D) Inhibition of fear extinction engram cells in pBLA caused significant recovery of fear
8 response on Day 4. Unpaired t-test. mCherry group $n = 13$; ArchT group $n = 12$. (E) When pBLA
9 extinction engram neurons were inhibited during extinction retrieval on Day 4, percentage of *Fos*⁺
10 neurons within BLA *Rspo2*⁺ neurons across A/P axis, -0.8 mm to -2.0 mm. (F) Average of the
11 percentage of *Fos*⁺/*Rspo2*⁺ shown in (E). Unpaired t-test. Ext-mCherry group $n=4$; Ext-ArchT
12 group $n=4$. (G) Double smFISH of Fos (red) and *Rspo2* (green) in aBLA of Ext-mCherry mice
13 (left) and Ext-ArchT mice (right). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. Data
14 are presented as mean \pm SEM.



1 **Figure 6. Fear extinction engram neurons and reward neurons overlap extensively on the**
2 **cellular level**

3 (A) Virus-based activity-dependent labeling scheme. C56BL/6 mice were injected with AAV₉-c-
4 fos-tTA and AAV₉-TRE-EYFP. (B) Experimental protocol for activity-dependent labeling of
5 reward-responsive neurons and detect extinction-activated neurons (W/Ext, blue). Water stimulus
6 (W/W, black), food stimulus (W/F, orange) and fear memory retrieval (W/NonExt, red) were used
7 as controls. (C) Similar proportions of neurons were labeled by water reward in all four groups.
8 One-way ANOVA. W/W $n = 5$; W/F $n = 5$; W/Ext $n = 5$; W/NonExt $n = 5$. (D) Percentage of c-
9 Fos⁺ cells in pBLA that were activated by water, food, extinction retrieval or fear retrieval. One-
10 way ANOVA. (E) Reactivation levels (c-Fos⁺EYFP⁺/EYFP⁺) in W/W group, W/F group and
11 W/Ext group were significantly higher than that in W/NonExt group. No significant difference
12 was observed between W/F group and W/Ext group. One-way ANOVA. (F) Double
13 immunofluorescence staining for EYFP (green) and c-Fos (red) in pBLA. * $P < 0.05$, ** $P < 0.01$,
14 **** $P < 0.0001$. Data are presented as mean \pm SEM.



1 **Figure 7. Fear extinction engram neurons and reward neurons are behaviorally**
2 **interchangeable**

3 (A) Virus-based activity-dependent labeling scheme. C56BL/6 mice were injected with AAV₉-c-
4 fos-tTA and AAV₉-TRE-ChR2-EYFP or AAV₉-TRE-EYFP virus and implanted with optical
5 fibers bilaterally targeting pBLA. (B) Experimental protocol to label and activate fear extinction
6 engram neurons (Ext-ChR2 or Ext-EYFP group), water-reward neurons (Water-ChR2 group) and
7 fear retrieval engram neurons (CFC-ChR2 group) in optogenetic self-stimulation behavior or
8 optogenetic place preference (OptoPP) behavior on Day 4. (C) 50% of Ext-ChR2 group mice and
9 67% of Water-ChR2 group mice exhibited self-stimulation behavior, with a threshold of (active
10 ports–inactive ports) ≥ 20 . No self-stimulation behavior was exhibited in Ext-EYFP or CFC-ChR2
11 groups. Ext-ChR2 $n = 12$; Ext-EYFP $n = 12$; Water-ChR2 $n = 11$; CFC-ChR2 $n = 8$. (D) Ext-ChR2
12 and Water-ChR2 group showed preference towards stimulation paired side. Ext-ChR2 $n = 12$; Ext-
13 EYFP $n = 11$; Water-ChR2 $n = 12$; CFC-ChR2 $n = 11$. Paired t-test. (E) Heatmaps showing time
14 spent in each side of the chamber. Blue bar showing the stimulation paired side. (F) Experimental
15 protocol to activate reward neurons during fear extinction (Day 4). Optical laser was off for ChR2-
16 OFF group during extinction training on Day 4. Optogenetic activation of water-reward neurons
17 accelerated fear extinction learning. EYFP Light-ON, $n = 8$; ChR2 Light-ON, $n = 11$; ChR2 Light-
18 Off, $n = 10$. Two-way RM ANOVA. $*P < 0.05$, $**P < 0.01$, $****P < 0.0001$. Data are presented
19 as mean \pm SEM.

1
2
3
4
5
6
7
8

Supplementary Materials for

Amygdala Reward Neurons Form and Store Fear Extinction Memory

Xiangyu Zhang¹, Joshua Kim¹ & Susumu Tonegawa^{1,2*}

Correspondence to: tonegawa@mit.edu

1 **Materials and Methods**

2 **Subjects**

3 The Ppp1r1b-Cre (also called Cartpt-Cre) mice were obtained from GENSAT (stock
4 number: 036659-UCD). The Rspo2-Cre mice were generated in house. Cre-expressing mice were
5 genetic knock-in mice or have been previously been validated for genetic specificity (16). All Cre
6 transgenic mice were bred using a heterozygous male with females of C57BL/6 background and
7 maintained as heterozygous. Mice had access to food and water *ad libitum* and were socially
8 housed in numbers of two to five littermates until surgery. After surgery, mice were singly
9 housed. 2- to 6-month old male mice were used for all experiments. For engram labeling, C57/B6
10 mice were raised on food containing 40 mg kg⁻¹ DOX 1 week before stereotaxic surgery and were
11 maintained on Dox diet for the remaining experiment period except for the target Dox-OFF days.
12 All mice were maintained and cared in accordance with protocols approved by the Massachusetts
13 Institute of Technology (MIT) Committee on Animal Care (CAC) and guidelines by the National
14 Institutes of Health (NIH).

15

16 **Viral constructs**

17 AAV₅-Efl1a-DIO-eArch3.0-EYFP (AV5257), AAV₅-Efl1a-DIO-ChR2-EYFP (AV5226B),
18 and AAV₅-Efl1a-DIO-eYFP (AV4310D) were obtained from the University of North Carolina at
19 Chapel Hill Vector Core. AAV₉-c-fos-tTA, AAV₉-TRE-EYFP, AAV₉-TRE-ChR2-EYFP, AAV₉-
20 TRE-mCherry, and AAV₉-TRE-eArchT-mCherry were generated by and acquired from the Gene
21 Therapy Center and Vector Core at the University of Massachusetts Medical School as previously
22 described (Kitamura et al., 2017; Liu et al., 2012; Redondo et al., 2014). The AAV₅-hsyn:DIO-
23 GCaMP6f virus was generated by and acquired from the University of Pennsylvania Vector Core.

1 **Stereotactic surgery**

2 Mice undergoing stereotactic injections were anesthetized under isoflurane. Standard
3 stereotactic procedures were used. Viruses were injected using a mineral oil-filled glass
4 micropipette attached to a 1 μ l Hamilton microsyringe. Virus was injected at a speed of 80 nl/min,
5 which was controlled by a microsyringe pump. The needle was lowered to the target site and
6 remained for 5 min after the injection. The incision was closed with sutures. Mice were given 0.1
7 mg/kg slow release Buprenex as analgesic and remained on a heating pad until full recovery from
8 anesthesia. The histology of all injections was verified after behavior experiments.

9 For optogenetic behavior experiment, 200 μ L of AAV₅ Cre-dependent virus was bilaterally
10 injected into the anterior BLA of Rspo2-Cre mice (AP -1.4 mm, ML \pm 3.4 mm, DV -4.9 mm) and
11 the posterior BLA of Ppp1r1b-Cre mice (AP -2.0 mm, ML \pm 3.4 mm, DV -4.9 mm) and incubated
12 for 3 – 4 weeks before behavioral experiments.

13 For activity-dependent labeling and activation/inhibition of pBLA engram cells, 200 nL
14 of $\sim 2.0 \times 10^9$ GC of AAV₉-cFos-tTA and AAV₉-TRE-ChR2-EYFP or AAV₉-TRE-eArch3.0-
15 mCherry (1:1 mixture) were bilaterally injected into the pBLA (AP -2.0 mm, ML \pm 3.4 mm, DV
16 -4.9 mm) of doxycycline (Dox)-fed C57/B6 mice and incubated for 7 days before subsequent
17 experiments. AAV₉-TRE-EYFP or AAV₉-TRE-mCherry was used as corresponding controls.

18 For optical implant surgery, two optic fibers (200 μ m core diameter, 5.5 mm length; Doric
19 Lenses) were bilaterally lowered above the injection sites (aBLA: -1.4 mm AP, \pm 3.4 mm ML,
20 -4.7 mm DV; pBLA: -2.0 mm AP, \pm 3.4 mm ML, -4.7 mm DV). The implants were secured to
21 the skull with two jewelry screws and adhesive cement (C&B Metabond). A protective cap, made
22 using a 1.5 mL black Eppendorf tube, was fixed onto the implant using dental cement (Kim et al.,
23 2016; Kim et al., 2017).

1 For Ca²⁺ imaging surgery, AAV₅-hsyn:DIO-GCaMP6f was unilaterally injected into right
2 aBLA of Rspo2-Cre mice and right pBLA of Ppp1r1b-Cre mice with the following coordinates:
3 aBLA was targeted at -1.4 mm AP, +3.4 mm ML, -4.9 mm DV and pBLA was targeted at
4 -2.0 mm AP, +3.4 mm ML, -4.9 mm DV. Two weeks following AAV₅-hsyn:DIO-GCaMP6f virus
5 injection, a GRIN lens (0.5 mm diameter, 8 mm length; Inscopix) was implanted targeting aBLA
6 or pBLA (aBLA: -1.4 mm AP, +3.4 mm ML, -4.7 mm DV; pBLA: -2.0 mm AP, +3.4 mm ML,
7 -4.7 mm DV).

8

9 **Contextual fear extinction behavior protocol**

10 Behavioral context was 29 × 25 × 22 cm chambers (Med Associates) with grid floors,
11 opaque ceilings, white lighting, and scented with 5% benzaldehyde. Before fear conditioning, mice
12 were habituated to investigator handling for 5 min on three consecutive days in the holding room
13 where the mice were housed. On Day 1, mice were habituated in the behavioral context for 3 min,
14 followed by three footshocks (0.60 mA, 2 s) delivered at 180 s, 250 s and 320 s. Mice remained in
15 the behavior chamber for 80 s after the third footshock and then returned to their home cages in
16 the holding room. 24 hrs after fear conditioning, mice returned to the fear conditioning chambers
17 to receive 45 min extinction training without any footshocks. Then mice were put back to their
18 home cages for another 24 hrs. On Day 3, mice were placed in the conditioned context for 5 min
19 extinction retrieval test. Behavior videos were recorded with VideoFreeze software and freezing
20 level was scored manually by experimenters who were blinded to conditions.

21

22 **Single molecular fluorescent *in situ* hybridization**

1 To examine the expression of *Fos* gene (Figures 1 and 5), single molecule fluorescence *in*
2 *situ* hybridization (smFISH) was performed using RNAscope Fluorescent Multiplex Kit
3 (Advanced Cell Diagnostics, ACDBio) as previously described (16, 18). Isoflurane anesthetized
4 mice were decapitated, their brains harvested and flash frozen on aluminum foil on dry ice. Brains
5 were stored at -80°C . Prior to sectioning, brains were equilibrated to -16°C in a cryostat for
6 30 min. Brains were coronally sectioned at $20\ \mu\text{m}$ with cryostat and thaw-mounted onto Superfrost
7 Plus slides (25 x 75 mm, Fisherbrand). Sections from a single brain were serially thaw-mounted
8 onto 10 slides through the entire BLA (anterior-posterior distance from Bregma, $-0.8\ \text{mm}$ to
9 $-2.6\ \text{mm}$). Slides were air-dried for 60 min at room temperature prior to storage at -80°C . smFISH
10 probes for all genes examined were obtained from ACDBio, *Fos* (Cat #421981), *Ppp1r1b* (Cat
11 #405901) and *Rspo2* (Cat #402001). Slides were counterstained for the nuclear marker DAPI using
12 ProLong Diamond Antifade mounting medium with DAPI (ThermosFisher) (Kim et al., 2016;
13 Kim et al., 2017).

14

15 **Immunohistochemistry**

16 Mice were perfused transcardially with phosphate buffered saline (PBS), followed by 4%
17 paraformaldehyde (PFA). Brains were post-fixed with 4% PFA solution overnight and then
18 sectioned coronally $50\ \mu\text{m}$ at with a vibratome. Free floating brain sections were washed in PBST
19 (PBS, 3% TritonX) three times for 10 min, blocked for 1 hr in blocking buffer (PBST, 5% normal
20 goat serum), and incubated in primary antibody in blocking buffer overnight at 4°C . On the next
21 day, brains were washed with three 10 min washes of PBST and incubated in secondary antibody
22 in blocking buffer at room temperature for 2 hrs. Primary antibodies used were chicken anti-GFP
23 (Invitrogen, A10262, 1:1,000), rabbit anti-c-Fos (Santa Cruz, sc-52, 1:1,000) and rabbit anti-

1 PPP1R1B (Abcam, ab40801, 1:1,000). Secondary antibodies used were goat anti-chicken Alexa
2 Fluor 488 (Invitrogen, A11039, 1:1,000) and goat anti-rabbit Alexa Fluor 555 (Invitrogen,
3 A21428, 1: 1,000). After three more 10-min PBST washes, slides were covered with coverslips
4 and mounted using VectaShield mounting solution containing DAPI (Vector Laboratories).

6 **Calcium imaging**

7 Calcium imaging of BLA *Rspo2*⁺ neurons and *Ppp1r1b*⁺ neurons was performed on
8 *Rspo2*-Cre and *Ppp1r1b*-Cre mice, respectively. AAV₅-hsyn:DIO-GCaMP6f was unilaterally
9 injected into right aBLA of *Rspo2*-Cre mice and right pBLA of *Ppp1r1b*-Cre mice with the
10 following coordinates: aBLA was targeted at -1.4 mm AP, +3.4 mm ML, -4.9 mm DV and pBLA
11 was targeted at -2.0 mm AP, +3.4 mm ML, -4.9 mm DV. Two weeks following AAV₅-hsyn:DIO-
12 GCaMP6f virus injection, a GRIN lens (0.5 mm diameter, 8 mm length; Inscopix) was implanted
13 targeting aBLA or pBLA (aBLA: -1.4 mm AP, +3.4 mm ML, -4.7 mm DV; pBLA: -2.0 mm AP,
14 +3.4 mm ML, -4.7 mm DV). Two weeks after GRIN lens implant, a baseplate for the miniaturized
15 microscope camera was attached above the GRIN lens using ultraviolet-light curable glue (Loctite
16 4305) (Mukamel et al., 2009; Ziv et al., 2013).

17 For contextual fear extinction behavior, recording was performed during the light cycle.
18 Inscopix endoscope and calcium events were captured at 20 Hz with an Inscopix miniature
19 microscope (nVista HD). Inscopix endoscope was synchronized with Med Associate behavior
20 chamber via a TTL signal, which was used for all the other fear extinction behavior experiments
21 in this study. Mice underwent the three-day contextual fear extinction procedure as previously
22 described (Figure 1). During contextual fear conditioning (CFC) on Day 1, freezing behavior and
23 Ca²⁺ signals of subjects were recorded for 3 min habituation and 3 min when three footshocks were

1 delivered. During 45 min fear extinction training, Ca^{2+} signals was recorded during the 0-3 min,
2 5-8 min, 10-13 min, 15-18 min, 20-23 min, 25-28 min, 30-33 min, 35-38 min and 40-43 min
3 timestamps. The 0-3 min epoch was considered as fear retrieval stage. The other eight epochs
4 together were considered as extinction training stage. In the heatmap and Ca^{2+} traces (Figure 2E-
5 2I), extinction training referred referred to 40-43 min time bin recording on Day 2. Recorded
6 calcium-imaging movies were first processed with ImageJ to subtract base line activity (dividing
7 each image by a low-pass, $r = 20$ pixels, and high pass, $r = 1000$ pixels), followed by motion
8 correction using Inscopix Mosaic software (Mukamel et al., 2009). Subsequently, $\Delta F/F$ ((F-
9 F0)/F0) signal was calculated in Mosaic software (Inscopix), in which F0 was calculated over the
10 entire recording session. A stacked image of $\Delta F/F$ recording movie was acquired by maximum
11 intensity projection. Cell identifications were semi-automatically identified and cross-registered
12 for longitudinal recording by visual inspection (Kitamura et al., 2017). The corresponding Ca^{2+}
13 activity was extracted by applying PCA/ICA analysis in Mosaic software (Inscopix). In total, we
14 recorded 179 BLA *Ppp1r1b*⁺ cells from 5 *Ppp1r1b*-Cre mice and 169 BLA *Rspo2*⁺ cells from 4
15 *Rspo2*-Cre mice.

16

17 **Data analysis**

18 Ca^{2+} events were detected with a threshold 15% $\Delta F/F$. We calculated the average time of
19 Ca^{2+} events, i.e., the sum of all Ca^{2+} events duration/recording time. To identify cells with
20 increased or decreased activity in response to footshocks during CFC on Day 1, we compared the
21 mean $\Delta F/F$ of 3 min after delivery of footshocks with that of 3 min habituation stage on Day1. To
22 identify cells with changed neuronal activities during fear retrieval, we compared the mean $\Delta F/F$
23 of the fear retrieval stage (0-3 min) on Day 2 with that of the 3 min habituate period before

1 footshocks on Day 1. To identify cells with increased or decreased activity during fear extinction
2 training, we compared the mean $\Delta F/F$ of extinction training with that of fear retrieval stage (0-3
3 min) (12). To identify cells with increased or decreased activity during extinction retrieval, we
4 compared the mean $\Delta F/F$ of 5 min extinction retrieval on Day 3 with that of the 3 min fear retrieval
5 stage on Day 2. For each comparison, after shuffling all individual Ca^{2+} transients 5000 times
6 between the two recording sessions to be compared, a distribution of $\Delta F/F$ difference was
7 generated. If the actual difference between mean $\Delta F/F$ of two recording sessions was higher than
8 97.5% of the distribution, then the cell was considered as significantly activated (Up cells). If the
9 actual difference between mean $\Delta F/F$ of two recording sessions was lower than 2.5% of the
10 distribution, then the cell was considered as significantly inhibited (Down cells). The others were
11 considered as no change cells. All the data were analyzed using custom code written in MATLAB
12 (Mathworks)

13

14 **Cell counting**

15 Images of smFISH counting were acquired using a Zeiss AxioImager.Z2 under a 10X
16 objective. Consecutive sections with 0.2 mm interval (anterior-posterior distance from Bregma,
17 -0.8 mm to -2.6 mm) were used for quantification. Neurons stained for *Fos* (red) and against gene
18 marker (green), *Rspo2* or *Ppp1r1b*, were counted using image analysis software (HALO, Indica
19 Labs), which was compatible with RNAscope Fluorescent Multiplex Kit.

20 Images of immunohistochemistry counting were acquired using confocal microscopy
21 (Zeiss LSM700) under a 10X objective. Maximum intensity projections were generated using ZEN
22 Blue software (Zeiss). Neurons stained against GFP (green), c-Fos (red), mCherry (red), PP1R1B

1 (green or red) and DAPI (blue) were automatically using cell counting software (ImageJ) with
2 manual adjustment of detection thresholds.

3 In Figure 6, For DAPI counts, thresholds were set to only include larger nuclei and to
4 exclude non-neuronal small nuclei as previously described. The region of interest (ROI) was
5 defined manually based on anatomical landmarks. All counting was performed blind as to the
6 group and condition that the specimen belonged to.

7 8 **Optogenetic activation/inhibition**

9 To apply optogenetic activation/inhibition during fear extinction in Figure 3, mice were
10 put back to conditioned behavior chamber to receive 45 min fear extinction training. After the first
11 5 min, the mice received 8 cycles of 3 min optogenetic activation (450-470 nm, 8-12 mW, 20 Hz)
12 or 3 min optical inhibition (520-550 nm, 8-12 mW, constant) with 2 min interval. 450 nm and 520
13 nm lasers were generated with LED drivers (Doric).

14 In Figure. 1J-IM, AAV9-DIO-ArchT-EYFP virus was injected into pBLA of Ppp1r1b-Cre
15 mice or Ppp1r1b-Cre-negative mice as control. During fear extinction retrieval on Day 3, the mice
16 received 3 cycles of 1 min optical inhibition (520-550 nm, 8-12 mW, constant) with 1 min interval.

17 18 **Labeling and optogenetic stimulation of fear extinction engram (Related to Figure 4 and 5)**

19 To label fear extinction engram neurons in Figure 4, we injected AAV₉-c-Fos-tTA together
20 with AAV₉-TRE-ChR2-EYFP or AAV₉-TRE-EYFP into the pBLA of C57BL/6 mice, and
21 subjected them to a series of engram labeling and behavioral steps (Figure 4D). This genetic
22 manipulation permits expression of ChR2-EYFP or EYFP in the cells that were activated during
23 fear extinction memory retrieval through the transcriptional promoter of c-Fos gene when

1 doxycycline (Dox) was removed from the diet. Mice were kept on Dox diet after surgery. In Ext-
2 ChR2 group and Ext-EYFP group, mice were switched to Dox-OFF diet for 24 hrs following fear
3 extinction training on Day 2. 60 min after extinction retrieval test on Day 3, mice returned to Dox-
4 ON diet and stayed at home cage on Day 4. In HC-ChR2 group, mice stayed on Dox diet
5 throughout the first round of contextual fear extinction protocol and were switched to Dox-OFF
6 diet for 24 hrs and stayed at home cage on Day 4. On Day 5, mice in all three groups received one
7 immediate footshock (0.75 mA) in the original conditioned context for fear reinstatement. On Day
8 6, mice received opto-extinction protocol, in which 8 cycles of 3 min-long blue lasers (470 nm, 20
9 Hz, 8-12 mW) were delivered after the first 5 min. On Day 7, mice were put back to the behavior
10 context for 5 min retrieval test.

11 In Figure 5, fear extinction engram neurons were labeled in the same fashion as Figure 4,
12 except that the fear extinction engram neurons expressed ArchT-mCherry or mCherry. On Day 4,
13 mice were put back to the conditioned context for 5 min opto-extinction retrieval test, in which
14 mice received three green laser stimulation (520-550 nm, 8-12 mW, constant) at 0-1 min, 2-3 min
15 and 4-5 min time points. For Fos quantification experiment in Figure 5E-5G, mice were sacrificed
16 and the brain were harvested and subjected to smFISH procedure.

17

18 **Labeling and reactivation of reward responsive neurons (Related to Figure 6B, 7B-7G)**

19 To label reward responsive neurons in Figure 6B, One-week post-surgery, mice were
20 deprived of water for 24 hrs while on Dox diet. Water-deprived mice were given regular chaw
21 food without Dox (Dox-OFF diet) for 24 hrs, during which water was given as a reward. 60 min
22 after drinking water, subjects returned to Dox-ON diet to close the labeling window.

1 After being back to Dox-ON diet, mice were divided into four different groups. Mice in
2 the W/W (Water/Water) group remained on water deprivation for subsequent three days. On Day
3 5, W/W group mice received *ad libitum* water as a reward 90 min before perfusion. Mice in the
4 W/F (Water/Food) group were deprived of food for subsequent three days and received *ad*
5 *libitum* food as a reward 90 min before perfusion. During food deprivation, mice received water
6 that contained 200 ul/L Doxycycline to remain on Dox. Mice in the W/Ext (Water/Extinction)
7 group went through three-day contextual fear extinction protocol and were perfused 90 min after
8 extinction retrieval on Day 5. Mice in the W/NonExt group (Water/NonExtinction) group went
9 through the same procedure as the W/Ext group except that they stayed in home cage (HC) instead
10 of extinction training on Day 4. After perfusion, brains of these four groups of mice were sliced
11 and stained for c-Fos staining followed by quantification.

12 In Figure 7F, water reward responsive neurons were labeled with EYFP or Chr2-EYFP
13 depending on the virus injection combination. Mice went through three-day contextual fear
14 extinction protocol. For EYFP-ON group and Chr2-ON group, mice received opto-extinction
15 protocol, in which 8 cycles of 3 min-long blue lasers (470 nm, 20 Hz, 8-12 mW) were delivered
16 after the first 5 min of extinction training. For Chr2-OFF group, mice didn't receive any
17 optogenetic stimulation during extinction training.

18

19 **Optogenetic self-stimulation test (Related to Figure 7B and 7C)**

20 The mice were water restricted overnight before the experiment. The optogenetic self-
21 stimulation test was conducted over 60 min session on the following day and no prior training was
22 required. The water-restricted mice were placed in an operant conditioning chamber (Med
23 Associates) equipped with two nose ports (16, 18). At the beginning of each trial, each nose port

1 contained 0.05 ml sugar water to initiate the nose poke. No additional water reward was given
2 throughout the trial. One of the two nose ports was randomly assigned as active port, which could
3 deliver 5 s duration of optical stimulation (8-12 mW, 20 Hz, 473 nm) upon nose poke, while the
4 other one did not (inactive). The optical laser was generated through a 470 nm LED light source
5 (XLED1, Lumen Dynamics). The number of nose pokes of active port and inactive port were
6 recorded with Med-PC IV software during 60 min test session. If the difference between active
7 pokes and inactive pokes, (active-inactive), was larger than 20 pokes, it was counted as self-
8 stimulation behavior.

9

10 **Optogenetic place preference (Related to Figures 7B, 7D and 7E)**

11 Mice were placed in a custom-made behavioral arena (L45 × W15.5 × H30 cm, white
12 plexiglass), where each side of the box contained distinct wall cues. Each trial was 20 min and one
13 side was randomly assigned as the stimulation side for each trial. Mice were placed in the non-
14 stimulated side at the onset of the experiment. Each time the mouse crossed to the stimulation side
15 of the chamber, mice received continuous 20 Hz laser stimulation (8-12 mW, 473 nm) until the
16 mouse crossed back into the non-stimulation side. Behavioral data were recorded and analyzed
17 with EthoVision XT video tracking software (Noldus Information Technologies).

18

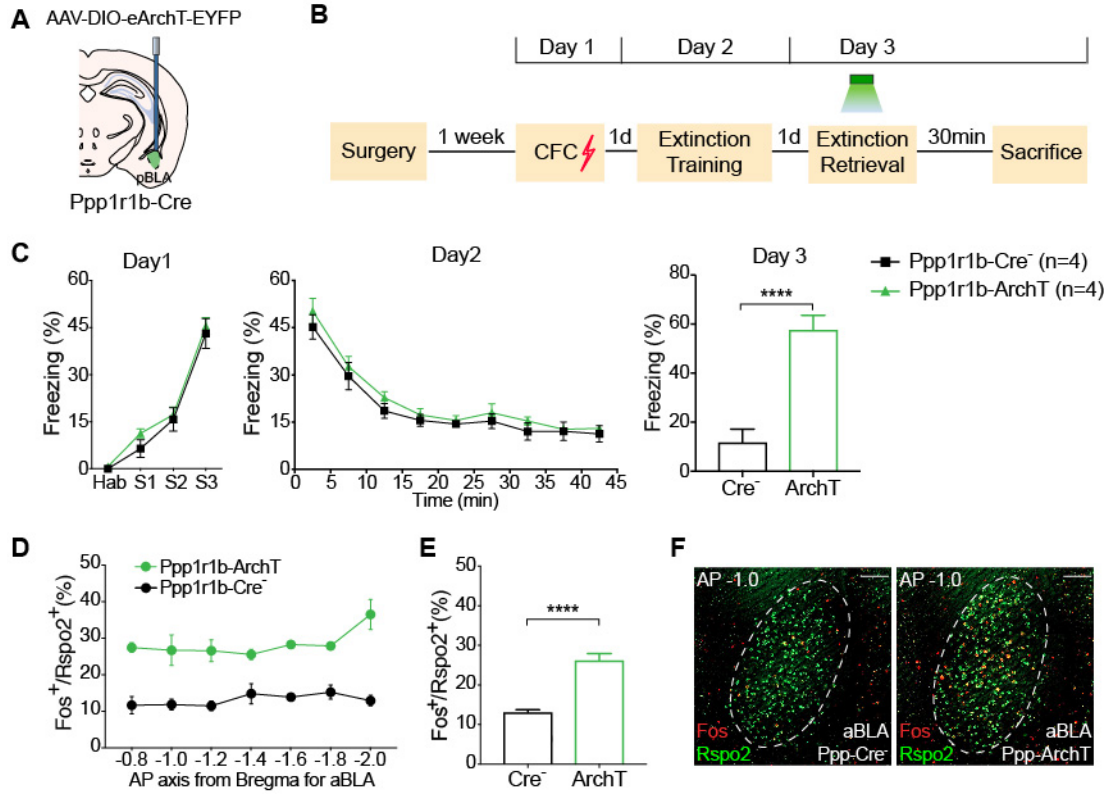
19 **Systemic injection of Kainic acid (Related to Figure S4A and S4B)**

20 To induce seizure, mice were injected intraperitoneally with 15 mg kg⁻¹ kainic acid (KA,
21 Tocris Cat. No 0222) after being on Dox off diet for 24 hrs. 3 hrs after KA injection, mice were
22 returned to Dox diet and perfused next day for immunohistochemistry staining.

23

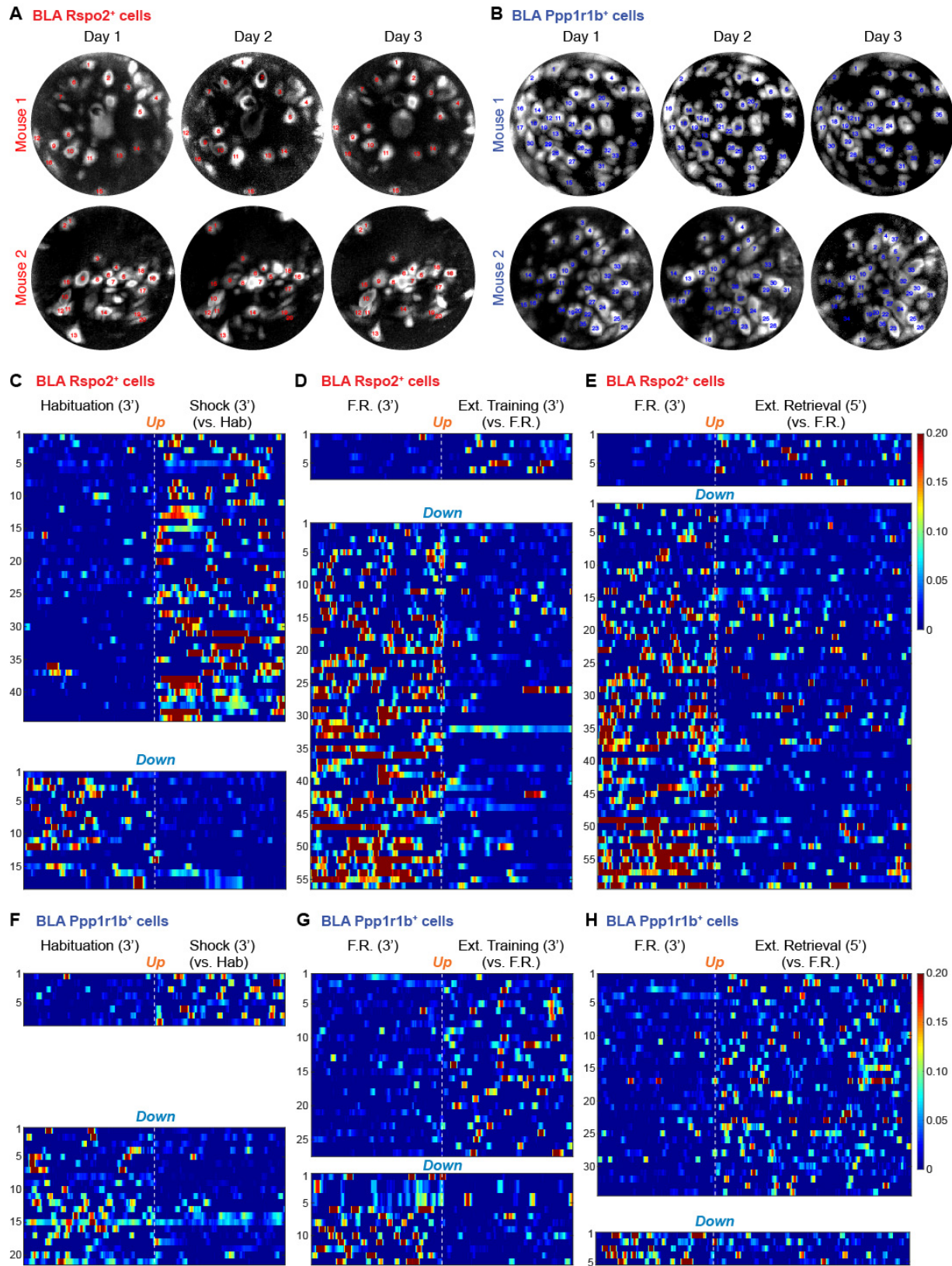
1 **Statistical Analysis**

2 GraphPad Prism (version 7.0c for Mac OS X, GraphPad Software, La Jolla California
3 USA) was used for statistical analysis. All data are presented as mean \pm SEM. *n* indicates number
4 of animals or number of cells. Comparisons between two-group data were analyzed by two-tailed
5 unpaired t-test or two-tailed paired t-test. Multiple group comparisons were assessed using a One-
6 way ANOVA or Two-way RM (Repeated Measures) ANOVA, followed by the Tukey's multiple
7 comparison test when significant main effects or interactions were detected. The null hypothesis
8 was rejected at the $P < 0.05$ level. Data met assumptions of statistical tests. Sample sizes were
9 chosen on the basis of previous studies (Kim et al., 2016; Kim et al., 2017; Liu et al., 2012; Nishi
10 et al., 1997).



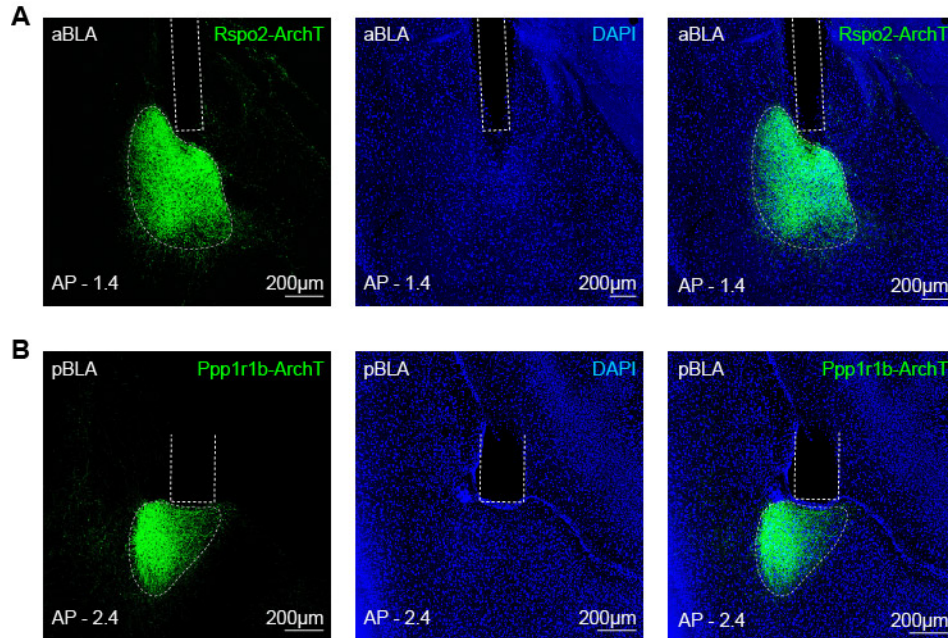
1

1 **Figure S1. Inhibition of BLA *Ppp1r1b*⁺ neurons during extinction retrieval leads to**
2 **increased *Rspo2*⁺ neuronal activity (Related to Figure 1)**
3 (A) Schematic diagram of bilateral injection of AAV₉-DIO-eArchT-EYFP and optical fiber
4 implant in pBLA of Ppp1r1b-Cre mice or Ppp1r1b-Cre⁻ control mice. (B) Experimental protocol.
5 BLA *Ppp1r1b*⁺ neurons were inhibited during extinction retrieval on Day 3 and mice were
6 sacrificed 30 min after extinction retrieval test for smFISH staining. (C) Inhibition of *Ppp1r1b*⁺
7 neurons led to increased freezing level during extinction retrieval on Day 3. Ppp1r1b-Cre- and
8 Ppp1r1b-ArchT mice exhibited similar freezing levels during CFC on Day 1 and extinction
9 training on Day 2. Unpaired t-test. Ppp-Cre⁻ *n* = 4; Ppp-ArchT *n* = 4. (D) When BLA *Ppp1r1b*⁺
10 neurons were inhibited during extinction retrieval on Day 3, percentage of *Fos*⁺ neurons within
11 BLA *Rspo2*⁺ neurons across A/P axis, -0.8 mm to -2.0 mm. (E) Average of the percentage of
12 *Fos*⁺/*Rspo2*⁺ shown in (D). Unpaired t-test. (F) Double smFISH of Fos (red) and *Rspo2* (green)
13 in aBLA in Ppp-Cre⁻ mice (left) and Ppp-ArchT mice (right). **P* < 0.05, ***P* < 0.01, ****P* < 0.001,
14 *****P* < 0.0001. Data are presented as mean ± SEM. Scale bars: 200 μm (M)



1

1 **Figure S2. Longitudinal recording of BLA *Rspo2*⁺ and *Ppp1r1b*⁺ cells (Related to Figure 2)**
2 (A) Stacked field of view (FOV) images of BLA *Rspo2*⁺ neurons acquired on Day 1 CFC, Day 2
3 Extinction training and Day 3 Extinction retrieval. Same cell numbers within each mouse are
4 tracked across three days. (B) Stacked FOV images of BLA *Ppp1r1b*⁺ neurons acquired on Day 1
5 CFC, Day 2-Extinction training and Day 3-Extinction retrieval. Same cell numbers within each
6 mouse are tracked across three days. (C) Heat maps of Ca²⁺ responses of BLA *Rspo2*⁺ neurons
7 that were activated (top-Up, *n* = 44) and inhibited (bottom-Down, *n* = 18) by shocks during CFC
8 on Day 1. (D) Heat maps of Ca²⁺ responses of BLA *Rspo2*⁺ neurons that were activated (top-Up,
9 *n* = 7) and inhibited (bottom-Down, *n* = 56) during extinction training. Extinction training shows
10 the recording of 35-38 min timestamp on Day 2. (E) Heat map of Ca²⁺ responses of BLA *Rspo2*⁺
11 neurons that were activated (top-Up, *n* = 8) and inhibited (bottom-Down, *n* = 59) during extinction
12 retrieval. (F) Heat map of Ca²⁺ responses of BLA *Ppp1r1b*⁺ neurons that were activated (top-Up,
13 *n* = 8) and inhibited (bottom-Down, *n* = 21) after shocks during CFC on Day 1. (G) Heat map of
14 Ca²⁺ responses of BLA *Ppp1r1b*⁺ neurons that were activated (top-Up, *n* = 28) or inhibited
15 (bottom-Down, *n* = 14) during extinction training. Extinction training shows the recording of 35-
16 38 min timestamp on Day 2. (H) Heat map of Ca²⁺ responses of BLA *Ppp1r1b*⁺ neurons that were
17 activated (top-Up, *n* = 34) or inhibited (bottom-Down, *n* = 5) during extinction retrieval.

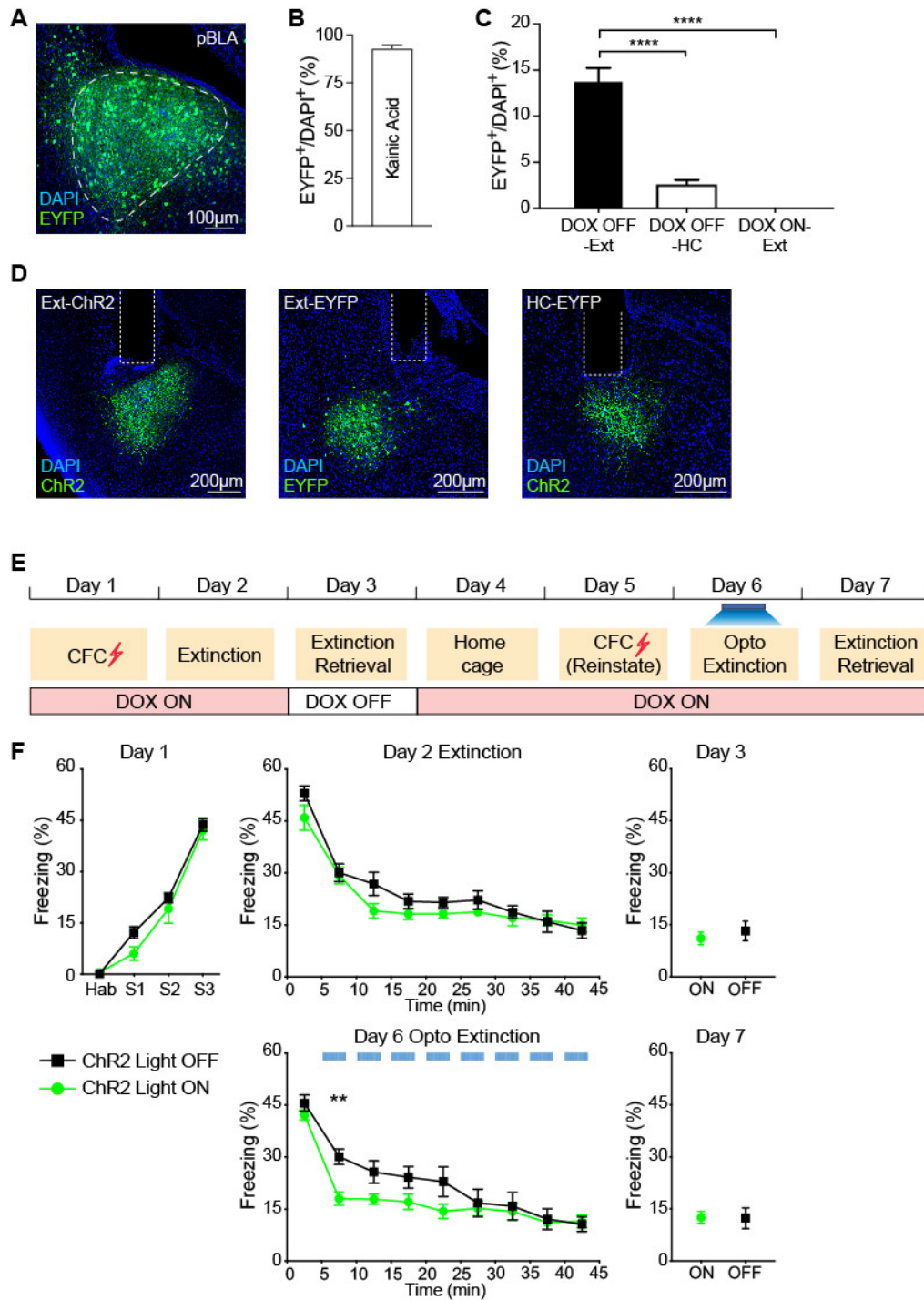


1

2 **Figure S3. Expression of eArchT-EYFP and fiber placement for targeting BLA $Rspo2^+$ and**
3 **$Ppp1r1b^+$ neurons (Related to Figure 3)**

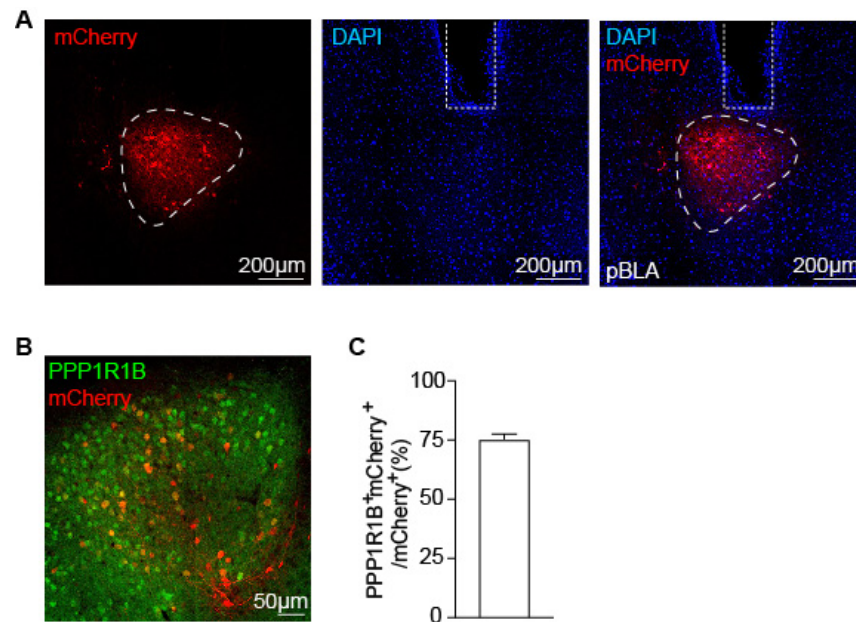
4 (A) Representative histology images showing the expression of eArchT-EYFP and optical fiber
5 implant in $Rspo2$ -Cre mice. (B) Representative histology images showing the expression of
6 eArchT-EYFP and optical fiber implant in $Ppp1r1b$ -Cre mice.

7

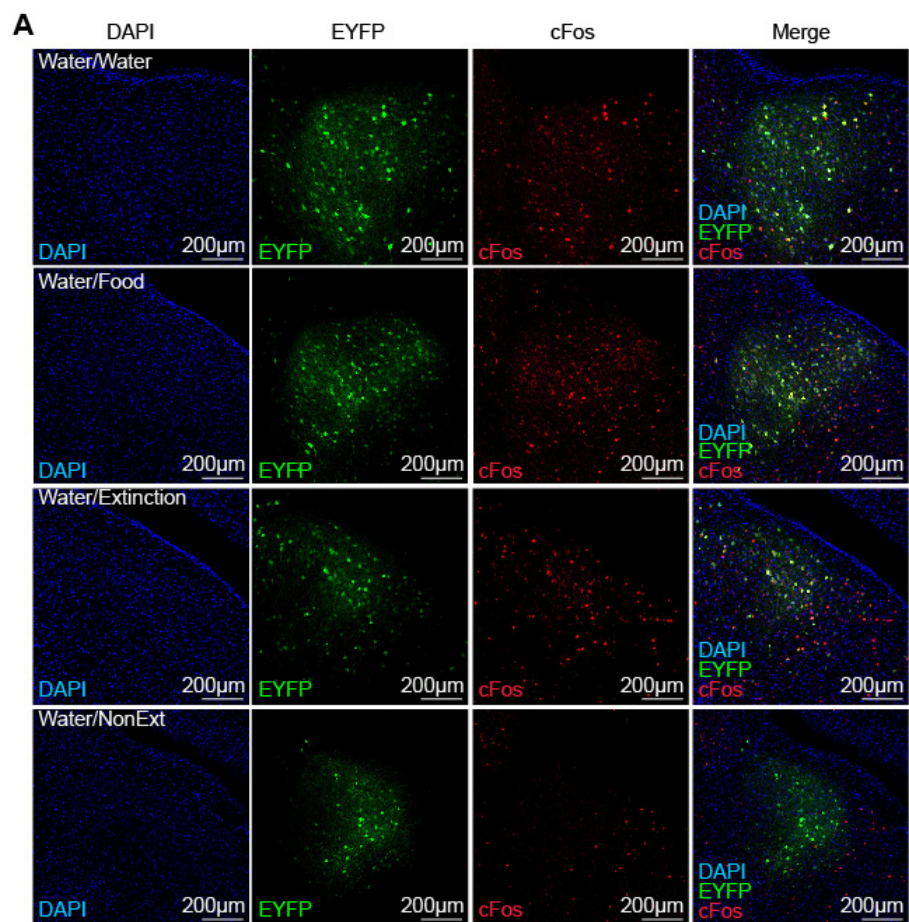


1 **Figure S4. Characterization of fear extinction engram cells in pBLA (Related to Figure 4)**

2 (A) AAV₉-c-fos-tTA and AAV₉-TRE-EYFP virus combination were injected into pBLA of
3 C57BL/6 mice. After 24 hrs Dox-OFF diet, mice were injected with Kainic Acid to induce seizure.
4 Image showing efficient labeling in pBLA. (B) EYFP⁺ cell counts that were activated by Kainic
5 Acid-induced seizure from pBLA section ($n = 6$ mice). $93 \pm 3\%$ cells were labelled. (C) EYFP⁺
6 cell counts from pBLA sections under three different conditions: Dox OFF during extinction
7 retrieval (Dox OFF-Ext, $n = 4$ mice), Dox OFF at home cage (Dox OFF-HC, $n = 4$ mice) and Dox
8 ON during extinction retrieval (Dox ON-Ext, $n = 4$ mice). One-way ANOVA. (D) Representative
9 histology image showing expression of ChR2-EYFP and EYFP and optical fiber implants. (E)
10 Behavior paradigm as described in Figure 4D. (F) When ChR2-expressing extinction engram
11 neurons were activated, Light ON group exhibited accelerated fear extinction learning on Day 6
12 compared to light OFF group. ChR2 Light OFF, $n = 7$, ChR2 Light ON, $n = 12$. Two-way RM
13 ANOVA. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$. Data are presented as mean \pm
14 SEM.



1
2 **Figure S5. Inhibition of fear extinction engram in pBLA (Related to Figure 5)**
3 (A) AAV₉-c-fos-tTA and AAV₉-TRE-ArchT-mCherry virus combination were injected into
4 pBLA of C57BL/6 mice. Mice underwent fear extinction retrieval during 24 hrs Dox OFF period.
5 Image showing efficient labeling of eArchT-mCherry in pBLA. (B) Image showing activity-
6 induced mCherry-expressing neurons in pBLA stained with PPP1R1B antibody (green). (C) 75 ±
7 2% mCherry⁺ cells were labeled as PPP1R1B⁺. Data are presented as mean ± SEM.

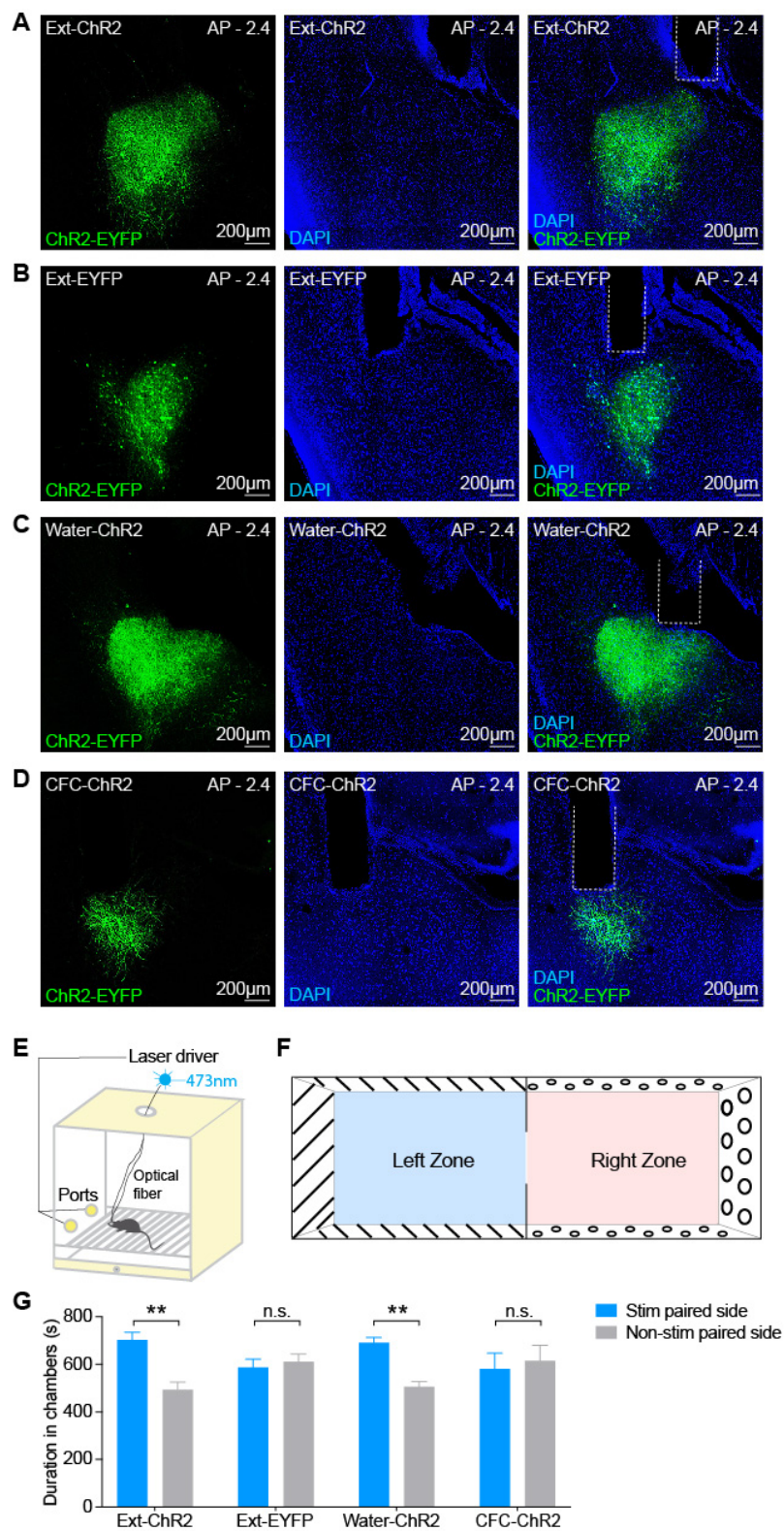


1

1 **Figure S6. Cellular overlap of water-reward neurons and extinction engram neurons in**

2 **pBLA (Related to Figure 6)**

3 (A) pBLA neurons activated by water reward was labeled by EYFP (green) and pBLA cells
4 reactivated by water reward (Water/Water group), food reward (Water/Food group), extinction
5 retrieval (Water/Extinction group) and fear retrieval (Water/NonExtinction group) were stained by
6 c-Fos (red). The merged images with DAPI staining are shown on the right.



1

1 **Figure S7. Behavioral overlap of water-reward neurons and extinction engram neurons in**
2 **pBLA (Related to Figure 7)**

3 (A) Image showing efficient labeling of ChR2-EYFP and optical fiber implant in pBLA of Ext-
4 ChR2 group in Figure 7B. (B) Image showing efficient labeling of EYFP and optical fiber implant
5 in pBLA of Ext-EYFP group in Figure 7B. (C) Image showing efficient labeling of ChR2-EYFP
6 and optical fiber implant in pBLA of Water-ChR2 group in Figure 7B. (D) Image showing labeling
7 of ChR2-EYFP and optical fiber implant in pBLA of CFC-ChR2 group in Figure 7B. (E)
8 Schematic diagram of operant chamber for optogenetic self-stimulation behavior in Figure 7C. (F)
9 Schematic diagram of chamber used for optogenetic place preference behavior in Figure 7D. (G)
10 Total duration spent in the stimulation paired side chamber and non-stimulation paired side. Ext-
11 ChR2 and Water-ChR2 group spent significantly more time in the stimulation paired side chamber
12 but not Ext-EYFP and CFC-ChR2 group. Paired t-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$,
13 **** $P < 0.0001$. Data are presented as mean \pm SEM.

1 **References**

- 2 Kim, J., Pignatelli, M., Xu, S., Itohara, S., and Tonegawa, S. (2016). Antagonistic negative and
3 positive neurons of the basolateral amygdala. *Nat Neurosci* *19*, 1636-1646.
- 4 Kim, J., Zhang, X., Muralidhar, S., LeBlanc, S.A., and Tonegawa, S. (2017). Basolateral to
5 Central Amygdala Neural Circuits for Appetitive Behaviors. *Neuron* *93*, 1464-1479 e1465.
- 6 Kitamura, T., Ogawa, S.K., Roy, D.S., Okuyama, T., Morrissey, M.D., Smith, L.M., Redondo,
7 R.L., and Tonegawa, S. (2017). Engrams and circuits crucial for systems consolidation of a
8 memory. *Science* *356*, 73-78.
- 9 Liu, X., Ramirez, S., Pang, P.T., Puryear, C.B., Govindarajan, A., Deisseroth, K., and Tonegawa,
10 S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall.
11 *Nature* *484*, 381-385.
- 12 Mukamel, E.A., Nimmerjahn, A., and Schnitzer, M.J. (2009). Automated analysis of cellular
13 signals from large-scale calcium imaging data. *Neuron* *63*, 747-760.
- 14 Nishi, A., Snyder, G.L., and Greengard, P. (1997). Bidirectional regulation of DARPP-32
15 phosphorylation by dopamine. *J Neurosci* *17*, 8147-8155.
- 16 Redondo, R.L., Kim, J., Arons, A.L., Ramirez, S., Liu, X., and Tonegawa, S. (2014).
17 Bidirectional switch of the valence associated with a hippocampal contextual memory engram.
18 *Nature* *513*, 426-430.
- 19 Ziv, Y., Burns, L.D., Cocker, E.D., Hamel, E.O., Ghosh, K.K., Kitch, L.J., El Gamal, A., and
20 Schnitzer, M.J. (2013). Long-term dynamics of CA1 hippocampal place codes. *Nat Neurosci* *16*,
21 264-266.
- 22