Dissecting the role of defined neuronal populations in fear learning

Inauguraldissertation

zur Erlangung der Würde eines Doktors der Philosophie vorgelegt der Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel von

Steffen Benjamin Eggert Wolff

aus Emden, Deutschland

Basel, 2014

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von:

Prof. Dr. Andreas Lüthi
(Fakultätsverantwortlicher und Dissertationsleiter)

Prof. Dr. Carl Petersen
(Korreferent)

Basel, den 15.10.2013

Prof. Dr. Jörg Schibler
(Dekan)

Namensnennung-Keine kommerzielle Nutzung-Keine Bearbeitung 3.0 Schweiz
(CC BY-NC-ND 3.0 CH)

Sie dürfen: **Teilen** — den Inhalt kopieren, verbreiten und zugänglich machen

Unter den folgenden Bedingungen:

Namensnennung — Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen.

Keine kommerzielle Nutzung — Sie dürfen diesen Inhalt nicht für kommerzielle Zwecke nutzen.

Keine Bearbeitung erlaubt — Sie dürfen diesen Inhalt nicht bearbeiten, abwandeln oder in anderer Weise verändern.

Wobei gilt:

- **Verzichtserklärung —** Jede der vorgenannten Bedingungen kann **aufgehoben** werden, sofern Sie die ausdrückliche Einwilligung des Rechteinhabers dazu erhalten.
- **Public Domain (gemeinfreie oder nicht-schützbare Inhalte) —** Soweit das Werk, der Inhalt oder irgendein Teil davon zur Public Domain der jeweiligen Rechtsordnung gehört, wird dieser Status von der Lizenz in keiner Weise berührt.
- **Sonstige Rechte —** Die Lizenz hat keinerlei Einfluss auf die folgenden Rechte:
    - Die Rechte, die jedermann wegen der Schranken des Urheberrechts oder aufgrund gesetzlicher Erlaubnisse zustehen (in einigen Ländern als grundsätzliche Doktrin des **fair use** bekannt);
    - Die **Persönlichkeitsrechte** des Urhebers;
    - Rechte anderer Personen, entweder am Lizenzgegenstand selber oder bezüglich seiner Verwendung, zum Beispiel für **Werbung** oder Privatsphärenschutz.
- **Hinweis —** Bei jeder Nutzung oder Verbreitung müssen Sie anderen alle Lizenzbedingungen mitteilen, die für diesen Inhalt gelten. Am einfachsten ist es, an entsprechender Stelle einen Link auf diese Seite einzubinden.

Quelle: http://creativecommons.org/licenses/by-nc-nd/3.0/ch/ Datum: 12.11.2013

TABLE OF CONTENTS

ABBREVIATIONS **7**

1. **ABSTRACT** **9**

2. **INTRODUCTION** **11**

2.1. Learning and memory – neuronal correlates and mechanisms **11**
2.1.1. Neuronal circuits 12
2.1.2. Selection of cells for a memory trace 13

2.2. Associative learning – Fear conditioning **14**

2.3. Neuronal circuitry of fear learning **16**
2.3.1. The amygdala 16
2.3.2. Cytoarchitecture of the basolateral amygdala 18
2.3.3. The role of the basolateral amygdala in fear learning 19
2.3.4. Synaptic plasticity in the basolateral amygdala 21
2.3.5. Inhibition in the basolateral amygdala 23
2.3.6. The fear circuitry beyond the amygdala 24
2.3.7. Auditory cortex 24
2.3.8. Medial Prefrontal cortex 24

2.4. Technical advances for the study of neuronal populations in associative learning **26**

2.5. Aims of this thesis **30**

3. **RESULTS** **31**

3.1. Publication I: A disinhibitory microcircuit for associative fear learning in auditory cortex (Letzkus, Wolff et al. 2011, Nature) **31**

3.2. Publication II: Distinct subtypes of amygdala interneurons control fear learning through stimulus-specific disinhibition (Wolff et al. 2014, Nature, accepted for publication) **67**

3.3. Publication III: Long-range connectivity defines behavioral specificity of amygdala neurons (Senn, Wolff et al. 2014, Neuron) **101**

4. **DISCUSSION** **138**

4.1. The role of inhibition and disinhibition in associative learning **138**
4.1.1. Regulation of Generalization 140
4.1.2. Disinhibition in the extinction and expression of fear 140

4.1.3. Regulation of disinhibition 141

4.2. Association of non-overlapping CS and US 143

4.3. Selection of cells for a memory trace and memory allocation 144

4.3.1. Testing factors for the incorporation of neurons into a memory trace 145

4.4. Diversity of neuronal subpopulations 148

4.5. Future and caveats of optogenetics 151

4.6. Conclusions 155

5. **ACKNOWLEDGEMENTS 156**

6. **REFERENCES 158**

ABBREVIATIONS

AAV	Adeno-associated virus
AC	Auditory Cortex
AMPAR	α-Amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid-receptor
ARCH	Archaerhodopsin
BA	Basal Amygdala
BLA	Basolateral Amygdala
BMA	Basomeidal Amygdala
CAV	Canine adeno virus
CEA	Central Amygdala
CEl	Lateral subdivision of the CEA
CEm	Medial subdivision of the CEA
ChR2	Channelrhodopsin-2
CR	Calretinin
CREB	cAMP response element-binding protein
CS	Conditioned stimulus
CS-	CS, not paired with a US
CS+	CS, paired with a US
EPSP	Excitatory post-synaptic potential
GABA	Gamma-aminobutyric acid
HSV	Herpes simplex virus
ICM	Intercalated cell masses
IL	Infralimbic cortex
IN	Interneuron
LA	Lateral Amygdala
LTD	Long-term depression

LTP	Long-term potentiation
MAPK/ERK	Mitogen-activated protein kinase/extracellular signal-regulated kinase
MGm/PIN	Magnocellular medial geniculate/posterior intralaminar nucleus
mPFC	Medial prefrontal cortex
NMDAR	N-methyl-D-aspartate receptor
NpHR	*Natronomonas pharaonis* halorhodopsin
NR2B	N-methyl D-aspartate receptor subtype 2B
PAG	Periaqueductal gray
PL	Prelimbic cortex
PN	Principal neuron
PTSD	Post-traumatic stress disorder
PV	Parvalbumin
PVN	Paraventricular nucleus of the hypothalamus
SOM	Somatostatin
STDP	Spike-timing dependent plasticity
US	Unconditioned stimulus
VGCC	Voltage-gated calcium channel
VIP	Vasointestinal peptide
WGA	Wheat germ agglutinin

1. ABSTRACT

The astonishing plasticity of the brain and its ability for continuous learning and memory formation are among its most essential functions, since constant adaptation to the environment is critical for an animal's survival. The individual brain areas mediating distinct types of learning and the cellular and molecular mechanisms underlying synaptic plasticity have been studied in great detail in the past. In order to understand learning and memory, though, it is necessary to bridge the gap between our insights on the macro- and the micro-scale. Neuronal circuits are the link between individual cells and entire brain areas and learning manifests as profound changes in their processing of information. However, the identity, function and mechanisms of neuronal circuits mediating learning are relatively unexplored. The study of neuronal circuits will not only reveal the actual sites of plasticity, but also the control mechanisms for the acquisition and expression of memories. To dissect the organization and function of neuronal circuits, both locally and across multiple brain areas, it is key to identify their fundamental building blocks – defined neuronal populations. Their specific roles and interplay will provide crucial insight into the function of neuronal circuits and the mechanisms of learning and memory.

During my PhD, I focused on the dissection of neuronal circuits underlying associative fear learning. Fear conditioning is one of the most powerful model systems to investigate plasticity of neuronal circuits and mechanisms of associative learning. The amygdala has been identified as a key brain structure where associative plasticity is induced through pairing of a neutral tone and a mild aversive footshock. Importantly, the amygdala is embedded in a vast network of brain structures which play distinct roles in fear learning, such as sensory cortices, the medial prefrontal cortex, and the hippocampus.

To determine the role of defined neuronal populations in these circuits, I established a broad portfolio of optogenetic techniques for monitoring and manipulating neuronal activity. Furthermore, I refined these methods, by developing hardware, genetic strategies, and viral approaches, and by synergistically integrating and combining electrophysiological recordings with light-induced perturbations of neuronal activity.

In three distinct projects, I applied these techniques to study different aspects of the neuronal fear circuitry. I revealed that distinct disinhibitory microcircuits both in the auditory cortex (AC) and the basolateral amygdala (BLA) control the acquisition of fear and that the function of BLA principal neurons (PNs) is related to the circuits in which they are embedded.

In the auditory cortex, a footshock leads to the acetylcholine-mediated excitation of interneurons (INs) in layer 1. Layer 1 interneurons, in turn, inhibit parvalbumin-expressing (PV^+) INs in layer 2/3. This disinhibitory microcircuit allows for the enhanced excitation and plasticity of pyramidal neurons in response to the footshock, and is necessary for learning.

The necessity of footshock-induced plasticity in AC and its disinhibitory control represent novel aspects of the circuitry underlying fear learning.

In the BLA, a different disinhibitory microcircuit controls the strength of acquired fear memories via stimulus-dependent mechanisms. During the shock, both PV^+ and somatostatin-expressing (SOM^+) INs are inhibited, which results in a general disinhibition of PNs along their somatodendritic axis, allowing for plasticity. During the tone however, PV^+ INs are excited, while SOM^+ INs are inhibited – most likely by directly connected PV^+ INs. This causes dendritic disinhibition of PNs, which leads to a boosting of the impact of auditory inputs and enhanced plasticity and learning. This demonstrates that BLA PV^+ and SOM^+ INs exert bidirectional control over fear acquisition through differential changes in inhibition along the somatodendritic axis of PNs.

PNs in the BLA project to different target regions. This differential connectivity also relates to differences in their function. I demonstrated that BLA neurons projecting to the infralimbic or prelimbic cortex are oppositely involved in fear acquisition and extinction, and exhibit distinct plastic changes. This not only shows a relationship between function and connectivity, but also represents an example for the control of learning in neuronal circuits which span across several brain areas.

In summary, I established a combination of optogenetic and electrophysiological tools and applied them to dissect the neuronal circuits of fear. I revealed several fundamental mechanisms in fear learning and made significant contributions towards the understanding of defined inhibitory and excitatory neuronal subpopulations both in auditory cortex and amygdala.

2. INTRODUCTION

2.1 Learning and memory – neuronal correlates and mechanisms

One of the most astonishing abilities of the brain is to constantly learn and form new memories. This remarkable capability to adapt to the environment also poses a truly fundamental question – how does the brain learn?

As there is not only one form of learning, there is also not only one "learning center" in the brain, which controls all learning processes and stores all memories. Rather, the entire brain functions as a "learning organ". Depending on the form of learning, distinct processes, mechanisms and brain regions are recruited. Research has focused for a long time on determining which brain areas are involved in discrete forms of learning. For example, studies on patients with restricted brain damage as well as experimental lesions, inactivations and electrical stimulations have revealed that the hippocampus is important for spatial learning and episodic memories (Martin and Clark 2007), the basal ganglia are crucial for motor learning (Olveczky 2011) and the amygdala is necessary for fear learning (LeDoux 2000).

However, this view is still too limited and restricted. Even for a discrete form of learning, not only one brain region is responsible. For instance, motor learning is based on the interplay between at least the basal ganglia, the motor cortex and the cerebellum (Olveczky 2011). The amygdala is embedded in a vast network of brain regions involved in fear learning, including the hippocampus, the medial prefrontal cortex, the thalamus, sensory cortices and several brain stem areas (LeDoux 2000). Although important insights have been gained by assigning individual brain regions to discrete functions, this approach cannot account for interactions and synergies between multiple brain structures.

Research has also been successful in uncovering the details of learning and memory at the other end of the scale, by investigating its cellular and molecular underpinnings. Changes in the strength of synapses have been repeatedly suggested as the cellular mechanisms underlying memory formation (Cajal 1909 - 1911; Hebb 1949; Eccles 1965; Kandel and Spencer 1968). Perhaps the most influential of these proposals is the Hebbian cell assembly theory (Hebb 1949), which posits that the population of neurons co-activated during learning undergoes plastic changes to strengthen its connections, thereby becoming the engram of that memory. However, only the discovery of long-term potentiation (LTP) of synaptic transmission (Bliss and Lomo 1973) and its counterpart, long-term depression (LTD, Lynch et al. 1977) provided physiological support for this theory. Sparked off by these discoveries, research made enormous progress. Today we have a detailed understanding of the mechanisms of synaptic plasticity at excitatory synapses (Martin et al. 2000; Malenka and Bear 2004; Sjostrom et al. 2008). A multitude of exquisitely fine-tuned mechanisms are in place that enable projection neurons to strengthen or weaken presynaptic inputs,

depending on their impact on firing in the postsynaptic cell. Even more striking support for the idea of Hebbian plasticity is provided by spike-timing-dependent plasticity (STDP), where, in numerous cell types, synapses active just before the postsynaptic action potential undergo potentiation, whereas inputs arriving immediately after the postsynaptic discharge are depressed (Markram et al. 1997; Bi and Poo 1998; Caporale and Dan 2008). Although it has not been demonstrated so far that synaptic plasticity alone is sufficient for learning, overwhelming evidence suggests that synaptic plasticity is necessary for memory formation (Martin et al. 2000; Whitlock et al. 2006; Martin and Clark 2007; Neves et al. 2008; Sah et al. 2008).

Our understanding of synaptic plasticity is broad, but without knowledge about the sites of plasticity, the selection of the modified cells and underlying control mechanisms, a description of the processes in individual synapses and cells cannot provide a full picture of the memory trace. Even manipulations of plasticity in discrete brain areas cannot fully explain learning, since a multitude of neuronal circuits, participating in different behaviors, are affected to different extents. To understand learning and memory, the sole study of synaptic and cellular plasticity is therefore not enough.

2.1.1 Neuronal circuits

While the study of brain areas and synaptic plasticity has been successful in revealing the underpinnings of learning and memory, recent years have seen a shift in the focus of memory research. The most pressing unresolved issues today are to bridge the gap between these two fields and to elucidate the role of neuronal circuits in learning and memory (Neves et al. 2008). Neuronal circuits are anatomically and functionally interconnected networks of neurons, both locally and across brain areas, which control distinct aspects of behavior. To this end, neuronal circuits integrate a broad range of internal and external signals and information, like sensory input, previous experiences, and the internal state of the animal. Associative learning manifests at this mesoscopic level as a change in the processing of information by neuronal circuits (**Fig. 1**). The exact nature of these changes, the mechanisms of their induction and how the function of a circuit is altered by memory formation are only beginning to emerge. Importantly, neuronal circuits are not made from uniform building blocks. As described already by Ramón y Cajal, the brain is comprised of a multitude of different neurons (Cajal 1909 - 1911), which not only have diverse morphologies and connections, but also distinct functions. To understand the complex neuronal circuits underlying associative learning, it is crucial to identify the individual circuit elements and to reveal their roles, connectivity and interactions. However, only recently the study of defined neuronal populations, which account for these individual elements, has become feasible. Novel, sophisticated techniques allow the selective targeting of neuronal populations for the monitoring and manipulation of their activity during behavior. Thanks to these advances, the specific function of defined neuronal populations can be revealed and a

more detailed dissection and deeper understanding of neuronal circuits can be obtained. This thesis represents some of the first examples of how these techniques can be used to determine the roles of individual components of neuronal circuits, both locally and across different brain areas. Modern neuroscience will focus more and more on neuronal circuits, and this will ultimately tie together the role of brain areas and synaptic plasticity, thereby leading to a better understanding of learning and memory.

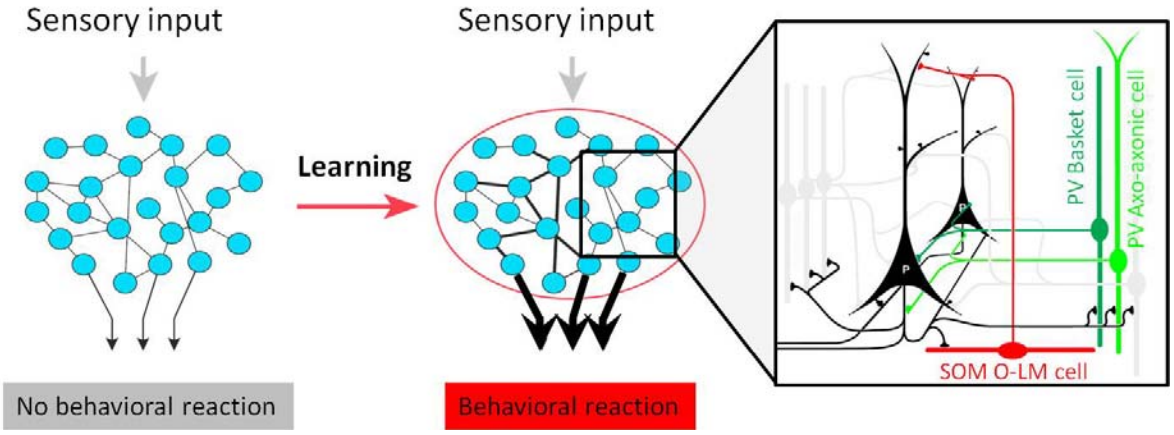


Figure 1 Learning induces long-term changes in information processing in neuronal circuits.
A neutral sensory stimulus, with no specific meaning for the animal, can enter a neuronal circuit without triggering a behavioral reaction (left). During associative learning, the stimulus can acquire significance for the animal, e.g. by pairing with a salient cue. Such learning leads to persistent changes in the connectivity and in the information processing in the involved neuronal circuit. Upon re-presentation, the sensory stimulus triggers now different processes in the changed neuronal circuit and can thereby evoke a behavioral reaction (middle). The modifications in a neuronal circuit during learning manifest in the interplay of its individual elements (right). Most neuronal circuits consist of both excitatory (black) and inhibitory (green, red, gray) neurons, respectively. Inhibitory interneurons, which are crucial for the regulation of circuit function, comprise multiple different subclasses with different properties and function. Two prominent examples are parvalbumin-expressing (PV, green) neurons targeting the perisomatic region of principal neurons and dendrite-targeting somatostatin-expressing (SOM, red) interneurons. Figure adapted from Klausberger & Somogyi 2008.

2.1.2 Selection of cells for a memory trace

Associative learning causes specific changes in neuronal circuits, thereby creating a trace of the acquired memory. Neurons which are recruited to this memory trace undergo synaptic plasticity - the molecular and cellular underpinning of learning. However, only a fraction of cells in a circuit is recruited to a memory trace. Furthermore, partially overlapping memory traces, which represent different experiences, exist in parallel. How the appropriate cells are selected to enter the individual traces and how the memory is allocated remains elusive.

Different lines of research suggest that the level of activation of a neuron during learning is a key factor for its incorporation into a memory trace. In a series of experiments, it could be

shown that exogenous expression of the transcription factor CREB, which increases excitability, makes neurons more likely to be recruited to a memory trace (Han et al. 2007; Silva et al. 2009; Zhou et al. 2009; Kim et al. 2013). Selective deletion of these cells leads to loss of the specific memory, without affecting older memories or the ability to learn (Han et al. 2009). The CREB-mediated increase in excitability enhances activation by sensory inputs and synaptic plasticity (Kim et al. 2013). These studies demonstrate that the levels of excitability and activity determine which cells are incorporated into a memory trace.

These findings are supported and extended by research addressing the allocation of contextual fear memories in the hippocampus (Liu et al. 2012; Ramirez et al. 2013). It was demonstrated that neurons in the dentate gyrus, which are active during contextual fear conditioning, form a fear memory trace for the specific conditioning context. Optogenetic excitation of these cells reactivates the memory trace and elicits fear responses even in a neutral context (Liu et al. 2012). Remarkably, it is even possible to merge two contextual memories into one (Ramirez et al. 2013). Optogenetic excitation of neurons forming the memory trace for a neutral context, during conditioning in another context, creates a merged memory trace. Information about the neutral context becomes part of the fear memory. Therefore, the “neutral” context evokes a partial fear response after this training. Strikingly, also the conditioning context only evokes a partial fear response, which can be recovered by additional optogenetic activation of the memory trace for the “neutral” context. This shows that the complete fear memory is a combination of the two context memory traces (Ramirez et al. 2013). These experiments demonstrate that the activity of neurons is crucial for their recruitment to a memory trace during learning.

Studies, as described in this thesis, revealing the mechanisms which control activity, plasticity and learning in neuronal circuits will also lead to a better understanding of the formation of memory traces and of memory allocation.

2.2 Associative learning – Fear conditioning

Classical fear conditioning is one of the most powerful model systems to study the function and plasticity of neuronal circuits as well as the mechanisms of associative learning in the mammalian brain (LeDoux 2000). During fear conditioning, the subject is exposed to the conditioned stimulus (CS), which is initially neutral, like a tone or a light, paired with the unconditioned stimulus (US), a noxious stimulus like a foot-shock (**Fig. 2**). Thereby, the CS gains aversive properties and, on subsequent presentation, triggers multifaceted fear responses. In rodents, fear responses comprise changes in heart rate and blood pressure, release of stress hormones, analgesia and facilitation of reflexes (LeDoux 2000; Fanselow and Poulos 2005). A range of active and passive defensive behaviors can also be triggered by fearful stimuli, depending on proximity, context, and intensity (Adolphs 2013). Aversive stimuli presented in inescapable contexts mainly trigger freezing behavior. Freezing is an

innate defensive behavior evolved to avoid detection by predators (LeDoux 2000; Fanselow and Poulos 2005). Since freezing is manifested as an immobile posture, it is easily and robustly measured and used as the main experimental readout to quantify fear responses (LeDoux 2000; Fanselow and Poulos 2005). The ability to precisely control stimuli and the robust readout of a behavioral response render classical fear conditioning a reliable and physiologically relevant model system.

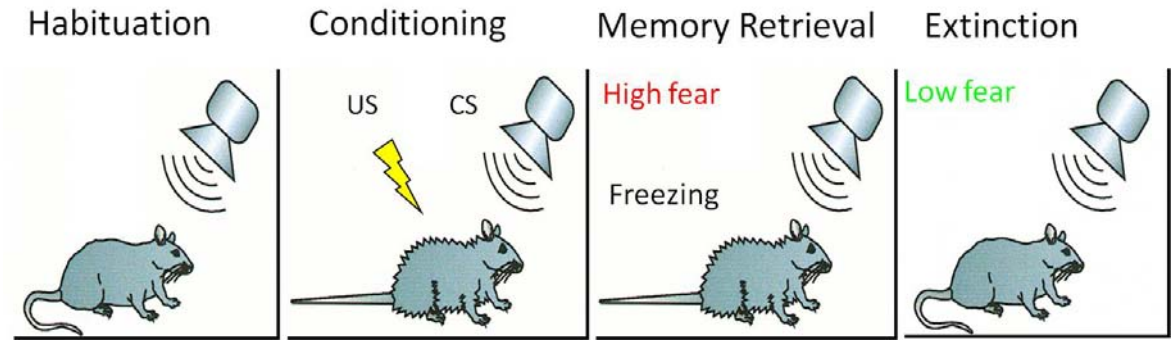


Figure 2 Behavioral paradigm for classical auditory fear conditioning and extinction.
During a habituation session, an initially neutral tone (the future conditioned stimulus (CS)) is presented, but does not induce a behavioral reaction. During the conditioning phase, the CS is repeatedly paired with the unconditioned stimulus (US), a mild foot-shock. This pairing leads to the formation of an association between the CS and the US and the creation of a memory trace. During a test session on the next day, the presentation of the CS triggers the retrieval of the previously acquired memory and evokes fear responses. Fear levels are quantified by measuring the duration of freezing during the CS. Subsequent repeated presentation of the CS without reinforcement by the US leads to fear extinction. A new context-dependent inhibitory memory is formed, which suppresses the fear response in the extinction context.

Repeated presentations of the CS without reinforcement by the US induce extinction of conditioned fear, resulting in a progressive reduction of the fear response (**Fig. 2**). Importantly, this process is not a mere erasure of the conditioned fear memory, but an active learning process (Rescorla 2001; Myers and Davis 2004). A new, independent memory is formed, which inhibits the fear memory. The notion that the fear memory remains present is supported by the facts that fear extinction is highly context-dependent, conditioned fear responses can spontaneously reappear over time, and can be recovered by re-exposure to the US (reinstatement) (Rescorla and Heth 1975; Bouton and King 1983; Quirk 2002). Thus, fear and extinction memory traces co-exist and can be retrieved, depending on the environmental context and internal state of the animal.

Besides their major relevance in the study of the neuronal circuits underlying associative learning, fear conditioning and extinction are also important model systems for human pathologies. Anxiety disorders, phobias, obsessive compulsive disorders and post-traumatic stress disorder (PTSD) are caused by malfunctions of the fear circuitry, inducing excessive fear responses or deficient discrimination between threatening and neutral sensory stimuli

(Kent and Rauch 2003). A better understanding of the neuronal substrates underlying physiological fear learning and extinction may help to identify targets for interventions in these diseases and to develop potential pharmacological and behavioral therapies.

2.3 Neuronal circuitry of fear learning

2.3.1 The Amygdala

The basic mechanisms and brain areas underlying fear learning have been intensively studied in the last decades (Davis 2000; LeDoux 2000; Maren and Quirk 2004; Fanselow and Poulos 2005). This allows the identification and investigation of the precise neuronal circuits, involved in the different phases and aspects of fear learning. The amygdala (**Fig. 3**) has been identified as a key structure underlying both fear conditioning and extinction in animals and humans (LeDoux 2000; Maren 2001; Herry et al. 2006; Sotres-Bayon et al. 2007).

The amygdala is a non-layered structure located deep in the medial temporal lobe and consists of several nuclei with differences in cytoarchitecture, histochemistry and connectivity (Pitkanen et al. 2000) (**Fig. 3**). Nevertheless, certain subnuclei can be grouped, based on their characteristics (Sah et al. 2003; LeDoux 2007). Most important for associative fear learning are the basolateral amygdala (BLA) and the central amygdala (CEA). The BLA is in the focus of this thesis. It is a cortex-like area, consisting of the lateral (LA), the basal (BA) and the basomedial (BMA) nuclei and represents the main input structure of the amygdala. The CEA in contrast, comprised of the central lateral (CEl), central capsular (CElc) and central medial (CEm) nuclei, is striatum-like and is considered as the main output center. Several superficial nuclei, like the anterior cortical nucleus or the bed nucleus of the accessory olfactory tract are also part of the amygdala, involved rather in innate fear responses. The intercalated cell masses, clusters of GABAergic neurons surrounding the BLA, are likely involved in fear learning, as well (Pare and Smith 1993). The dense connectivity within and between subnuclei determines the general flow of information in the amygdala. For example, sensory input is mainly received at the LA and further transmitted to the BA. Both LA and BA project to the CEA, where the information is integrated with further sensory input. The CEl finally targets the CEm, which is the main amygdala output nucleus (Pitkanen et al. 1997; Sah et al. 2003) (**Fig. 3**).

Besides intra-amygdala connectivity, all subnuclei are strongly interconnected with other cortical and sub-cortical structures. The LA is the main input nucleus and receives multimodal sensory information from the thalamus and the cortex (Turner and Herkenham 1991; McDonald 1998) (**Fig. 3**). The direct thalamic pathway enters the LA via the internal capsule (LeDoux et al. 1990), and the indirect thalamo-cortico pathway via the external capsule (LeDoux et al. 1991; Amaral and Insausti 1992) (**Fig. 3**). The LA is the first integration

site for CS and US inputs and LA neurons exhibit enhanced CS-evoked responses after fear conditioning (Quirk et al. 1995; Quirk et al. 1997; Rogan et al. 1997; Rosenkranz and Grace 2002) (**Fig. 3**). Importantly, the LA is connected with other memory-related structures, like prefrontal and perirhinal cortices, and the hippocampus (LeDoux 2000). The connectivity of the LA is mostly reciprocal, with the exceptions of the unidirectional projections from thalamus to LA and from LA to CEl (Turner and Herkenham 1991; Pitkanen et al. 1997).

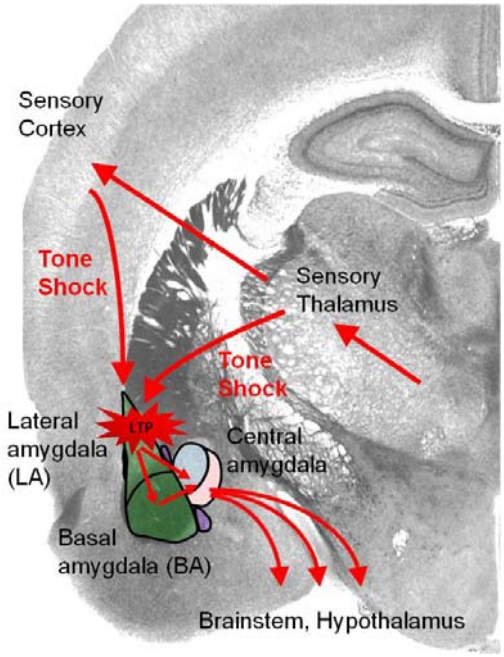


Figure 3 Anatomy of the fear circuit and flow of sensory information.
Tone and shock inputs are sent from the periphery to different thalamic nuclei. The thalamus directly projects to the lateral amygdala (LA) and conveys sensory information via this "low road" pathway. Simultaneously, the thalamus projects via the "high road" to sensory cortices, like the auditory cortex, where the sensory information is further processed and subsequently also conveyed to the LA. Coactivation of LA neurons by tone and shock inputs leads to long-term potentiation (LTP) at both thalamic and cortical afferents in the LA. Information is transmitted to the basal amygdala (BA), which is important for switches in the emotional state of an animal during conditioning and extinction. The LA and the BA together form the basolateral amygdala (BLA). Both the BA and the LA project to the lateral subdivision of the central amygdala (CEl), but only the BA also to its medial subdivision (CEm). The CEm is the final output nucleus of the amygdala and projects to the hypothalamus and several brainstem nuclei, where the physiological fear responses are triggered.

The extra-amygdala connectivity of the BA is similar to the LA, although most connections are not as dense. Strong reciprocal connections are made with the prefrontal and perirhinal cortices (Berendse et al. 1992; Bacon et al. 1996). Important connections are also formed with the hippocampus, the thalamus, basal forebrain and nucleus accumbens (McDonald 1998; Pikkarainen et al. 1999; Pitkanen et al. 2000). The BA receives its main intra-amygdala input from the LA and projects to the LA, CE and other amygdala nuclei (Savander et al. 1997) (**Fig. 3**).

The CEA receives both strong intra-amygdala input and sensory inputs from the thalamus, olfactory bulb and the nucleus of the solitary tract in the brainstem (Veinante and Freund-Mercier 1998). The CEA, especially the CEm, sends projections to the hypothalamus and to different brainstem structures, such as the PAG or the PVN, orchestrating conditioned autonomic and motor responses (Krettek and Price 1978; Veening et al. 1984; LeDoux et al. 1988) (**Fig. 3**). For a long time, the CEA was considered to be a mere relay and output nucleus. However, in recent years it was shown that the CEA is a major site of plasticity and necessary for learning and expression of fear memories (Ciocchi et al. 2010; Gozzi et al. 2010; Haubensak et al. 2010; Duvarci et al. 2011).

2.3.2 Cytoarchitecture of the basolateral amygdala

Based on morphological, neurochemical and physiological features, the BLA is a cortex-like structure and consists of two main neuronal populations (**Fig. 4**). The majority of neurons (about 80%) in the BLA are spiny, glutamatergic projection neurons, the so-called principal neurons (PNs), (McDonald 1982). PNs extend axon collaterals and form synaptic contacts inside the amygdala, but project also to other brain regions (Herry et al. 2008). PNs have large dendritic arborizations spanning across sub-nuclear boundaries (Pare and Gaudreau 1996). So far, no molecular markers are known to differentiate between distinct subclasses of PNs in the BLA. However, PNs differ in their projection target which is strongly correlated with their function (Senn et al. 2013).

Similar to the neocortex, the BLA contains about 20% of GABAergic inhibitory interneurons (**Fig. 4**). INs are aspiny, have smaller somata compared to PNs, and form a heterogeneous population in regard to their dendritic and axonal arborization. Further heterogeneity exists in the expression of molecular markers (**Fig. 4**), connectivity, subcellular targeting, cellular properties and function (Freund and Buzsaki 1996; Sah et al. 2003; Somogyi and Klausberger 2005; Ehrlich et al. 2009; Pape and Pare 2010; Fishell and Rudy 2011; Spampanato et al. 2011). So far, almost all of our knowledge about the function of the BLA is based on studies investigating BLA PNs. The role of inhibition and especially of different subtypes of INs remains to be elucidated, but they are likely to play differential roles during fear behavior (see 2.3.5 Inhibition in the amygdala). First steps to dissect the function and interactions of BLA interneurons are part of this thesis (Wolff et al. 2013).

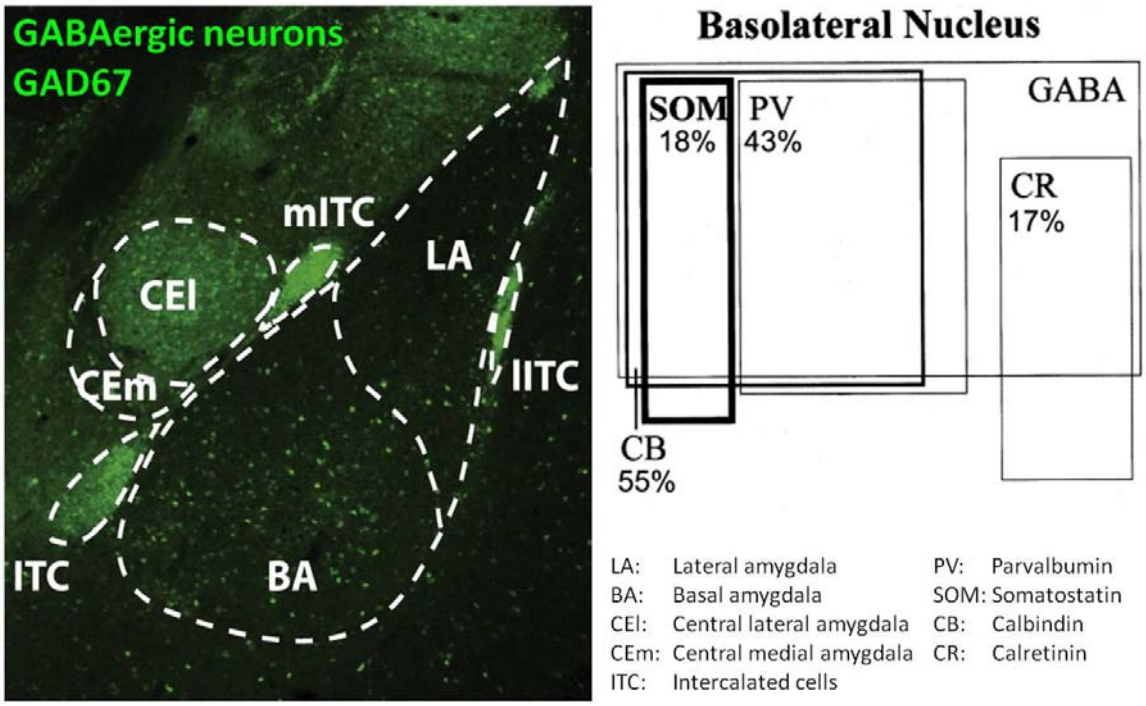


Figure 4 GABAergic neurons in the amygdala.
Left: The central amygdala is a striatum-like structure, mostly consisting of GABAergic neurons which resemble striatal medium spiny neurons. In contrast, the basolateral amygdala (BLA), consisting of lateral (LA) and basal (BA) amygdala is a cortex-like area. It is a non-layered structure, but its cellular composition resembles the cortex. The majority of neurons (80%) are glutamatergic and excitatory and only about 20% are GABAergic and inhibitory (green – GFP label under the control of the GAD67 promoter). Right: The population of GABAergic inhibitory neurons in the BLA consists of several subtypes, in properties and proportions resembling cortical or hippocampal interneurons (McDonald 1982). The major subtype are parvalbumin-expressing (PV^+) fast-spiking interneurons, which target the perisomatic region of principal neurons. PV^+ cells account for about half of the BLA interneurons. Another important subtype are dendrite-targeting somatostatin-expressing (SOM^+) interneurons. Importantly, the interneuron subtypes differ not only in the expression of molecular markers, but also in morphology, subcellular targeting, cellular properties, developmental origin and function. Figure adapted from Ehrlich et al. (Ehrlich et al. 2009) (left) and McDonald & Mascagni (McDonald and Mascagni 2001) (right).

2.3.3 The role of the basolateral amygdala in fear learning

The BLA is a key player for fear conditioning and extinction and is essential to form the association between CS and US (**Fig. 3**) (LeDoux 2000). The BLA was first implicated in fear behavior in the 1950s, when bilateral amygdala lesions made monkeys less fearful (Weiskrantz 1956), while electrical amygdala stimulation elicited strong fear responses (Delgado et al. 1956).

More recent studies, using neurotoxic lesions in rodents, support these findings, showing that pre-training lesions prevent fear acquisition (Cousens and Otto 1998), while post-

training lesions impair memory recall (Campeau and Davis 1995b; Cousens and Otto 1998). These lesions do not affect locomotion or pain sensitivity, supporting a role of the BLA in learning (Campeau and Davis 1995b; Maren 1999). These findings were reproduced using acute pharmacological inactivations with the $GABA_A$ receptor agonist muscimol, excluding compensatory effects after lesions. Pre-training or pre-recall BLA muscimol injections prevented the acquisition or expression of fear memories, respectively (Wilensky et al. 1999). Importantly, learning was only impaired upon pre-training treatment, while injections immediately after training had no effect on fear acquisition (Wilensky et al. 1999), suggesting that neuronal activity in the BLA during training is necessary for fear learning. Furthermore, pharmacological manipulations targeting NMDARs revealed that synaptic plasticity in the BLA is necessary both for the acquisition and the extinction of fear (Falls et al. 1992; Kandel 2001; Lu et al. 2001; Rodrigues et al. 2001; Schafe et al. 2001; Desgranges et al. 2008).

Importantly, learning-induced plasticity could indeed be observed in extracellular recordings of LA neurons as an enhancement of short-latency CS-evoked activity (Quirk et al. 1995; Quirk et al. 1997; Rogan et al. 1997). Thalamic, not cortical afferents to LA neurons are likely the initial site of this plasticity. Not only is the short-latency, thalamic component of the CS response potentiated first, but plasticity is also observed earlier in LA than in cortical neurons. This plasticity is stimulus-specific, given that only CS+, and not CS-, responses are enhanced after a discriminative fear conditioning paradigm (Collins and Pare 2000).

While plasticity in the BLA is needed for learning and CS responses are altered after fear conditioning, the initial plasticity could occur upstream of the amygdala. Indeed, neurons in the medial geniculate nucleus, the main source of thalamic afferents to the BLA, exhibit plasticity upon fear conditioning (Gerren and Weinberger 1983). However, this plasticity is prevented by inactivation of the BLA during conditioning, suggesting that the BLA induces the plastic changes in the thalamus and not vice versa (Maren et al. 2001). These findings support the notion of the BLA as the key player in fear learning.

Fear extinction is mediated by the concerted action of a network of highly interconnected brain areas, including the BLA, the hippocampus and the mPFC (Canteras and Swanson 1992; McDonald et al. 1996). The hippocampus provides environmental information to determine the context in which extinction takes place and in which fear responses will be inhibited (Corcoran et al. 2005). The mPFC is implicated in general behavioral inhibition (Quirk et al. 2006; Graybiel 2008). Neuronal activity in the IL subdivision of the mPFC is correlated with the acquisition and expression of extinction and electrical stimulations accelerate extinction (Milad and Quirk 2002; Maren and Quirk 2004) (see chapter 2.3.8).

Embedded in a vast network of brain areas, the BLA is a critical player for fear extinction. Both human and rodent studies show an enhanced activation of the BLA by extinction (Herry and Mons 2004; Phelps et al. 2004). Neuronal activity in the BA is necessary for the acquisition of extinction and for switches between low and high fear states (Herry et al.

2008). Furthermore, extinction is dependent on synaptic plasticity in the BLA. Pharmacological activation or blockade of NMDAR function enhances or impairs fear extinction, respectively (Falls et al. 1992; Walker et al. 2002). Pharmacological blockade of the Ca^{2+} permeable NR2B subunit of NMDARs or interference with the MAPK/ERK pathway also impair extinction learning (Herry et al. 2006; Sotres-Bayon et al. 2007). An important question is whether fear conditioning and extinction are mediated by the same neurons and circuits within the amygdala, or whether they are based on distinct neuronal substrates. Growing evidence suggests that these processes have different correlates. Recent work has shown for instance that the basal amygdala (BA) harbors fear and extinction neurons – principal cells which exhibit distinct activity after conditioning and extinction (Herry et al. 2008). An important step to determine how this functional classification correlates with specific projection targets of these neurons is described in chapter 3.3 (Senn et al. 2013).

Importantly, the activity and thereby also the plasticity of the neuronal circuits in the BLA is regulated by local inhibitory interneurons (see chapter 2.3.5). While the recruitment of inhibitory circuits has been implicated in fear extinction (Harris and Westbrook 1998; Rosenkranz et al. 2003; Chhatwal et al. 2005; Heldt and Ressler 2007), their role in conditioning is elusive. An investigation of the differential roles of distinct interneuron subtypes in the acquisition of fear is described in chapter 3.2 (Results – Publication II) (Wolff et al. 2013).

2.3.4 Synaptic plasticity in the basolateral amygdala

How does the pairing of a tone and a footshock during fear conditioning lead to learning on a cellular level? Numerous studies provided considerable evidence that LTP at sensory afferents to the BLA underlies the acquisition of fear memories (LeDoux 2000; Goosens and Maren 2004; Sah et al. 2008). Pharmacological and molecular manipulations that block or occlude NMDAR-dependent LTP also impair fear conditioning (Rogan and LeDoux 1995; McKernan and Shinnick-Gallagher 1997; Bauer et al. 2002; Goosens and Maren 2004; Rumpel et al. 2005; Humeau et al. 2007). Furthermore, highly-plastic cortical and thalamic inputs converge in the BLA (LeDoux et al. 1991; Romanski et al. 1993) and their synaptic transmission is enhanced after fear conditioning *ex vivo* (McKernan and Shinnick-Gallagher 1997; Tsvetkov et al. 2002) and *in vivo* (Quirk et al. 1997; Rogan et al. 1997; Goosens et al. 2003). Finally, LTP can be induced both *in vitro* and *in vivo* at these sensory afferents to the amygdala (Rogan and LeDoux 1995; Huang and Kandel 1998; Doyere et al. 2003). Together, this suggests that synaptic plasticity underlies fear learning and represents one of the strongest established links between LTP and behavioral learning.

But which molecular processes underlie LTP at the sensory afferents to individual BLA neurons upon fear conditioning (Blair et al. 2001)? PNs in the BLA receive thalamic and cortical auditory inputs at their dendrites (Farb and LeDoux 1997; McDonald 1998) (**Fig. 5**).

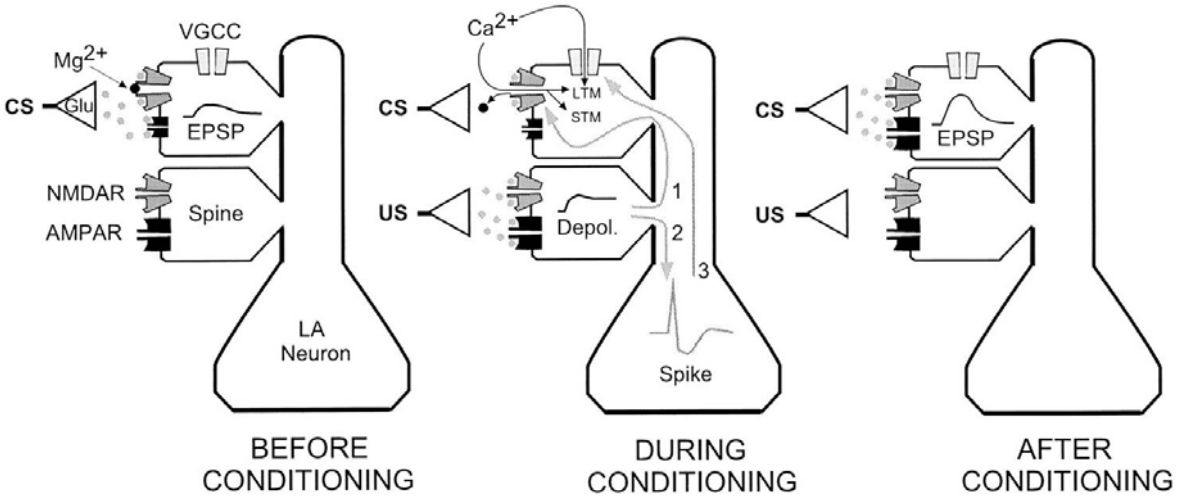


Figure 5 Shock-dependent strengthening of auditory inputs in the BLA.
Before fear conditioning (left), presentation of the tone CS induces glutamate release at auditory afferents to the BLA. Glutamate binds to AMPARs and NMDARs, eliciting only a small EPSP, without removing the Mg^{2+} block of the NMDARs. However, upon strong postsynaptic depolarization by the shock US and removal of the Mg^{2+} block, the bound glutamate allows for opening of the NMDARs and Ca^{2+} influx (middle, arrow 1). This entry of Ca^{2+} induces short-term plasticity at the synapse. For the induction of long-term memory, the additional activation of voltage-gated calcium channels (VGCCs) is necessary. VGCCs are opened, when a US-elicited spike (middle, arrow 2) backpropagates into the dendrites and collides with CS-evoked EPSPs (middle, arrow 3). The concomitant influx of Ca^{2+} via NMDARs and VGCCs triggers a signaling cascade which leads to long-term changes in synaptic strength. Long-term potentiation (LTP) can enhance AMPAR currents, so that after fear conditioning CS-evoked EPSPs are enlarged (right). This associative plasticity likely represents the cellular correlate of fear learning and memory formation. Figure adapted from Blair et al. (Blair et al. 2001).

An auditory stimulus leads to glutamate release at these synapses, causing opening of AMPARs and fast, but small excitatory postsynaptic potentials (EPSPs) (Hestrin et al. 1990). In contrast, NMDARs, which function as coincidence detectors of concerted neuronal activity, will not be opened by glutamate prior to fear conditioning, since the tone-mediated depolarization of the cell is not sufficient to release the NMDAR Mg^{2+} block. However, during conditioning, the temporally contingent membrane depolarization caused by the shock allows the opening of NMDARs and concomitant Ca^{2+} influx. In addition, US-induced backpropagating action potentials may collide with CS-triggered EPSPs, causing their amplification and the opening of voltage-gated calcium channels (VGCCs), allowing for further Ca^{2+} influx (Mermelstein et al. 2000; Blair et al. 2001; Stuart and Hausser 2001). The conjunctive Ca^{2+} entry via NMDARs and VGCCs leads to the activation of second messenger pathways and ultimately to a strengthening of the CS-activated input synapses and long-term plasticity (Malenka and Bear 2004) (**Fig. 5**).

In summary, synaptic plasticity at sensory afferents in the BLA underlies fear learning (Blair et al. 2001). These cellular and molecular changes are critically dependent on the activation of PNs by both the tone and the shock. This implies that mechanisms which regulate the activity of PNs, like neuromodulation or inhibition, also affect plasticity and learning.

2.3.5 Inhibition in the basolateral amygdala

The activity of BLA PNs, evoked by sensory stimuli, is a major factor in the acquisition of memories and in the underlying synaptic plasticity. A key mechanism to control this activity is inhibition (Ehrlich et al. 2009). Spontaneous inhibition in the BLA is strong and fear conditioning and extinction lead to opposite changes in inhibitory transmission and expression of inhibition-related genes (Harris and Westbrook 1998; Chhatwal et al. 2005; Heldt and Ressler 2007; Ehrlich et al. 2009). However, our knowledge about the detailed mechanisms of how inhibition regulates fear learning and especially about the roles of distinct subtypes of inhibitory interneurons is very limited. The BLA is a cortex-like structure and the composition of interneurons strongly resembles neocortical circuits in marker expression, basic properties and connectivity (Freund and Buzsaki 1996; Sah et al. 2003; Markram et al. 2004; Somogyi and Klausberger 2005; Ehrlich et al. 2009; Pape and Pare 2010; Fishell and Rudy 2011; Spampanato et al. 2011) (**Fig. 4**). Notably, salient sensory stimuli evoke distinct cellular responses in different IN subtypes in anaesthetized animals (Bienvenu et al. 2012). This suggests that different IN subtypes in the BLA also have different functions during fear learning and may control plasticity in distinct ways. To understand the neuronal circuitry underlying fear learning, it is therefore essential to elucidate the different roles of defined IN populations.

The most abundant IN subtype in cortex and the BLA expresses the calcium-binding protein parvalbumin (PV; **Fig. 4**) (Kawaguchi and Kubota 1997; Markram et al. 2004; Somogyi and Klausberger 2005; Rudy et al. 2011; Atallah et al. 2012). PV^+ INs are typically fast-spiking, exhibiting high firing rates and narrow spikes (Rainnie et al. 2006; Woodruff and Sah 2007b). They preferentially make perisomatic synapses, where they can control the activity and spike output of their target cells (McDonald and Betette 2001; Somogyi and Klausberger 2005; Muller et al. 2006; Atallah et al. 2012). PV^+ INs do not only contact glutamtergic PNs, but also other INs, including PV^+ INs (Muller et al. 2005; Woodruff and Sah 2007b). In the hippocampus PV^+ INs interact for instance with somatostatin-expressing (SOM^+) INs to regulate pyramidal cell output (Lovett-Barron et al. 2012). SOM^+ INs mainly target the distal dendrites of their target cells (Muller et al. 2007a), where they can exhibit efficient control over the impact of synaptic inputs (Gentet et al. 2012; Chiu et al. 2013).

A main part of this thesis addresses the role of BLA PV^+ and SOM^+ INs in fear learning and how they are integrated into neuronal circuits (Wolff et al. 2013). In line with the notion that inhibition is a general mechanism for the control of activity and plasticity, another part of this thesis focuses on the role of inhibition during fear learning in auditory cortex (Letzkus et al. 2011).

2.3.6 The fear circuitry beyond the amygdala

As a central player in fear learning, the amygdala is embedded in a vast network of highly interconnected brain areas, like different cortices, the hippocampus or several brainstem nuclei, which together mediate the acquisition, expression and extinction of fear memories. Compared to the amygdala, much less is known about the role of these up- and downstream areas in fear learning. Considering that neuronal circuits can span across the entire brain, it is essential to study the role of these structures and their interplay with the amygdala to dissect fear learning. I will highlight two brain areas, crucially involved in the acquisition and extinction of fear – the auditory cortex and the medial prefrontal cortex.

2.3.7 Auditory cortex

The amygdala receives auditory inputs via two pathways, the so called low road, which transmits auditory information via subcortical structures and the high road, which also employs cortical areas (LeDoux 1996; LeDoux 2000). On the low road, auditory inputs are processed in the auditory thalamus (MGm/PIN) and relayed directly to the LA. The high road includes a further step, in which the thalamus projects to the auditory cortex, which in turn projects to the LA (LeDoux 1996; LeDoux 2000). Both pathways are activated simultaneously and fear conditioning rapidly induces plasticity at thalamic and later also at cortical afferents (Quirk et al. 1995; Quirk et al. 1997; Rogan et al. 1997).
What is the function of the high road, if auditory information reaches the amygdala already on the low road? It was suggested that while for simple tones the low road is sufficient, complex or naturalistic tones require an additional processing step in the AC to be adequately perceived and recognized (LeDoux 1996; LeDoux 2000). However, the role of the AC could not be conclusively determined. Different studies provided evidence both for and against an essential function of the AC in fear learning (Romanski and LeDoux 1992; Campeau and Davis 1995a; Boatman and Kim 2006). Importantly, CS responses in the AC also exhibit plasticity, which is dependent on cholinergic afferents from the basal forebrain and can be elicited by direct basal forebrain stimulation (Suga and Ma 2003; Ji et al. 2005; Weinberger 2007a). This argues against a mere relay function of the AC, but instead suggests an active role of the AC. In one part of my thesis (see chapter 3.1) I addressed the role of AC, whether plasticity in AC is important for fear learning and how the plasticity is regulated (Letzkus et al. 2011).

2.3.8 Medial Prefrontal Cortex (mPFC)

The term mPFC loosely refers to a collection of brain areas in the anterior part of the frontal lobe, including the anterior cingulate (CG), the infralimbic (IL), the prelimbic (PL) and the

medial orbital cortices (Groenewegen et al. 1990; Uylings and van Eden 1990). The mPFC has been strongly associated with behavioral control and inhibition and with emotional regulation (Quirk et al. 2006; Sotres-Bayon et al. 2006). Furthermore, it has been implicated in decision making, habit formation and impulsivity (Graybiel 2008; Euston et al. 2012) as well as in drug seeking (Peters et al. 2009).

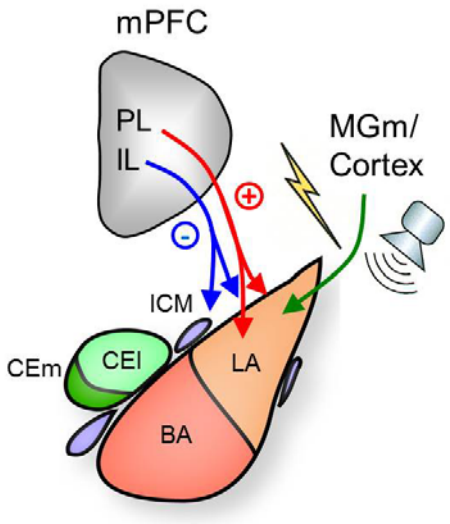


Figure 6 Opposite roles of prelimbic (PL) and infralimbic (IL) cortex in fear conditioning and extinction.
Both PL and IL send excitatory, glutamatergic projections to the basolateral amygdala (BLA). PL neurons exhibit fear related activity and participate in the activation of the BLA during high fear states, thereby inducing fear responses. In contrast, the IL shows fear extinction related activity. It likely suppresses fear responses by targeting of local BLA inhibitory interneurons and/or of intercalated cell masses (ICM). The ICMs are clusters of inhibitory GABAergic neurons. One of their projection targets is the amygdala output nucleus, the medial subdivision of the central amygdala (CEm), where they can probably suppress fear responses. Figure adapted from Ehrlich et al. (Ehrlich et al. 2009).

The role of the mPFC in adjusting behavior in response to behavioral stimuli suggests an involvement in fear conditioning and extinction (Quirk et al. 2006) (**Fig. 6**). In fact, early lesion studies showed effects on fear learning – however, with conflicting results, either enhancing or impairing fear learning and extinction (Rosen et al. 1992; Morgan et al. 1993). Specific targeting of the subdivisions of the mPFC could resolve these conflicts. It was shown that the IL and the PL exhibit opposing roles during the acquisition and extinction of fear (Quirk et al. 2006). Electrical microstimulation of the IL enhances, while PL stimulation impairs extinction learning (Vidal-Gonzalez et al. 2006). Pharmacological inactivation of the PL reduces fear expression (Corcoran and Quirk 2007), but lesions of IL specifically impair extinction learning (Quirk et al. 2000). These and other studies suggest that the PL is involved in fear acquisition, while the IL plays a role in extinction learning (**Fig. 6**). The differential effects of IL and PL on fear behavior may be due to their different activity and

projections. The PL shows increased activity during states of high fear and sends excitatory projections to the BA. These projections may excite BA cells, projecting to the CEA, in turn enhancing fear responses (Vertes 2004; Likhtik et al. 2005) (**Fig. 6**). In contrast, the IL exhibits extinction related neuronal activity. It was suggested that these IL responses inhibit fear-induced activity in the BA and fear behavior by activation of local BA interneurons or of the intercalated cell masses (Rosenkranz et al. 2003; Berretta et al. 2005; Likhtik et al. 2008) (**Fig. 6**).

Although it has been reported that the connectivity between the mPFC and the amygdala is bidirectional, the role of the BA-mPFC projections remained unclear. Electrical stimulations of the BA, inducing both monosynaptic excitatory and polysynaptic inhibitory responses (Floresco and Tse 2007), did not reveal any functional role of these connections. A part of this thesis (see chapter 3.3)investigates whether BA-mPFC projections play a role in fear learning and whether functional differences exist between BA neurons projecting to either the IL or the PL (Senn et al. 2013).

2.4 Technical advances for the study of neuronal populations in associative learning

In the past a broad range of techniques was very successfully used to study both the function of specific brain areas and the molecular and cellular underpinnings of learning in individual neurons. However, the study of neuronal circuits was hindered by the lack of appropriate techniques to target defined neuronal populations and to selectively monitor and manipulate their activity during behavior.

Classical techniques either lack cell type specificity, temporal resolution or cannot demonstrate causal relationships. Lesions and electrical stimulations do not allow to target specific neuronal populations, but only brain areas of different sizes. Pharmacological approaches usually lack the temporal specificity to dissect differential effects during a behavioral session and are often not specific for defined neuronal populations. Electrophysiological recordings *in vivo* can typically neither reveal the exact identity of the recorded cell, nor can they prove causal relationships between activity and behavior, but only show correlations. *Ex vivo/ in vitro* approaches like patch-clamping in brain slices provide important information about cellular functions and molecular mechanisms, but cannot mimic the situation *in vivo* during behavior or the activity in distributed neuronal networks.

Many of these limitations were overcome with the advent of optogenetics. Optogenetics describes the combined use of optical and genetic technologies to control cells and measure their activity in intact neuronal circuits (Deisseroth 2010). The heterologous expression of opsins - light-sensitive ion channels or pumps - can be used to activate or inhibit neurons

using light (**Fig. 7**). The most prominent opsin is Channelrhodopsin-2 (ChR2), a blue light activated, nonspecific cation channel which can be used to depolarize neurons upon illumination (Zhang et al. 2007) (**Fig. 7**). Until today a multitude of other opsins have been discovered and engineered, which are activated by different wavelengths of light and which have distinct properties (Yizhar et al. 2011). Besides excitatory also inhibitory opsins are available, the two most important ones being Halorhodopsin (NpHR) (Zhang et al. 2007) and Archaerhodopsin (ARCH) (Chow et al. 2010) (**Fig. 7**).

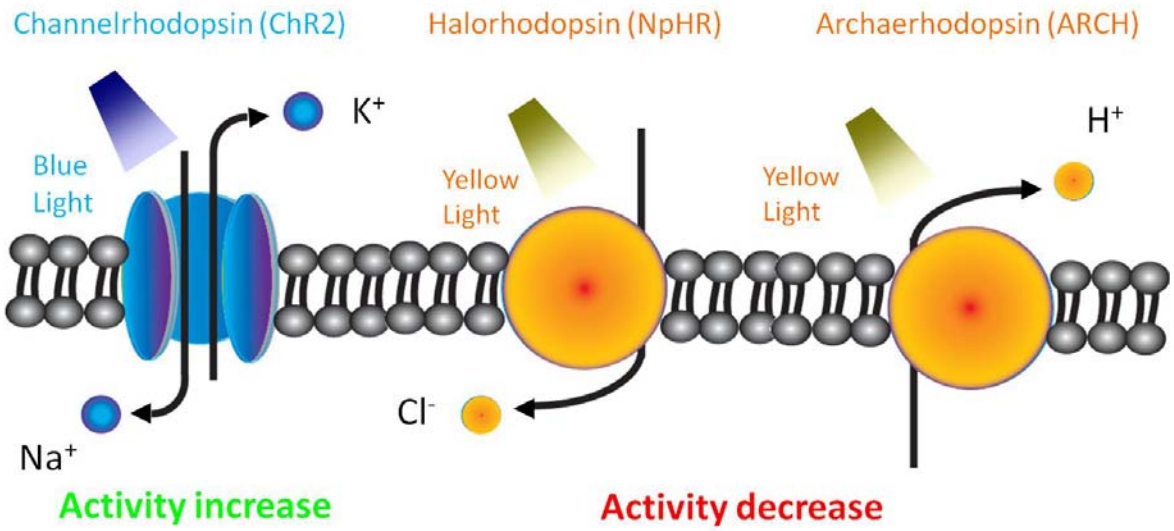


Figure 7 Opsins for optogenetic manipulations of neuronal activity.
Channelrhodopsin-2 (ChR2) is an unselective cation channel, which is activated by blue light. The influx of Na^+ cations upon its opening depolarizes the expressing neuron and induces firing. Halorhodopsin (NpHR) is a Cl^- pump, which is activated by yellow light. The active transport of Cl^- into the cell leads to a strong hyperpolarization and decreased activity. Archaerhodopsin (ARCH) is a proton pump, which is also activated by yellow light. Its active export of protons leads to membrane hyperpolarization and decreased neuronal activity.

Importantly, the expression of opsins can be targeted to defined neuronal populations, using different methods of DNA delivery. While transgenic mice offer stable expression in genetically defined neurons throughout the brain, more flexible viral approaches are commonly used (Yizhar et al. 2011; Johansen et al. 2012) (**Fig. 8**). The injection of viral vectors for opsin expression under a ubiquitous promoter into a certain brain area allows for locally restricted expression and optogenetic manipulations. An example for this approach is the study by Ciocchi et al. on the function of the CEA, in which targeted injections allowed for the specific manipulation of CEm activity (Ciocchi et al. 2010). For further restriction of the expression viral vectors with cell-specific promoters can be used (Yizhar et al. 2011; Johansen et al. 2012). However, only a small number of promoters is available for incorporation into viral genomes. Most commonly used for cell-type specific expression is the CRE/LoxP system. The combination of conditional, CRE-dependent viral

vectors and transgenic mice, expressing CRE under the control of specific promoters allows to express opsins locally in genetically defined neurons in the targeted brain area (Yizhar et al. 2011; Johansen et al. 2012) (**Fig. 8**). Finally, retrograde viruses like Herpes simplex virus (HSV) or Canine adeno virus (CAV), which can infect neurons via their axon terminals, can be used to express opsins in neurons which are defined by their anatomical projection target (Hnasko et al. 2006; Lima et al. 2009) (**Fig. 8**). In summary, viral and genetic strategies allow to specifically target defined neuronal populations for the expression of opsins and thereby for optogenetic manipulations.

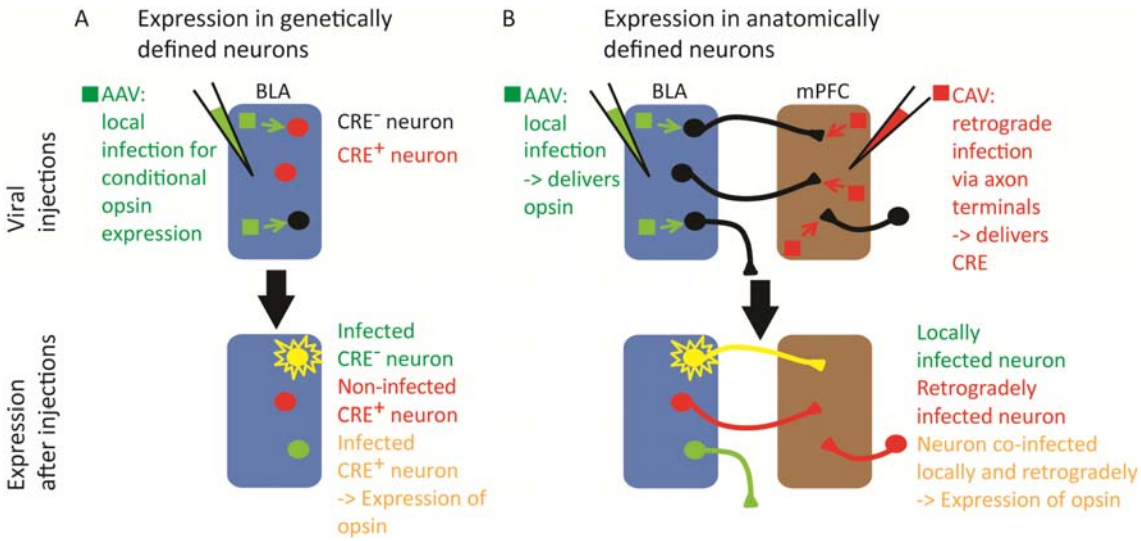


Figure 8 Viral approaches for specific targeting of genetically or anatomically defined BLA neurons. **A** Expression of opsins for optogenetic manipulations in genetically defined neurons can be achieved by a combination of transgenic mouse lines, expressing CRE recombinase only in distinct neuronal types, like PV^+ interneurons and conditional, CRE-dependent expression vectors. AAVs (green) are injected into the BLA of a CRE^+ mouse, delivering the construct for conditional opsin expression. Both CRE^+ (red) and CRE^- (black) neurons will be infected, but only CRE^+ BLA neurons will express the opsin (yellow), not CRE^- neurons (green). **B** Anatomically defined projection neurons are targeted by a combination of locally and retrogradely infecting viruses. An AAV which infects neurons locally (green) is injected into the BLA, delivering a construct for conditional, CRE-dependent opsin expression. CRE recombinase is provided by a second virus (CAV; red), which is injected into the mPFC and infects neurons via their axon terminals. Neurons which are infected by only one virus (green or red) do not express the opsin. Only BLA neurons projecting to mPFC can be infected by both viruses (yellow) and express the opsin, rendering them light-sensitive.

Importantly, optogenetics cannot only be used for the manipulation of neuronal activity, but also to identify specific neurons in electrophysiological recordings *in vivo* (Lima et al. 2009). Direct activation or inhibition of a neuron by a respective optogenetic manipulation marks it as a member of the targeted neuronal population. This allows to identify neurons e.g. in single unit recordings and to determine their physiological activity during behavior.

A central part of this thesis was the establishment and refinement of optogenetic approaches. I used molecular techniques to engineer several viral constructs for the expression of opsins and also their co-expression (Tang et al. 2009). I established optogenetic manipulations of defined neuronal populations *in vivo* during behavior and applied these to determine the role of the central medial amygdala during fear expression (Ciocchi et al. 2010). Finally, I refined the manipulations further and combined them with chronic single unit recordings for optogenetic identification of defined neurons (Letzkus et al. 2011; Rubehn et al. 2013; Senn et al. 2013; Wolff et al. 2013). All these improvements of the optogenetic approaches allowed me to perform sophisticated long-term experiments, including the simultaneous manipulation and monitoring of neuronal activity. The details of the application of optogenetics in fear conditioning are summarized in Johansen et al (Johansen et al. 2012).

2.5 Aims of this thesis

My research focused on the dissection of the neuronal circuits underlying fear learning. After establishing and refining a portfolio of techniques for both monitoring and manipulating the activity of defined neuronal populations, I aimed to reveal novel circuit mechanisms of fear learning.

In the three main studies presented here, I addressed different aspects of fear learning in the BLA and in the auditory cortex. The common questions of these studies are which role certain defined neuronal populations play during the acquisition and extinction of fear and how they are integrated in the underlying neuronal circuits. While two of the studies investigate how different GABAergic interneurons and their interactions control fear learning via disinhibitory mechanisms (Letzkus et al. 2011; Wolff et al. 2013), the third study targets anatomically defined PNs in the BLA and their differential involvement in fear acquisition and extinction (Senn et al. 2013).

The first study (Letzkus et al. 2011) addresses the role of auditory cortex in fear conditioning. In particular, it asks whether AC is a mere relay for auditory information to the amygdala, or whether plasticity in the AC is also involved in fear learning. The study mainly focuses on the mechanisms of how activity and plasticity in the AC are controlled during learning. To this end, it examines the role of inhibition and disinhibition and how these mechanisms regulate shock-induced plasticity and learning in the AC (chapter 3.1).

In my second study (Wolff et al. 2013), I investigated how activity and plasticity during the acquisition of fear are regulated on a network level. To this end, the study addresses the role of distinct interneuron subtypes in the BLA and how their interactions regulate fear learning. The study also investigates how the CS and the US are represented and processed in the BLA and what role the different interneuron types play. Finally, the role of disinhibitory mechanisms in the control of stimulus-evoked activity is tested (chapter 3.2).

The third study (Senn et al. 2013) addresses the relationship between the function of PNs in the BLA and their anatomical identity. In particular, this study investigates whether BLA neurons which are defined by their projection to different subdivisions of the mPFC have different functions. Since the IL and the PL play opposing roles in fear acquisition and extinction, the respective BLA projection neurons may also have converse functions. Furthermore, a potential overlap between the functionally defined fear and extinction neurons and the anatomically defined neurons targeted in this study was explored (chapter 3.3).

3. RESULTS

3.1 Publication I

A disinhibitory microcircuit for associative fear learning in auditory cortex

Johannes J. Letzkus[1,3]*, Steffen B.E. Wolff[1]*, Elisabeth M.M. Meyer[1], Philip Tovote[1], Julien Courtin[2], Cyril Herry[2], Andreas Lüthi[1,3]

*[1]Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland. [2]INSERM U862, Neurocentre Magendie, 146 Rue Léo-Saignat, 33077 Bordeaux, France. * These authors contributed equally to this work. [3] Authors for correspondence: johannes.letzkus@fmi.ch and andreas.luthi@fmi.ch*

Nature (2011) **480**(7377):331-5

Learning causes a change in how information is processed by neuronal circuits. While synaptic plasticity, an important cellular mechanism, has been studied in great detail, we know much less about how learning is implemented at the level of neuronal circuits and, in particular, how interactions between distinct types of neurons within local networks contribute to the process of learning. Here we show that acquisition of associative fear memories depends on the recruitment of a disinhibitory microcircuit in mouse auditory cortex. Fear-conditioning-associated disinhibition in auditory cortex is driven by foot-shock-mediated cholinergic activation of layer 1 interneurons, in turn generating inhibition of layer 2/3 parvalbumin-positive interneurons. Importantly, pharmacological or optogenetic block of pyramidal neuron disinhibition abolishes fear learning. Together, these data demonstrate that stimulus convergence in auditory cortex is necessary for associative fear learning to complex tones, define the circuit elements mediating this convergence and suggest that layer-1-mediated disinhibition is an important mechanism underlying learning and information processing in neocortical circuits.

INTRODUCTION

The transformation of sensory input to an appropriate behavioral output is accomplished by neuronal circuits that process information through dynamic interactions between distinct types of neurons. Learning as an adaptive change of an animal's behavior modifies these network computations. Important cellular mechanisms underlying learning are thought to be activity-dependent synaptic plasticity (Martin and Morris 2002) and subsequent structural modifications (Hubener and Bonhoeffer 2010). In contrast to our detailed understanding of these cellular mechanisms, the factors and circuit elements controlling neuronal activity and concomitant induction of plasticity during learning in the intact animal remain poorly understood.

Neocortical interneurons exert powerful control over circuit activity by supplying inhibition to surrounding pyramidal neurons and to other interneurons (Markram et al. 2004). The balance of excitation and inhibition is critical to circuit function, and is maintained under most conditions including sensory stimulation, when thalamocortical input typically recruits feed-forward inhibition a few milliseconds after direct excitation (Gabernet et al. 2005). This form of inhibition is mainly supplied by basket cells, the most abundant type of interneuron in rodent neocortex (Markram et al. 2004). Fast-spiking basket cells establish strong synapses with high release probability on the perisomatic region of pyramidal neurons where they control firing, and express the calcium-binding protein parvalbumin (PV) (Kawaguchi and Kubota 1997; Markram et al. 2004; Freund and Katona 2007).

On top of the glutamatergic recruitment of inhibition by sensory input, several types of interneurons are also major targets of neuromodulation (Metherate 2004; Bacci et al. 2005; Kruglikov and Rudy 2008; Lawrence 2008), suggesting that these systems can profoundly affect circuit computations by shifting the excitation-inhibition balance (Vogels and Abbott 2009). This has mainly been addressed *in vitro*, revealing a staggering complexity of effects which strongly depend on interneuron type (Metherate 2004; Bacci et al. 2005; Kruglikov and Rudy 2008; Lawrence 2008). Given that neuromodulation plays a key role in learning (Suga and Ma 2003; Weinberger 2007b), it is likely that interneurons are engaged in these processes. However, the contribution of different interneuron types to learning in the intact animal remains elusive.

Auditory fear conditioning is a form of associative learning acquired by temporal coincidence of a neutral conditioned stimulus (CS, a tone) with a mild foot-shock. Fear conditioning causes prominent, long-lasting plasticity of CS responses in auditory cortex (Quirk et al. 1997; Suga and Ma 2003; Weinberger 2007b), which depends on the activity of cholinergic afferents from the basal forebrain (Suga and Ma 2003; Ji et al. 2005; Weinberger 2007b). Pairing of tones with basal forebrain stimulation can elicit similar plasticity in auditory cortex (Suga and Ma 2003; Weinberger 2007b), and an important recent study suggests an involvement of pyramidal neuron disinhibition in this process (Froemke et al. 2007). Here, we investigate the contribution of identified types of auditory cortex

interneurons to acquisition of auditory fear memory. Our results indicate that foot-shocks, via recruitment of cholinergic basal forebrain afferents, activate a disinhibitory neocortical microcircuit that is required for fear learning.

RESULTS

Nicotinic activation of L1 interneurons by foot-shocks

Fear memory acquisition critically depends on the amygdala (LeDoux 2000). In contrast, the role of auditory cortex is contentious, with evidence both for (Campeau and Davis 1995a; Boatman and Kim 2006) and against an essential function in auditory fear conditioning (Romanski and LeDoux 1992; Campeau and Davis 1995a; Boatman and Kim 2006). Based on the hypothesis that stimulus complexity determines the engagement of neocortex (LeDoux 2000), we performed differential fear conditioning using trains of upward and downward frequency-modulated sweeps as conditioned stimuli (CS, **Fig. 1a**, modulation between 5 and 15 kHz). To directly test the role of auditory cortex in this paradigm, we inhibited neuronal activity during fear conditioning by local injection of the $GABA_A$-receptor agonist muscimol into auditory cortex (**Fig. 1a, Supplementary Fig. 1**). When tested 24h later in drug-free state, this manipulation resulted in strongly reduced fear levels (**Fig. 1b**), indicating that activity in auditory cortex is required for fear learning in this paradigm.

We next focused on the circuit mechanisms underlying acquisition of fear memory in auditory cortex (defined here as areas A1 and AuV). The role of auditory cortex could be that of an essential relay for tone information to downstream structures like the amygdala. Alternatively, foot-shocks alone might also elicit activity in auditory cortex, and convergence of tone and foot-shock information during fear conditioning could then contribute to fear learning. To distinguish these alternatives, we employed 2-photon calcium imaging of neurons in the superficial layers of auditory cortex in mice anesthetized with chlorprothixene and urethane (**Fig. 1c,d**). Trains of frequency-modulated sweeps used as CS strongly activated neurons in layer 2/3 (L2/3, **Supplementary Fig. 2**), whereas mild foot-shocks applied to the hindpaws evoked little response (**Fig. 1e, Supplementary Fig. 2**). In marked contrast, the population of neurons located in L1 displayed strong activation by foot-shocks (**Fig. 1e**). Layer 1 is unique in neocortex since it contains only very few neuronal somata, almost all of which are GABAergic interneurons (Hestrin and Armstrong 1996; Zhou and Hablitz 1996; Christophe et al. 2002; Chu et al. 2003). To investigate their activation in more detail, we used 2-photon targeted loose-seal cell-attached recordings (Margrie et al. 2003) (**Fig. 1f**). These experiments revealed that L1 interneurons are tonically active (baseline frequency 3.2±0.5 Hz, n=30), and confirmed that the majority of L1 interneurons are strongly activated by foot-shocks, while a minority (23%) displayed inhibition during and after foot-shocks (**Fig. 1g, Supplementary Fig. 3**). The excitatory response was also present when foot-shocks were paired with tones, whereas L1 interneurons did not show

pronounced responses to tones alone (**Supplementary Fig. 4**). Interestingly, we observed a very similar activation of L1 interneurons by foot-shocks in primary visual cortex (**Supplementary Fig. 5**), suggesting that the relevance of this pathway may extend beyond auditory memory acquisition.

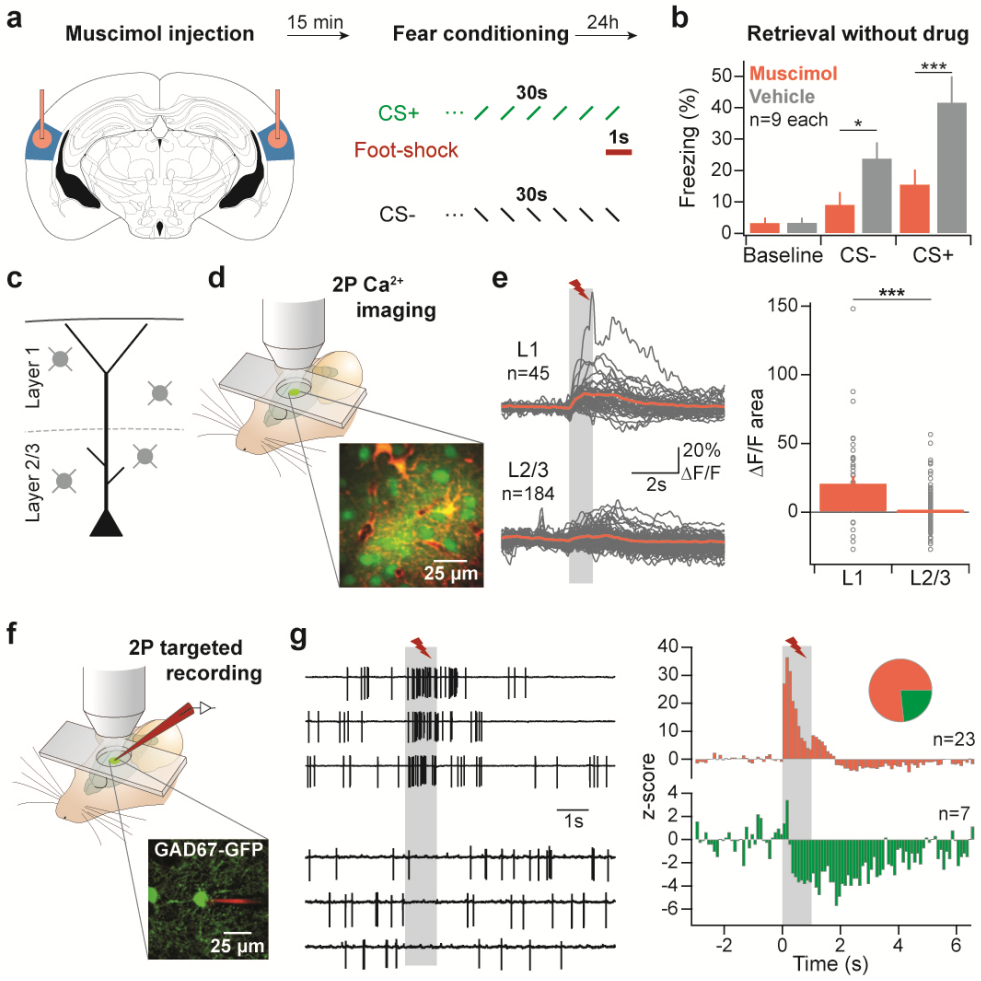


Figure 1: Foot-shock responses in auditory cortex L1 interneurons
a, Left: Injection of the $GABA_A$-receptor agonist muscimol (red) into auditory cortex (blue). Right: Differential fear conditioning protocol using frequency-modulated sweeps. **b**, Drug-free freezing one day after conditioning in a novel context. Compared to vehicle-injected mice (grey), muscimol reduced freezing to both CS^- and CS^+ (red). **c**, Cytoarchitecture of upper layers of auditory cortex (interneurons grey, pyramidal neuron black). **d**, 2-photon calcium imaging in head-fixed, anesthetized mice using OGB-1 AM (green) and sulforhodamine 101 (red, counterstains glial cells). **e**, Left: Responses in L1 and L2/3 to foot-shocks in single neurons (grey) and the population average (red). Right: L1 interneurons display much stronger activation than L2/3 neurons. **f**, 2-photon targeted loose-seal cell-attached recording of L1 interneuron (green). **g**, Example traces (left) and z-scored population peri-stimulus time histograms (right) of two types of foot-shock responses in L1 interneurons. Inset: Incidence of response type. Values are mean±s.e.m. *p<0.05, ***p<0.001. Statistical analysis in Supplementary Information.

Layer 1 is a prominent feed-back pathway in neocortex, containing both glutamatergic projections from higher cortical areas (Cauller et al. 1998; Gonchar and Burkhalter 2003) (and from non-specific thalamic nuclei (Rubio-Garrido et al. 2009)), and cholinergic afferents from the basal forebrain, the major source of acetylcholine in rodent neocortex (Wenk 1997; Mechawar et al. 2000; Weinberger 2007b; Hasselmo and Sarter 2011). We next sought to identify the afferent pathways mediating activation of L1 interneurons during foot-shocks. Local bath-application of the glutamate receptor antagonist NBQX (1 mM) strongly reduced baseline firing frequency (**Supplementary Fig. 6**), but left the foot-shock response in L1 interneurons intact (**Fig. 2a,b, Supplementary Fig. 6,** see below). In contrast, combined application of the nicotinic acetylcholine receptor (nAChR) antagonists mecamylamine and methyllycaconitine (1 and 0.1 mM, respectively) abolished the L1 interneuron response almost completely (**Fig. 2c, Supplementary Fig. 6**), and drastically reduced baseline firing (**Supplementary Fig. 6**). In agreement with the interpretation that cholinergic basal forebrain afferents generate foot-shock responses in L1 interneurons, electrical microstimulation of the basal forebrain caused strong excitation of L1 interneurons in the absence of foot-shocks (**Fig. 2d, Supplementary Fig. 6**). Activation of the basal forebrain by foot-shocks is expected to have a longer latency than thalamocortical signalling (Weinberger 2007b). Latency analysis showed that L1 interneuron activation was biphasic, with an early, glutamatergic peak (onset 10 to 20 ms after shock onset) which may originate from nonlemniscal thalamus (Rubio-Garrido et al. 2009), and a later, nicotinic peak (onset 40 to 50 ms after shock onset) that outlasted the foot-shock and contained the majority of spikes (**Fig. 2** insets). As previously reported (Christophe et al. 2002), recordings from brain slices showed that all L1 interneurons displayed responses to nicotine puffs (**Fig. 2 e,f**) which were blocked by the same antagonists that abolished foot-shock responses *in vivo* (**Fig. 2f**), and could fire L1 interneurons (**Supplementary Fig. 7**). In summary, these results indicate that activity of cholinergic basal forebrain neurons is both necessary and sufficient to fire L1 interneurons during foot-shocks, and that acetylcholine activates nAChRs on L1 interneurons. This implies that acetylcholine is released rapidly (<50 ms) after an aversive stimulus. Activation of L1 interneurons in turn is likely to play a central role in fear conditioning related plasticity in cortex.

L1 interneurons inhibit L2/3 PV$^+$-interneurons

How do foot-shock responses in L1 interneurons affect processing in the local microcircuit? There is evidence that L1 interneurons can inhibit interneurons in L2/3 during nicotinic activation (Christophe et al. 2002). Fast-spiking, PV$^+$-interneurons are the most abundant type of interneuron in neocortex (Markram et al. 2004). We injected a conditional adeno-associated virus (AAV) expressing a fluorescent marker (venus) into auditory cortex of PV-ires-Cre mice

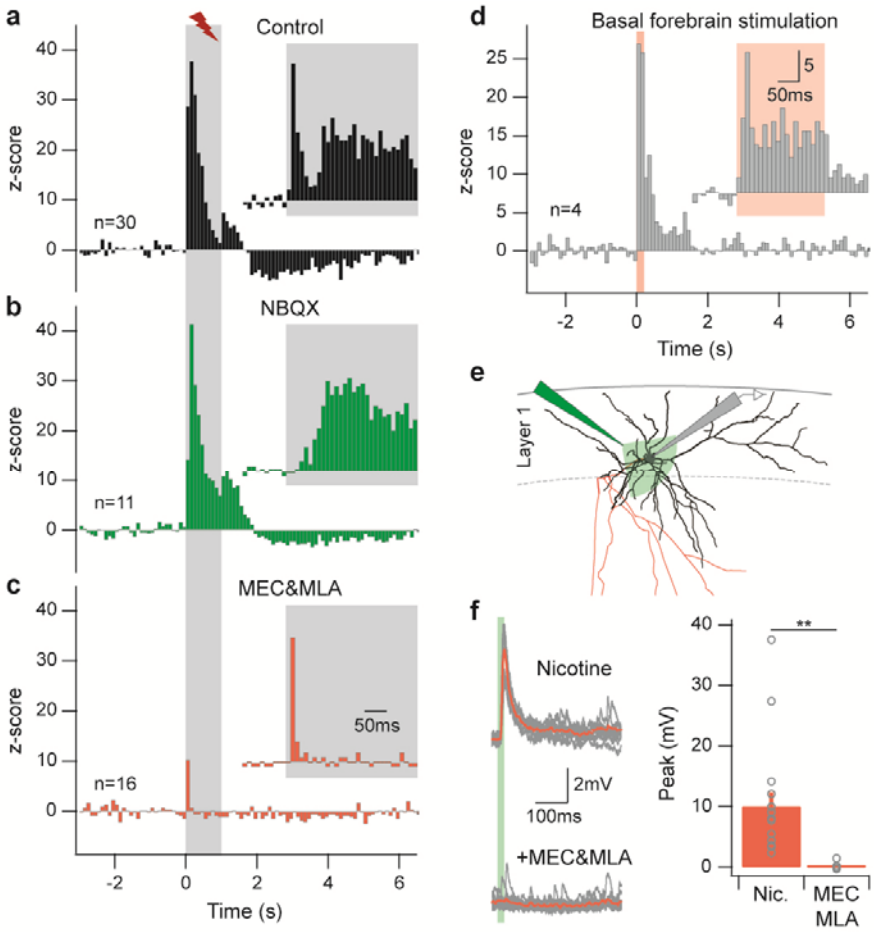


Figure 2: Nicotinic activation of L1 interneurons by foot-shocks
a, Population response of L1 interneurons to foot-shocks. Insets in **a-d** are at high temporal resolution (10 ms bins). The control response was biphasic (onset 10-20 ms after foot-shock onset). **b**, Local application of the glutamate receptor antagonist NBQX left the response intact, but selectively eliminated the early peak, indicating its glutamatergic nature (inset, onset latency 40-50 ms) **c**, Local block of nAChRs by mecamylamine and methyllycaconitine strongly reduced foot-shock responses, selectively eliminating the later, broader peak (inset). **d**, Electrical microstimulation of the basal forebrain activated L1 interneurons after 10 to 20 ms (inset, n=4). **e**, Whole-cell recording of a L1 interneuron (soma and dendrites black, axon red) in auditory cortex slices during puff application of nicotine (100 µM, 20 ms). **f**, Left: Example nicotine responses in control and after bath-application of mecamylamine (100 µM) and methyllycaconitine (0.1 µM; grey: single trials, red: average). Right: All L1 interneurons displayed nicotine responses (n=17), which were blocked by the nAChR antagonists (n=8). Values are mean±s.e.m. **p<0.01. Statistical analysis in Supplementary Information.

(Hippenmeyer et al. 2005) to selectively label these cells (**Fig. 3a**). Targeted recordings revealed that PV^+-interneurons in L2/3 are tonically active under baseline conditions (5.9±1.2 Hz, n=17), similar to fast-spiking interneurons in awake, head-fixed mice (Gentet et al. 2010). Foot-shocks caused prominent, long lasting inhibition of firing in the majority of PV^+-interneurons (88%, **Fig. 3b, Supplementary Fig. 8**), while the remaining 2 neurons displayed an excitatory response (**Supplementary Fig. 9**). Inhibition of PV^+-interneurons was

strongly reduced by the nAChR antagonists mecamylamine and methyllycaconitine (**Fig. 3b, Supplementary Fig. 8**), which also blocked excitation of L1 interneurons by foot-shocks (**Fig. 2c**). Since PV$^+$ fast-spiking interneurons lack intrinsic responses to acetylcholine (Kawaguchi and Kubota 1997; Kruglikov and Rudy 2008), this result is consistent with direct inhibition of PV$^+$-interneurons by L1 interneurons. In line with this interpretation, we observed morphological and functional synaptic contacts between L1 interneurons and PV$^+$-interneurons (**Supplementary Fig. 10**). In addition, the time-course of foot-shock responses in the two populations matches (**Supplementary Fig. 11**). Taken together, these data indicate that L2/3 fast-spiking PV$^+$-interneurons are inhibited by L1 interneurons during foot-shocks.

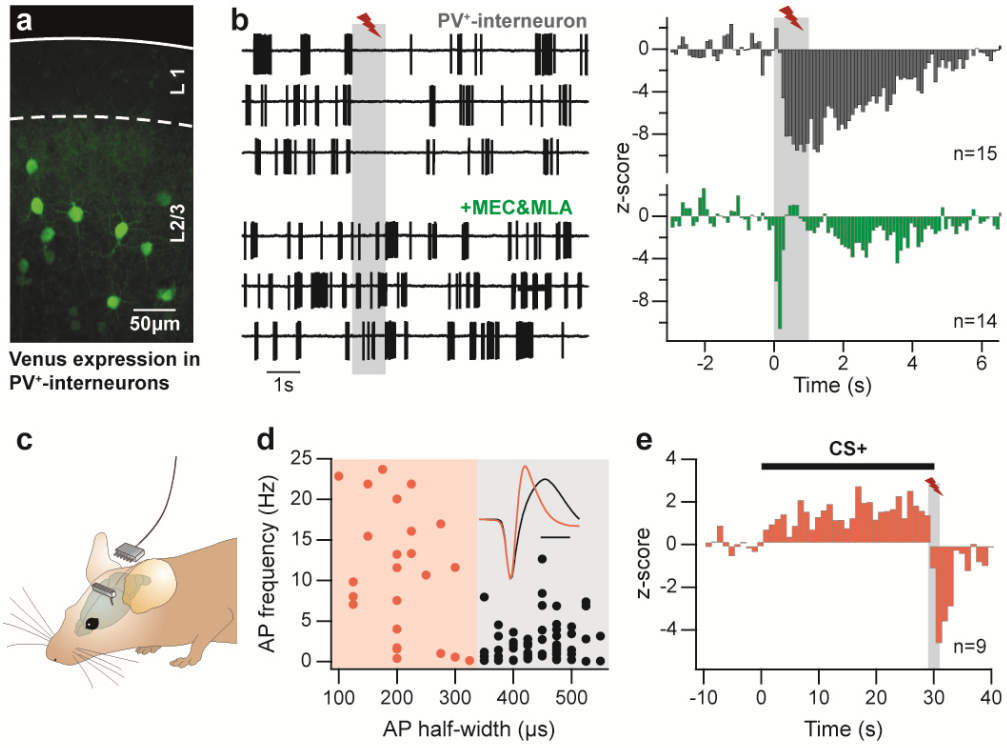


Figure 3: Aversive shocks inhibit layer 2/3 PV$^+$-interneurons
a, Expression of venus in L2/3 PV$^+$-interneurons. **b**, Cell-attached recordings from PV$^+$-interneurons in control (top) show strong inhibition of spontaneous firing by foot-shocks (left: example traces, right: population response). This response was strongly reduced in the presence of mecamylamine and methyllycaconitine (bottom). **c**, Single-unit recordings in the superficial layers of auditory cortex in freely-behaving mice. **d**, Putative interneurons (red) were separated from putative pyramidal neurons (black) by unsupervised cluster analysis. Inset: Example spike waveforms (scale bar 400 µs) **e**, z-scored activity of putative interneurons displaying inhibition by periorbital shocks during fear conditioning. Note recruitment by the CS$^+$ and subsequent inhibition by the shock (n=9 of 24 putative interneurons, **Supplementary Fig. 12**).

To test whether this mechanism is also engaged in awake, freely-moving animals, we implanted mice with single-unit recording electrodes in the superficial layers of auditory cortex (**Fig. 3c, Supplementary Fig. 12**). Putative interneurons were distinguished from

putative pyramidal neurons using unsupervised cluster analysis (**Fig. 3d, Supplementary Fig. 12**). Recordings during fear conditioning confirmed that a large population of putative interneurons is inhibited during and after an aversive shock (**Fig. 3e, Supplementary Fig. 12**). The same neurons were activated by the CS, indicating that the shock removes feed-forward inhibition in pyramidal neurons during auditory input. A similar proportion of putative interneurons displayed either excitation or no response to shocks (**Supplementary Fig. 12**). These data are consistent with the interpretation that excitation of L1 interneurons by aversive stimuli serves to remove both spontaneous and feed-forward inhibition provided by PV^+-interneurons to surrounding pyramidal neurons in behaving mice.

Disinhibition of L2/3 pyramidal neurons

PV^+ basket cells provide strong, perisomatic inhibition to local pyramidal neurons (Kawaguchi and Kubota 1997; Markram et al. 2004; Freund and Katona 2007; Kruglikov and Rudy 2008). To directly test whether disinhibition is the main effect of foot-shocks in auditory cortex L2/3 pyramidal cells, we employed whole-cell recordings. Foot-shocks elicited a depolarization from rest in all neurons tested (5±1.1 mV, n=6, **Fig. 4a**). The amplitude increased at more depolarized membrane potentials (**Supplementary Fig. 13**), consistent with a response mediated by strong disinhibition and weak excitation (Manookin et al. 2008). Application of nAChR antagonists converted the foot-shock response to a net inhibition (**Fig. 4a)**, suggesting an involvement of the disinhibitory circuit described here, and perhaps indicating that its block unmasks a component of inhibition recruited by foot-shocks. Finally, recordings under conditions isolating inhibitory postsynaptic currents (IPSCs) (Froemke et al. 2007) revealed a drastic reduction of IPSC frequency during and after the foot-shock (**Fig. 4b**). These results suggest that inhibition of PV^+-interneurons is a dominant influence shaping foot-shock responses in pyramidal neurons, and are in line with the observation that basal forebrain stimulation causes disinhibition of pyramidal cells (Froemke et al. 2007).

Given that acetylcholine can affect many stages of neocortical information processing (Metherate 2004; Bacci et al. 2005; Kruglikov and Rudy 2008; Lawrence 2008), we next aimed to determine how sensory input interacts with foot-shock-mediated disinhibition. We imaged calcium responses to tones in L2/3 of auditory cortex, and compared them to presentations of the same tones in conjunction with foot-shocks (**Fig. 4c, Supplementary Fig. 14)**. This experiment revealed that foot-shocks cause a strong enhancement of the calcium signal integral, suggesting that tone/shock compounds elicit much greater activity than tones alone. A similar observation was made with extracellular recordings in freely moving mice, where coincidence of tone and shock excited putative pyramidal neurons much more than tone alone (**Fig. 4d**). This effect was highly supralinear in both experiments since shocks alone elicited almost no activity in these neurons (**Supplementary Fig. 14**). In

summary, we provide evidence that L2/3 pyramidal neurons are disinhibited by aversive stimuli via inhibition of PV^+-interneurons.

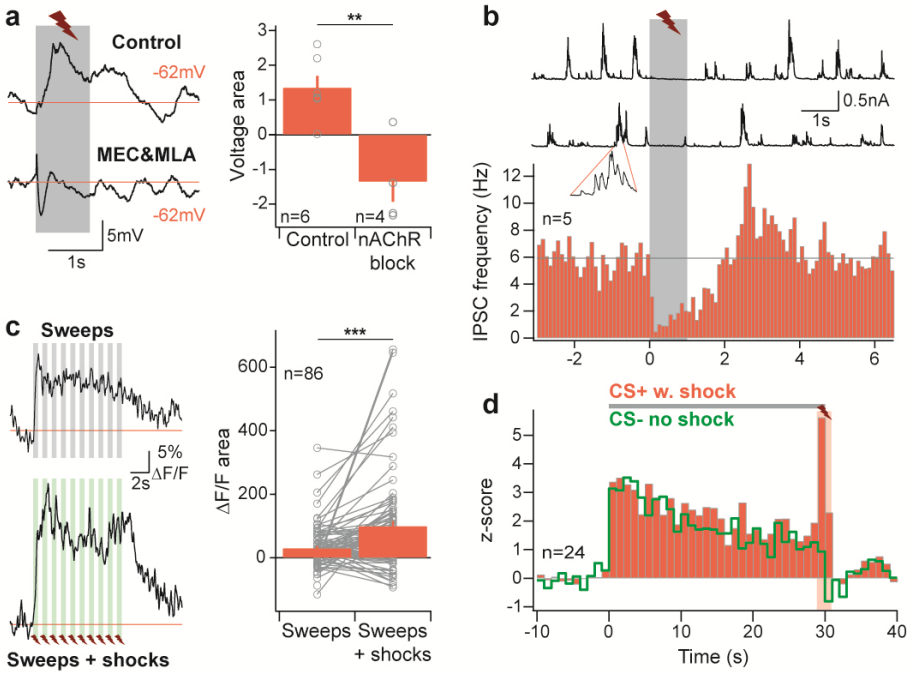


Figure 4: Aversive shocks disinhibit L2/3 pyramidal neurons
a, Top left: *In vivo* whole-cell current-clamp recording of an auditory cortex pyramidal neuron during foot-shocks. Bottom left: Foot-shock response in the presence of mecamylamine and methyllycaconitine. Response area recorded at rest was reduced by nAChR antagonists (right). **b**, Whole-cell voltage-clamp recordings of inhibitory postsynaptic currents in pyramidal neurons show a strong reduction in IPSC frequency during and after foot-shocks (top: example traces, bottom: population data, V_{hold} 0 to +20 mV). **c**, Left: Example response of a L2/3 neuron to frequency-modulated sweeps (top) and sweeps paired with foot-shocks (bottom) recorded with 2-photon calcium imaging. Right: Foot-shocks caused a threefold enhancement of sweep response area in L2/3. **d**, Single-unit recordings of putative pyramidal neurons with significant response to the CS^+ in freely-behaving mice during fear conditioning. Note strong activation by coincidence of sweep and shock. Values are mean±s.e.m. **p<0.01, ***p<0.001. Statistical analysis in Supplementary Information.

Auditory cortex disinhibition is required for fear learning

To determine whether activation of nAChRs contributes to fear learning, we locally applied mecamylamine and methyllycaconitine into auditory cortex prior to fear conditioning (**Fig. 5a, Supplementary Fig. 1**). When tested 24h later in drug-free state, this manipulation resulted in strongly reduced fear levels (**Fig. 5b**), consistent with a requirement for nicotinic activation of L1 interneurons. However, L1 interneurons are not the only circuit elements expressing nAChRs (Metherate 2004; Bacci et al. 2005; Lawrence 2008). To further test whether disinhibition specifically during the foot-shock contributes to fear learning, we employed channelrhodopsin-2 (ChR-2) (Zhang et al. 2007) expression in PV^+-interneurons

(**Fig. 5c, Supplementary Fig. 15, 16**). Mice with chronic, bilateral optic fibre implantation (**Fig. 5d**) were subjected to differential fear conditioning in which optogenetic stimulation of PV^+-interneurons occurred during and for 5 s after the foot-shock (**Fig. 5e**), the period during which we observed inhibition of these neurons (**Fig. 3b**). When tested 24 h later without optogenetic stimulation, these mice displayed strongly reduced fear responses to the conditioned stimulus compared to sham-injected littermates (**Fig. 5f**). Reconditioning without optogenetic manipulation yielded normal fear learning (**Fig. 5f**). In addition, we ruled out the possibility that foot-shock perception was perturbed by optogenetic stimulation and that laser illumination itself was perceived as a conditioned stimulus (**Supplementary Fig. 17**). Together, these data indicate that nicotinic disinhibition of auditory cortex selectively during the foot-shock is required for associative fear learning.

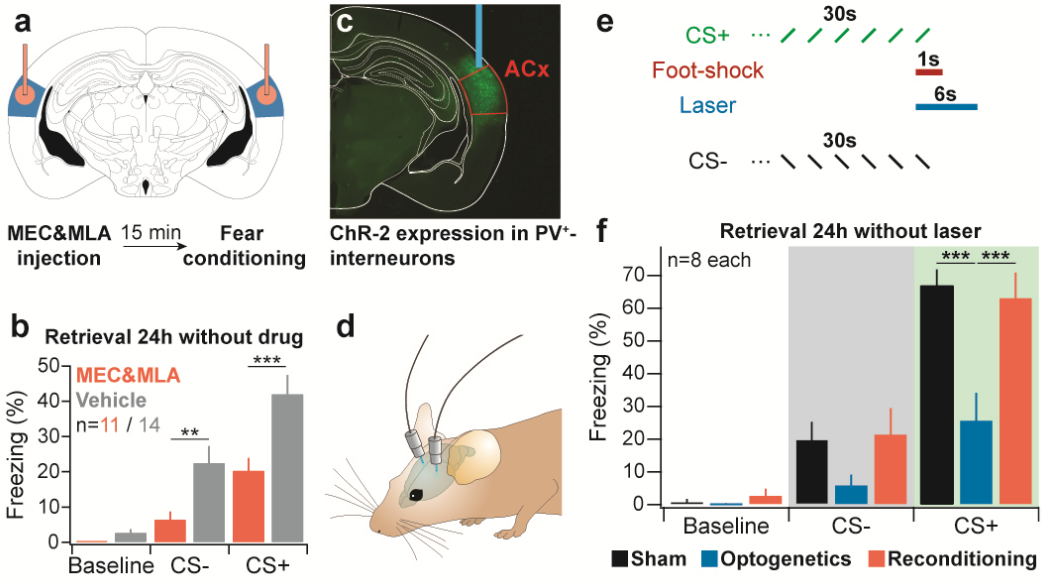


Figure 5: Auditory cortex disinhibition is required for fear learning
a, Injection of mecamylamine and methyllycaconitine (red) into auditory cortex (blue) prior to fear conditioning. **b**, Drug-free freezing in a novel context one day after conditioning. Compared to vehicle-injected mice (grey), nAChR antagonism reduced freezing to both CS^- and CS^+ (red). **c**, Stimulation of ChR-2 expressing PV^+-interneurons (green) in auditory cortex (red) via an optic fibre (blue). **d**, Optogenetic manipulation in freely-behaving mice. **e**, Differential fear conditioning protocol with optogenetic stimulation during and for 5 s after every foot-shock. **f**, Freezing in a novel context without laser stimulation one day after conditioning. Compared to identically treated sham-injected littermates (black), virus-injected mice (blue) exhibit drastically reduced freezing to the CS^+. Reconditioning without optogenetic stimulation yielded strongly enhanced freezing (red) that was indistinguishable from sham. Values are mean±s.e.m. **p<0.01, ***p<0.001. Statistical analysis in Supplementary Information.

DISCUSSION

Using targeted recordings from identified populations of auditory cortex neurons in conjunction with single-unit recordings, pharmacological and optogenetic manipulations in behaving mice, we have identified a disinhibitory microcircuit required for associative learning. Our data show that L1 interneurons play a central role in conveying information about an aversive stimulus to auditory cortex. Activity of L1 interneurons was tightly controlled by endogenous acetylcholine released from basal forebrain cholinergic projections, which determined baseline firing and acutely activated the majority of L1 interneurons during foot-shocks, while a sub-set responded with inhibition. Given that all L1 interneurons express functional nAChRs (**Fig. 2**, (Christophe et al. 2002)), a likely source of this inhibition are synaptic interactions within L1 (Chu et al. 2003). Layer 1 contains several morphologically distinct sub-types of interneurons (Hestrin and Armstrong 1996; Zhou and Hablitz 1996; Christophe et al. 2002; Chu et al. 2003; Wozny and Williams 2011), and it remains to be determined whether foot-shock response type correlates with morphology. A strikingly similar activation by foot-shocks was also observed in primary visual cortex, which could underlie aspects of contextual fear learning and suggests that cholinergic activation of L1 interneurons may be a general feature of neocortical organization. Together, these results add to mounting evidence that L1 is a prominent locus of feed-back signalling in neocortex (Cauller et al. 1998; Mechawar et al. 2000; Gonchar and Burkhalter 2003) and begin to define how feed-back signals interact with thalamocortical feed-forward information during memory acquisition. Interestingly, L1 interneurons also receive prominent corticocortical feed-back (Gonchar and Burkhalter 2003), and are responsive to dopamine (Wu and Hablitz 2005) and serotonin (Foehring et al. 2002), suggesting that, depending on the nature of the learning task, different systems can feed into the microcircuit described here.

Foot-shocks reduced both spontaneous and feed-forward perisomatic inhibition provided by basket cells, thus disinhibiting pyramidal neurons. Importantly, we cannot rule out that electrotonically remote dendritic sites received unchanged or even increased inhibition, for instance from L1 neurogliaform cells (Chu et al. 2003; Wozny and Williams 2011), which could serve to compartmentalize induction of synaptic plasticity. Nicotinic enhancement of inhibitory input to pyramidal neurons has been observed *in vitro* (Metherate 2004; Couey et al. 2007). Therefore, the level of ongoing inhibition onto pyramidal neurons, as a prerequisite for disinhibition, may determine whether the net effect of nAChR activation is inhibitory or disinhibitory. Disinhibition is an attractive mechanism for learning since it is *permissive* for strong activation and concomitant plasticity induction, but not *causative*. Rather, the available sensory input at the time of the aversive stimulus can select the circuit elements for plasticity induction in a bottom-up fashion. In addition, other cholinergic actions like enhancement of thalamocortical transmission (Gil et al. 1997; Metherate 2004) and reduced GABA release from basket cell synapses (Kruglikov and Rudy 2008) may have acted in synergy with the microcircuit described here to boost sensory responses.

Disinhibition of pyramidal neurons by foot-shocks in turn likely gated induction of activity-dependent plasticity in auditory cortex (Ji et al. 2005; Froemke et al. 2007) and at cortical afferents to the amygdala (LeDoux 2000). In parallel, cholinergic activation of L1 interneurons may also contribute to memory expression, since basal forebrain neurons acquire a conditioned response during learning (Acquas et al. 1996; Wenk 1997). Irrespective of the plasticity loci, our results delineate a role for auditory cortex in fear conditioning to complex tones that goes beyond mere signalling of tone information to the amygdala (Romanski and LeDoux 1992; Campeau and Davis 1995a; LeDoux 2000; Boatman and Kim 2006). Rather, we observe that stimulus convergence and concomitant auditory cortex disinhibition is essential for fear learning. It is likely that the use of complex, naturalistic tones has emphasized the role of auditory cortex (LeDoux 2000), since fear conditioning to pure tones is often unaffected by auditory cortex lesions (Romanski and LeDoux 1992; Campeau and Davis 1995a) (but see Refs. (Campeau and Davis 1995a; Boatman and Kim 2006)).

Beyond memory formation, cholinergic transmission is well known to mediate functions such as arousal and attention (Wenk 1997; Hasselmo and Sarter 2011). Interestingly, the transition from quiet wakefulness to active whisking is associated with a sharp drop in firing of fast-spiking interneurons in barrel cortex (Gentet et al. 2010), which is reminiscent of foot-shock responses in these neurons, and perhaps implies that cholinergic gating by inhibition of PV^{+}-interneurons is a general hallmark of active brain states.

METHODS SUMMARY

Male C57BL6/J mice (1.5 to 6 months old) were used. 2-photon calcium imaging was performed using the membrane permeant calcium dye OGB-1 AM (Stosiek et al. 2003) and sulforhodamine 101 to counterstain glial cells (Nimmerjahn et al. 2004). Targeted patch clamp recordings (Margrie et al. 2003) were performed using pipettes (4-10 MΩ) filled with extracellular solution (cell-attached), K-methanesulfonate or Cs-methanesulfonate-based intracellular solution supplemented with fluorescent dye. Neurons were identified as follows: L1 interneurons in wildtype and GAD-67-GFP mice based on location (30-70 µm below pia); PV^+-interneurons by expression of the fluorescent protein venus; Pyramidal neurons based on morphology and spiny dendrites after dye filling. *In vitro* current-clamp recordings were performed in acute slices of auditory cortex (6-8 week old mice). Chronically implanted electrodes were used for extracellular recordings in freely-behaving mice. In differential fear conditioning, the CS^+ but not the CS^- was paired with a mild foot-shock. Freezing was quantified automatically (Ciocchi et al. 2010). For optogenetic intervention, mice were stereotaxically injected with an AAV2/7 coexpressing ChR-2 (Zhang et al. 2007) and venus. Bilateral activation of ChR-2 was performed using chronically implanted optic fibres coupled to a laser.

Acknowledgements

We thank all members of the Lüthi lab for discussions and comments. We thank Björn Kampa and Thomas Oertner for advice on 2-photon imaging, Rainer Friedrich and Jan Gründemann for comments on the manuscript, Paul Argast, Julia Lüdke and Christian Müller for excellent technical assistance, Hanna Zielinska for preparation of artwork and Silvia Arber, Yuchio Yanagawa and Karl Deisseroth for generously sharing materials. This work was supported by grants from the Swiss National Science Foundation (J.J.L. Ambizione; A.L.), the French National Research Agency (C.H., ANR-2010-BLAN-1442-01), a Marie-Curie fellowship (J.J.L.), a Schering Foundation fellowship (S.B.E.W.), a Fonds AXA pour la Recherche fellowship (J.C.), the Aquitaine Regional Council (C.H.) and the Novartis Research Foundation.

Author contributions

J.J.L. initiated the project and performed most experiments and data analyses. S.B.E.W. established optogenetic manipulation. S.B.E.W. and P.T. helped with optogenetic experiments. E.M.M.M. performed and analyzed *in vitro* experiments. P.T. performed and analyzed *in vivo* pharmacology. J.C. and C.H. performed and analyzed single-unit recordings. J.J.L. and A.L. conceived the project and wrote the manuscript. All authors contributed to the experimental design and interpretation, and commented on the manuscript.

SUPPLEMENTARY DATA

Statistical analysis

Figure 1b: Effects of muscimol injection on acquisition of fear memory. Comparison of freezing during memory retrieval (24h after conditioning, drug-free) between a group of mice injected with muscimol during training (n=9) and a vehicle-injected control group (n=9). Two-way ANOVA indicated significant main effects between treatment groups: $F(1,53)=12.536$, $p<0.001$; as well as between test phases (baseline, CS-, CS+): $F(2,53)=14.451$, $p<0.001$. Significant interaction was detected between test phase and treatment: $F(2,53)=3.823$, $p<0.05$. Vehicle-injected control mice exhibited significantly higher freezing behaviour during CS+ presentation compared to baseline ($p<0.001$) and during CS-presentation compared to baseline ($p<0.05$), whereas there was no significant increase in freezing during CS-or CS+ presentation in muscimol treated mice. Post-hoc pair-wise Tukey's tests revealed significantly reduced freezing during CS+ presentation ($p<0.001$) and CS-presentation ($p<0.05$) in muscimol injected animals compared to vehicle.

Figure 1e right: Comparison of ΔF/F area in response to a foot-shock in L1 interneurons (n=45) and L2/3 neurons (n=184). Two-tailed, unpaired Student's t-test revealed significant difference ($p<0.001$).

Figure 2f right: Comparison of peak response to a nicotine puff in L1 interneurons in control (n=17) and after bath application of mecamylamine (100 µM) and methyllycaconitine (0.1 µM, n=8). Two-tailed, unpaired Student's t-test revealed significant difference ($p<0.01$).

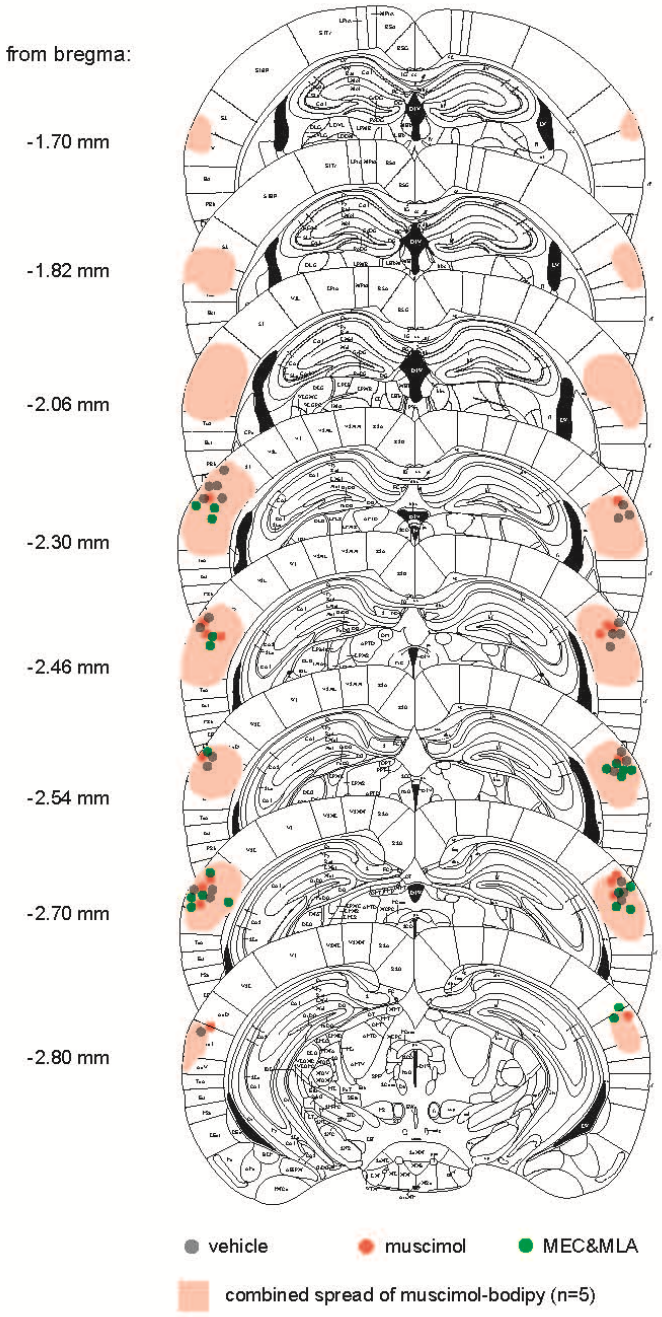
Figure 4a right: Comparison of voltage response area elicited by a foot-shock in L2/3 pyramidal neurons in control (n=6) and in the presence of mecamylamine (1 mM) and methyllycaconitine (0.1 mM, n=4). Two-tailed, unpaired Student's t-test revealed significant difference ($p<0.01$).

Figure 4c right: Comparison of the response (ΔF/F area) of L2/3 neurons (n=86) to a train of frequency-modulated sweeps, and to an identical train in which each sweep was paired with a foot-shock. Two-tailed, paired Student's t-test revealed significant difference ($p<0.001$).

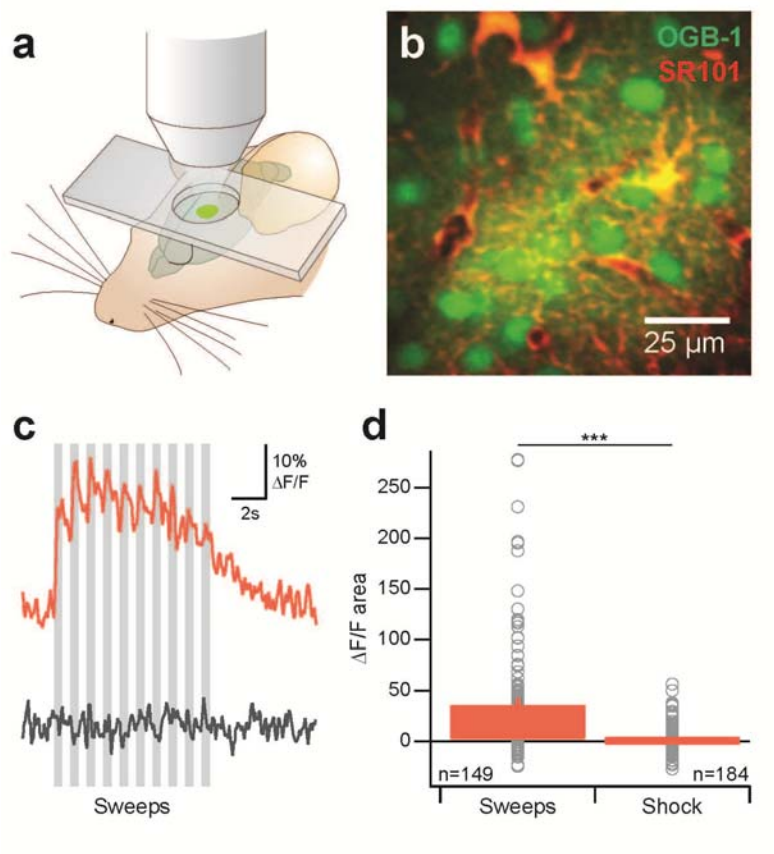
Figure 5b: Effects of mecamylamine and methyllycaconitine injection on acquisition of fear memory. Comparison of freezing during memory retrieval (24h after conditioning, drug-free) between a group of mice injected with drug during training (n=11) and a vehicle-injected control group (n=14). Two-way ANOVA indicated significant main effects between treatment groups: $F(1,74)=20.060$, $p<0.001$; as well as between test phases (baseline, CS-, CS+): $F(2,74)=32.591$, $p<0.001$. Significant interaction was detected between test phase and treatment: $F(2,74)=3.497$, $p<0.05$. Vehicle-injected control mice exhibited significantly higher freezing behaviour during CS+ presentation compared to baseline ($p<0.001$) and during CS-presentation compared to baseline ($p<0.001$). Drug-injected mice exhibited significantly higher freezing behaviour during CS+ presentation compared to baseline

($p<0.01$), but not during CS-presentation compared to baseline ($p>0.05$). Post-hoc pair-wise Tukey's tests revealed significantly reduced freezing during CS+ presentation ($p<0.001$) and CS-presentation ($p<0.01$) in drug injected animals compared to vehicle.

Figure 5f: Effects of optogenetic stimulation of PV+-interneurons during and after foot-shocks on acquisition of fear memory. Comparison of freezing during memory retrieval (24h after conditioning, no optogenetic manipulation) in three groups: optogenetics (n=8), reconditioning (n=8), sham (n=8). Two-way ANOVA indicated significant main effects between groups: $F(2,71)=12.348$, $p<0.001$; as well as between test phases (baseline, CS-, CS+): $F(2,71)=73.043$, $p<0.001$. Significant interaction was detected between test phase and treatment: $F(4,71)=4.409$, $p<0.05$. Post-hoc pairwise Tukey's tests revealed significantly reduced freezing during CS+ presentation in the optogenetics group compared to both sham and reconditioning ($p<0.001$), whereas mice in sham and reconditioning groups exhibited similarly high CS+-induced freezing ($p>0.05$). During CS-presentation there were no significant differences between optogenetics, sham and reconditioning groups.

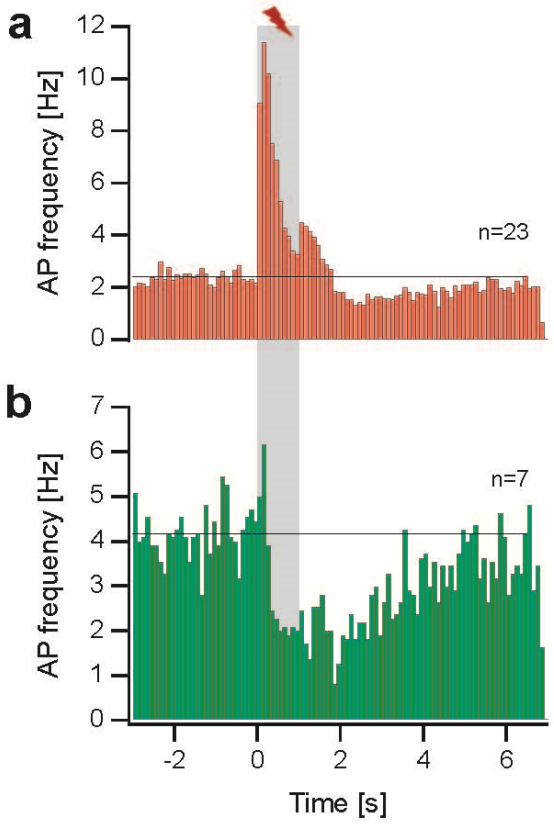


Supplementary Figure 1: Placement of injector tips and spread of muscimol
Coronal schematics of the mouse brain at the indicated distance posterior of bregma. Symbols indicate location of injector tips used for bilateral injection of vehicle (grey, n=14 mice), muscimol (red, n=9 mice) or mecamylamine and methyllycaconitine (green, n=11 mice) into auditory cortex before fear conditioning (**Fig. 1 a,b; Fig. 5 a,b**). Light red areas indicate the maximal spread of fluorescent muscimolbodipy (n=5 mice) injected after completion of memory retrieval to reveal the area of drug effect.



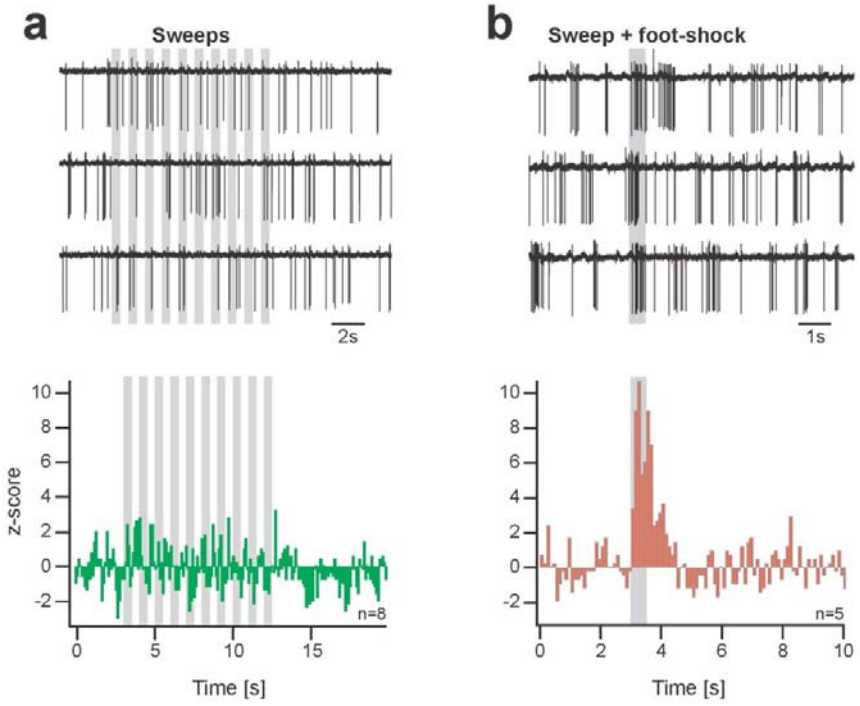
Supplementary Figure 2: Conditioned stimulus and foot-shock responses in auditory cortex L2/3 neurons

a, 2-photon calcium imaging in head-fixed, anesthetized mice. **b**, Time averaged frame showing L2/3 neurons (green) stained with the membrane permeant dye Oregon green BAPTA-1 AM (OGB-1). Glial cells are counterstained with sulforhodamine 101 (Ref. 1, red). **c**, Example responses of two L2/3 neurons to a train of frequency-modulated sweeps (averages of 3 presentations). **d**, Population data indicating that sweeps (n=149) elicit much greater activity than a foot-shock (n=184) in layer 2/3 neurons (single neurons grey, average red, $p<0.001$, two-tailed unpaired Student's t-test). Values presented as mean±s.e.m.

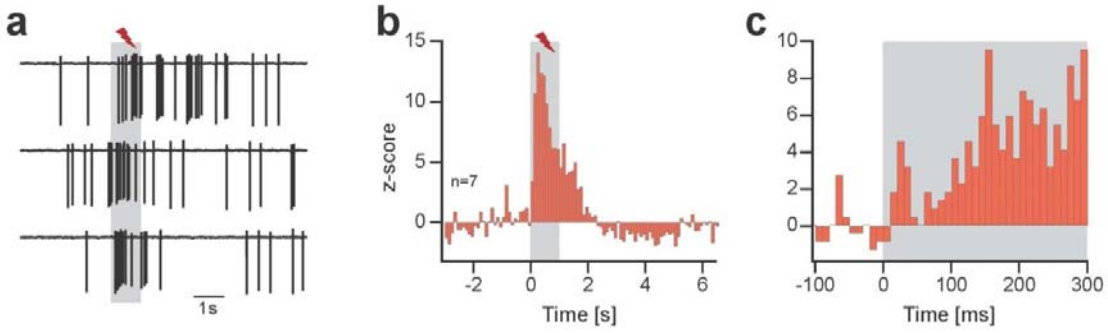


Supplementary Figure 3: Action potential frequency during foot-shocks in auditory cortex L1 interneurons

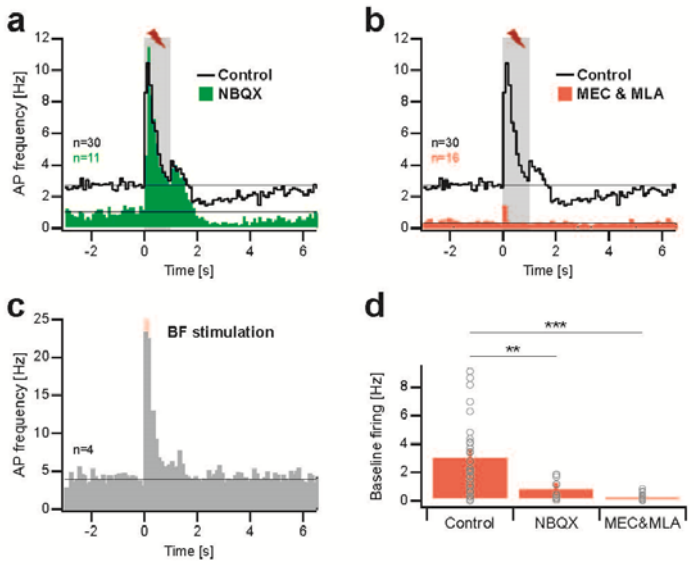
a, Population response to foot-shocks in L1 interneurons that were activated by foot-shocks (n=23). **b**, Population response to foot-shocks in L1 interneurons that were inhibited by foot-shocks (n=7). Lines indicate baseline firing frequency.



Supplementary Figure 4: Responses to tones and tone/foot-shock compounds in L1 interneurons
a, Top: Example traces from cell-attached recording of a L1 interneuron during presentation of frequency-modulated sweeps. Bottom: z-scored population response of L1 interneurons to sweeps (n=8). **b**, Top: Example traces from cell-attached recording of a L1 interneuron during presentation of a single sweep coincident with a foot-shock (both 500ms). Bottom: z-scored population response of L1 interneurons to sweep/foot-shock compounds (n=5).

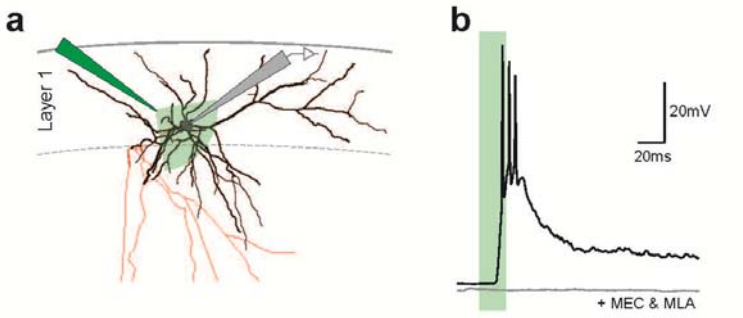


Supplementary Figure 5: Foot-shock responses in primary visual cortex L1 interneurons
a, Example traces from cell-attached recording of a L1 interneuron in primary visual cortex during foot-shocks. **b**, z-scored population response of L1 interneurons in visual cortex to foot-shocks (n=7). Note the similarity to foot-shock responses in auditory cortex (**Fig. 1g**). In addition, inhibitory responses were observed in two L1 interneurons in visual cortex (data not shown). **c**, The onset of the foot-shock response is biphasic, similar to L1 interneurons in auditory cortex (**Fig. 2a**, inset).



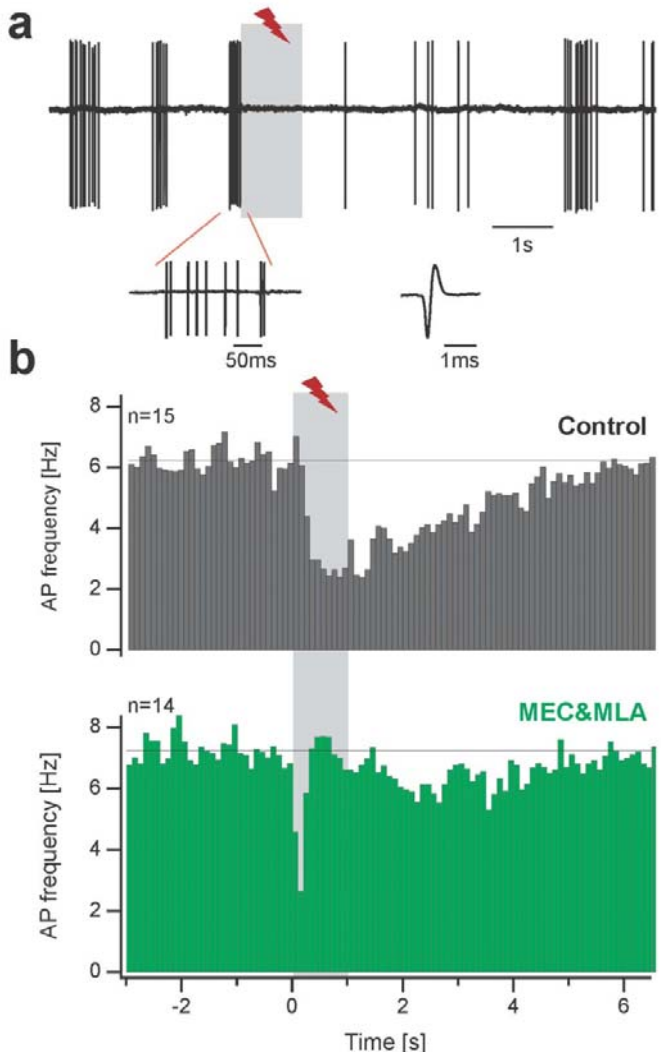
Supplementary Figure 6: L1 interneuron action potential frequency during foot-shocks under pharmacology, and during basal forebrain stimulation

a, Population response of L1 interneurons to foot-shocks in control (black, n=30; foot-shock excited and inhibited neurons were combined), and in the presence of the AMPA-receptor antagonist NBQX (1 mM, green, n=11). Note that NBQX decreases baseline action potential firing (line), but leaves the foot-shock response intact. **b**, Foot-shock responses in control (black, n=30), and in the presence of mecamylamine (1mM) and methyllycaconitine (0.1 mM, n=16). Block of nAChRs strongly affects both baseline firing (line) and foot-shock responses. **c**, Electrical microstimulation of the basal forebrain (red, 20 pulses of 0.2 ms duration delivered at 100 Hz) strongly activated L1 interneurons (n=4). **d**, Baseline firing frequency was reduced by both NBQX (n=11) and a combination of mecamylamine and methyllycaconitine (n=16) compared to control (n=30, average red, single neurons grey). One way ANOVA indicated significant differences between these groups ($F_{(2,56)}$=13.047, p<0.001). Post-hoc pairwise Tukey's tests revealed significant differences between control and NBQX (p<0.01) and between control and nAChR block (p<0.001). Values presented as mean±s.e.m.

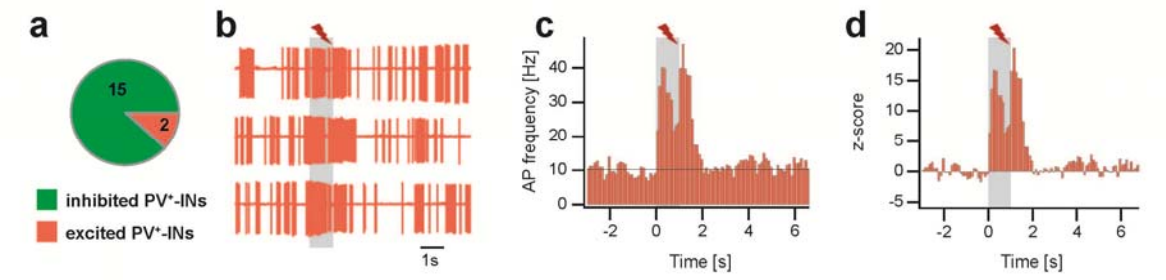


Supplementary Figure 7: Nicotine puffs can fire L1 interneurons

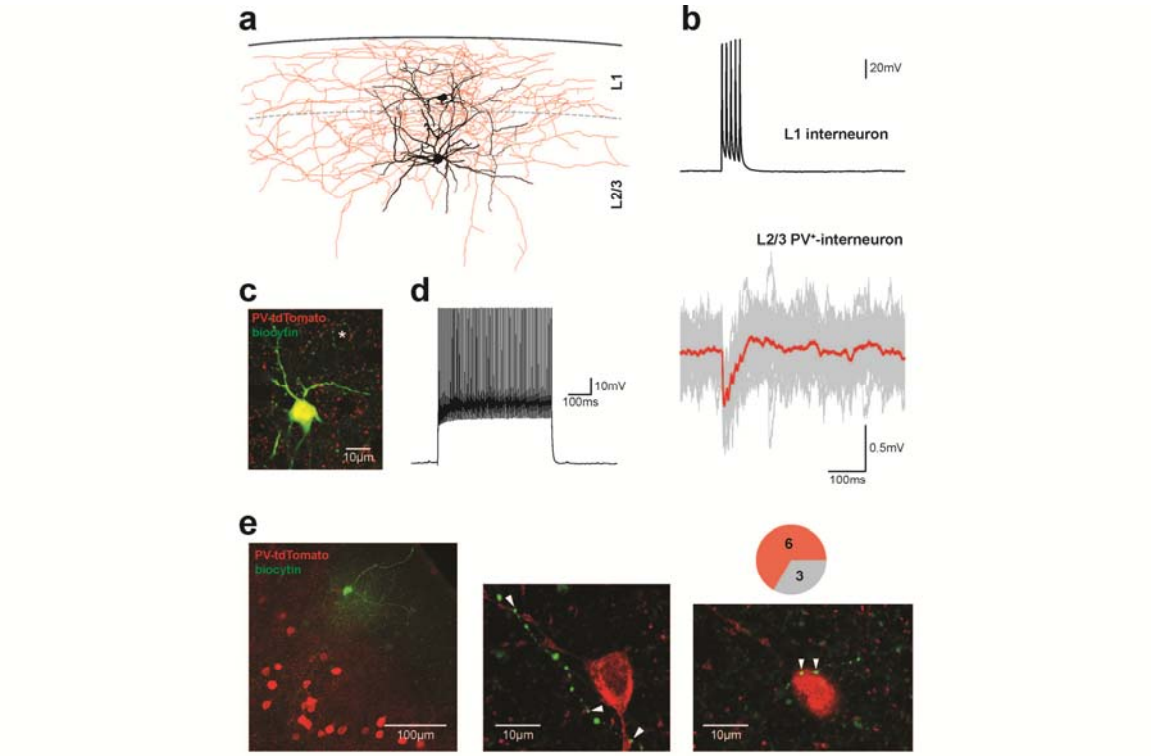
a, Whole-cell recording of a L1 interneuron (soma and dendrites black, axon red) in auditory cortex slice during puff application of nicotine (green, 100 µM). **b**, Example recording of a L1 interneuron which fired action potentials in response to the nicotine puff, and block by the nAChR antagonists mecamylamine (100 µM) and methyllycaconitine (0.1 µM) in grey.



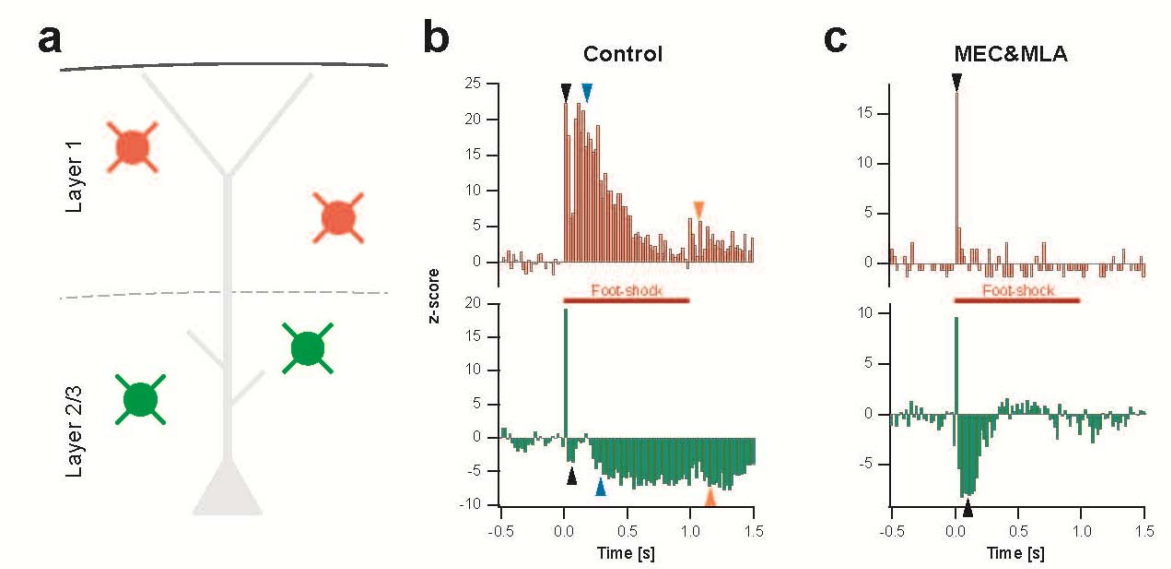
Supplementary Figure 8: Action potential frequency during foot-shocks in L2/3 PV+-interneurons
a, Top: Example cell-attached recording from a PV^+-interneuron during foot-shock. PV^+-interneurons typically fired high-frequency bursts (bottom left), and displayed fast action potentials with prominent afterhyperpolarization (right). **b,** Population response of PV^+-interneurons to foot-shocks in control (grey, n=15; 42% reduction of firing during the foot-shock), and in the presence of the nAChR antagonists mecamylamine (1 mM) and methyllycaconitine (0.1 mM, red, n=14; 11% reduction of firing during the foot-shocks). Note that nAChR block increased baseline firing rate (line), but strongly attenuated the foot-shock response.



Supplementary Figure 9: Excitatory foot-shock responses in a small subset of PV+-interneurons
a, PV$^+$-interneurons displaying excitatory responses to foot-shocks are a small minority. **b,** Example cell-attached recording from an excited PV$^+$-interneuron during foot-shocks. **c,** Population response of excited PV$^+$-interneurons to foot-shocks (n=2, line indicates baseline firing frequency). d, z-scored population response of excited PV$^+$-interneurons to foot-shocks (n=2).

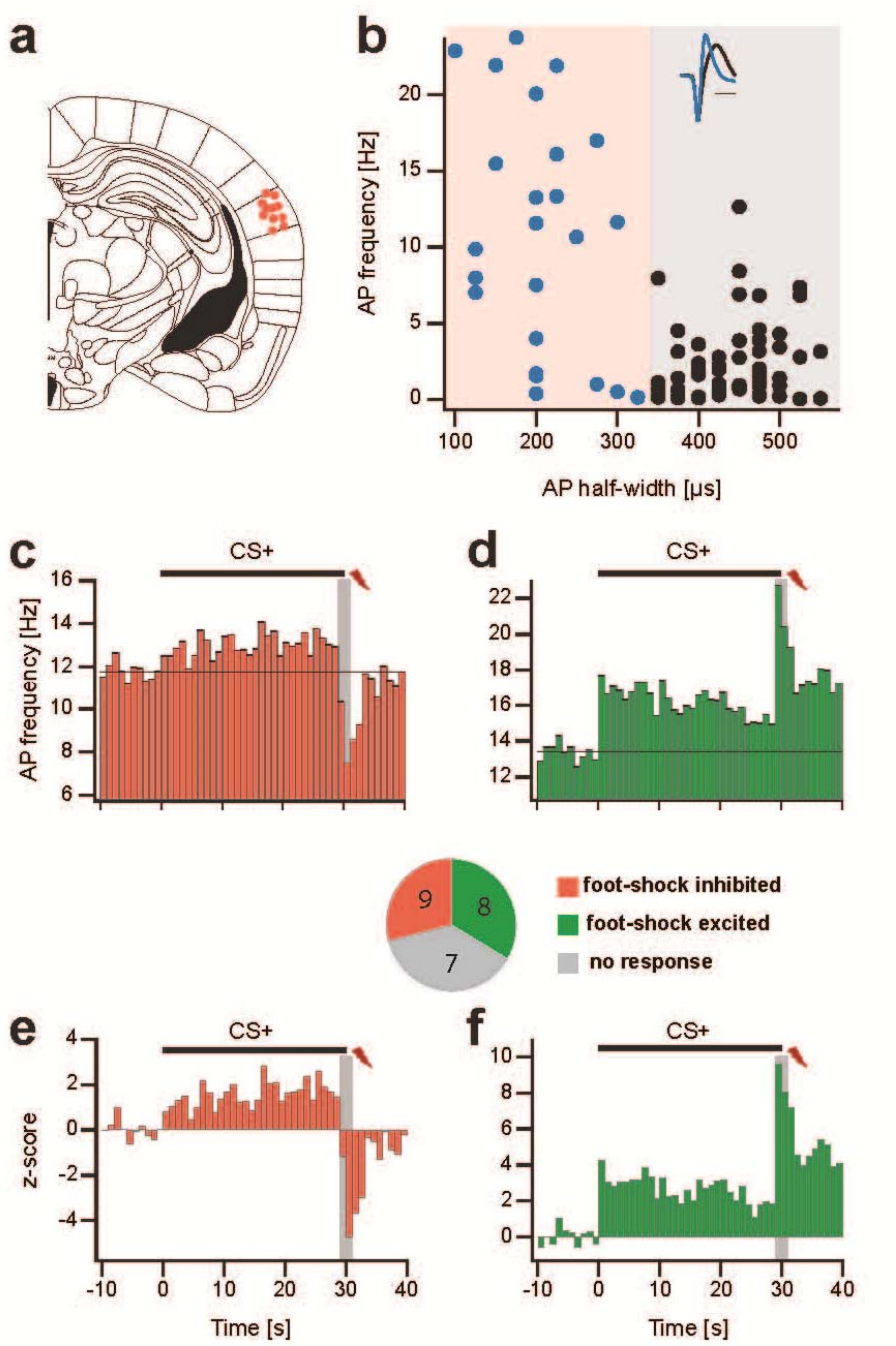


Supplementary Figure 10: Synaptic contacts between L1 interneurons and L2/3 PV$^+$-interneurons
a, Morphology of synaptically connected L1 interneuron and PV$^+$-interneuron (soma and dendrites black, axon red). **b**, A train of action potentials in the presynaptic L1 interneuron (top) elicited an inhibitory postsynaptic potential in the PV$^+$-interneuron held at -50 mV (bottom, single traces grey, average red). Two connections between L1 interneuron and PV$^+$-interneuron were observed in 9 pairs. **c**, The postsynaptic neuron was identified by expression of a fluorescent marker (tdTomato) after injection of conditional AAV into auditory cortex of PV-ires-Cre mice. The presence of synaptic baskets (asterisk marks centre of an example basket) identifies it as a basket cell. **d**, Firing pattern of postsynaptic neuron, consistent with fast-spiking phenotype. **e**, Left: Overview of a slice of auditory cortex with L1 interneuron filled with biocytin (green) and PV$^+$-interneurons identified by marker expression (red, tdTomato). Centre, right: Example sites of putative synaptic contacts (arrowheads) between the axon of the L1 interneuron (green) and proximal dendrites (centre) and the soma (right) of PV$^+$-interneurons (optical sections of 1 µm thickness). Inset: At least one putative synaptic contact between L1 interneuron and PV$^+$-interneurons was observed in 6 out of 9 experiments.

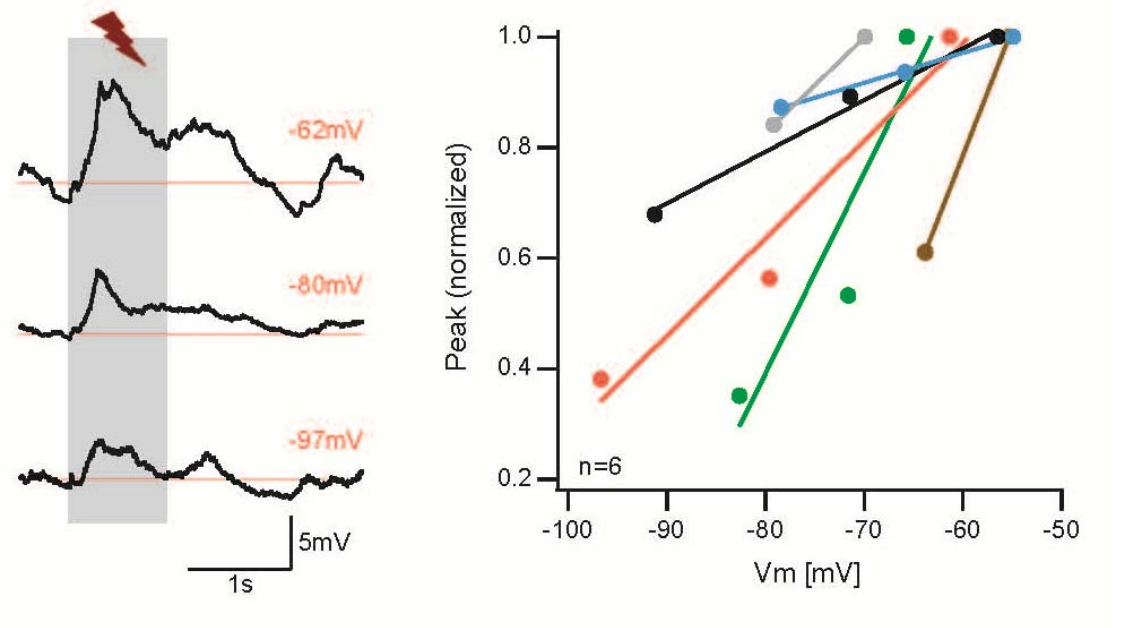


Supplementary Figure 11: Foot-shock response timecourse in L1 interneurons and L2/3 PV+-interneurons

a, Cytoarchitecture of upper layers of auditory cortex. **b,** z-scored population response to foot-shock at high temporal resolution (20 ms bins) in L1 interneurons (red) and L2/3 PV^+-interneurons (green) under control conditions. Both types of interneurons received short-latency excitation. In PV^+-interneurons, this was followed by inhibition (black arrowhead). L1 interneurons subsequently displayed a broad excitation (blue arrowhead), which was mirrored in PV^+-interneurons as broad inhibition. At foot-shock offset, L1 interneurons showed a second excitation (orange arrowhead), which was again reflected in PV^+-interneurons as inhibition. **c,** Foot-shock response in the two populations of interneurons during block of nAChRs by mecamylamine (1 mM) and methyllycaconitine (0.1 mM). Under these conditions, both types of interneurons received short-latency excitation, which was followed by short-latency inhibition of PV^+-interneurons (black arrow head). Together, these observations are consistent with the interpretation that L1 interneurons inhibit PV^+-interneurons during foot-shocks.

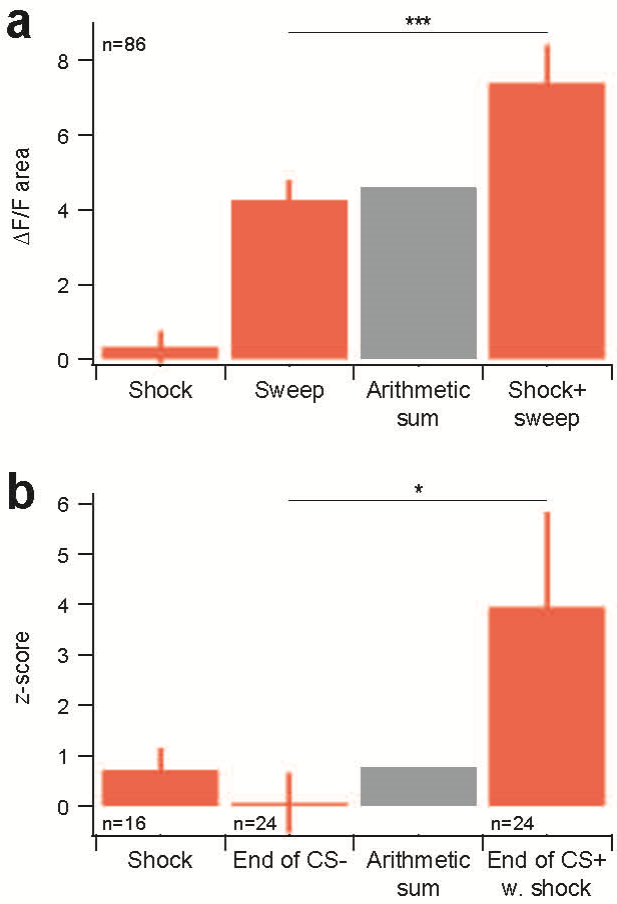


Supplementary Figure 12: Single-unit recordings of putative interneurons in freely behaving mice **a**, Location of recording sites in the superficial layers of auditory cortex (n=11). **b**, Putative interneurons (blue) were separated from putative pyramidal neurons (black) using unsupervised cluster analysis. Inset: Example traces (scale bar 500 µs). **c**, **d**, Population response of shock-inhibited (**c**) and shock-excited (**d**) putative interneurons to a train of frequency-modulated sweeps (CS+) with a shock at the end (line indicates baseline firing frequency). Inset: Putative interneurons displayed inhibition (n=9), excitation (n=8) or no response (n=7) to shocks. **e**, **f**, z-scored population response as in **c**, **d**.

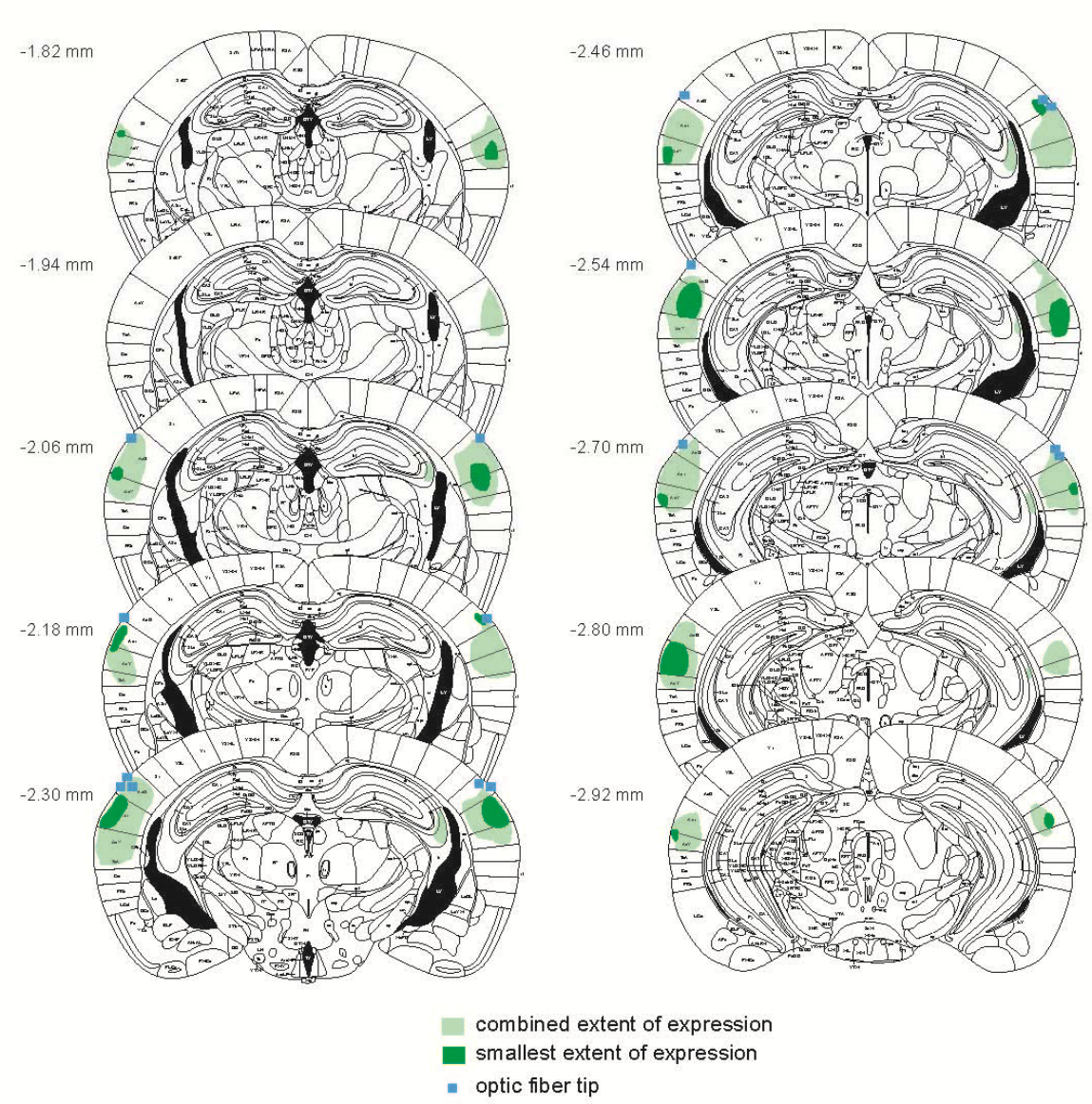


Supplementary Figure 13: Positive relationship between foot-shock response amplitude and membrane potential in pyramidal neurons

Left: Example whole-cell current-clamp recording of foot-shock responses in an auditory cortex pyramidal neuron at different membrane potentials. Right: Positive relationship between normalized response amplitude and membrane potential (n=6). This is consistent with disinhibition, because the effect of disinhibition increases with increasing distance from the $GABA_A$ reversal potential. However, in most of the pyramidal neurons recorded the reversal potential of the response is more hyperpolarized than expected for pure disinhibition (calculated ECl for the solutions used is -84 mV). This likely suggests that the foot-shock response is mediated by weak excitation along with strong disinhibition, since such compound responses can reverse at more hyperpolarized potentials than pure disinhibition.

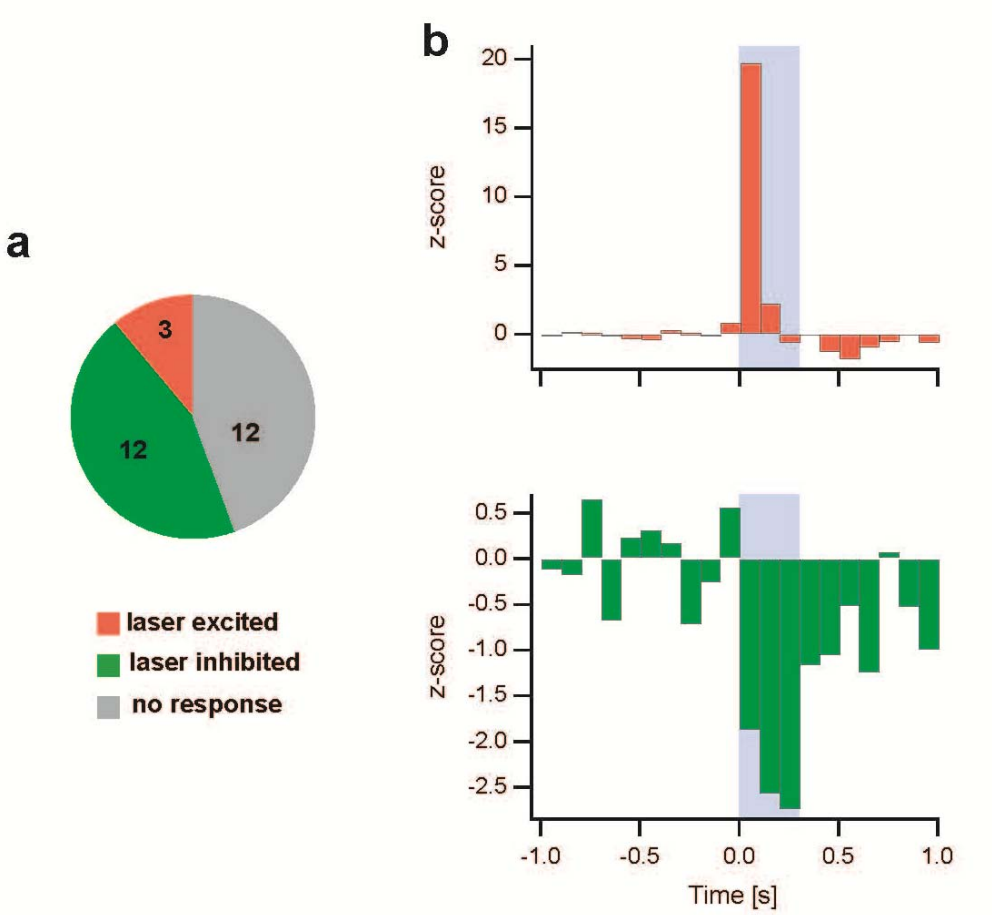


Supplementary Figure 14: Disinhibition of pyramidal neuron tone responses by aversive shocks
a, Comparison of average response area in 86 L2/3 neurons during calcium imaging in anesthetized mice to a single shock, a single sweep and a shock/sweep compound. Note strong boosting of the sweep response by the shock ($p<0.001$, two-tailed paired Student's t-test), exceeding the arithmetic sum of shock and sweep responses (grey). **b**, Comparison of average z-scores in putative pyramidal neurons during single-unit recordings in freely behaving mice in response to the shock alone (n=16), the end of the CS- without shock and the end of the CS+ with shock (n=24). Note strong boosting of the sweep response by the shock ($p<0.05$, two-tailed paired Student's t-test), exceeding the arithmetic sum of the responses to shock and the end of the CS- (grey). 6 of the neurons recorded during sweep and compound were also recorded during shock alone. Values presented as mean±s.e.m.



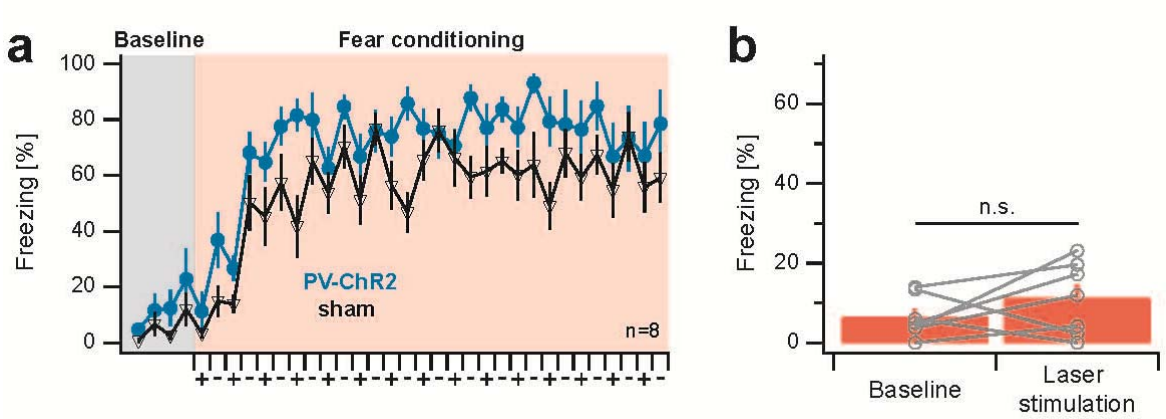
Supplementary Figure 15: Extent of ChR-2 expression in PV^{+}-interneurons in mice used for fear conditioning

Coronal schematics of the mouse brain at the indicated distance posterior of bregma. The combined extent from all animals (n=8; behaviour displayed in **Fig. 5f** in groups 'optogenetics' and 'reconditioning') is indicated in light green, the smallest extent from a single animal in dark green. Locations of optic fibre tips are shown in blue.



Supplementary Figure 16: Acute single-unit recordings from auditory cortex during optogenetic stimulation of PV$^+$-interneurons

a, Proportion of units that showed excitation (red), inhibition (green) and no response (grey) to stimulation of ChR-2 expressed in PV+-interneurons. **b,** Top: Three units displayed (presumably direct) excitation by a laser pulse (blue, 300ms). Bottom: Inhibition was the predominant response to laser stimulation (n=12), consistent with the interpretation that these neurons were inhibited by PV$^+$-interneurons during stimulation.



Supplementary Figure 17: Optogenetic intervention does not affect foot-shock processing, and is not perceived as a conditioned stimulus

a, Freezing during the fear conditioning session for data displayed in **Fig. 5f**. Both PV-ires-Cre mice injected with conditional AAV expressing ChR2 (n=8) and sham-injected littermates (n=8) were fear conditioned with laser stimulation during and after every foot-shock (Fig. 5e). During the baseline period (grey), mice from both groups displayed low freezing values. During the fear conditioning period (red), mice received alternating presentations of CS+ with foot-shock (+) and CS- alone (-). This caused similar freezing levels in both groups, suggesting that optogenetic stimulation did not affect foot-shock perception. **b,** After fear conditioning and memory retrieval, mice with ChR-2 expression in PV+-interneurons were subjected to optogenetic stimulation without auditory stimulus in a neutral context. This did not cause a significant enhancement in freezing (average red, single animals grey, $p>0.05$, two-tailed paired Student's t-test), suggesting that laser stimulation is not perceived as a conditioned stimulus associated with the foot-shock. Values presented as mean±s.e.m.

METHODS

Animals

Male C57BL6/J mice (1.5-6 months old, Harlan Ltd., Füllinsdorf, Switzerland and Janvier, France) were housed under a 12 h light/dark cycle, and provided with food and water ad libitum. Prior to fear conditioning, mice were individually housed for ≥ 7 days and habituated to handling ≥ 5 times. All animal procedures were executed in accordance with institutional guidelines and the prescribed authorities (Veterinary Department of the Canton of Basel-Stadt; Switzerland or Ministère de l'agriculture et de la forêt, 87-848, France and European Economic Community, EEC, 86-6091).

Head-fixed 2-photon imaging and targeted patch-clamp recordings

Mice were sedated with chlorprothixene (5 mg/kg, i.p. injection), kept warm on a hot-water bottle and after 10 minutes anesthetized with urethane (0.9 g/kg, i.p. injection). Core body temperature was maintained at 37.5ºC by a feed-back controlled heating pad (FHC, Bowdoinham, ME, USA). Mice were fixed in a stereotaxic frame (Kopf Instruments, Tujunga, USA), and local analgesia was provided by injection of ropivacain under the scalp (10 mg/ml, Naropin©, AstraZeneca, Switzerland). After retraction of the scalp, the location of the area of interest was determined as follows: auditory cortex: 2.46 mm posterior of bregma, 4.5 to 4.7 mm lateral of midline; primary visual cortex: 3.16 mm posterior of bregma, 2.5 mm lateral of midline. A custom-built head-plate was glued to the skull with cyanoacrylate glue (Ultra Gel©, Henkel, Düsseldorf, Germany) and dental cement (Paladur©, Heraeus, Hanau, Germany), and a small craniotomy was performed (2-3 mm diameter). The dura was retracted, the exposed cortical surface was superfused with normal rat ringer (NRR) containing (in mM): 135 NaCl, 5.4 KCl, 5 HEPES, 1.8 $CaCl_2$, pH 7.2 with NaOH), covered with agarose (1% type-III-A in NRR) and fixed with a cover-slip. For 2-photon calcium imaging, the membrane permeant dye Oregon Green BAPTA-1 acetoxymethyl ester (OGB-1, 1mM in NRR with 8% DMSO and 2% pluronic F-127, Invitrogen, Basel, Switzerland) was pressure ejected from a patch pipette under 2-photon visualization. After 1 hour, this lead to intracellular staining of virtually all cells in a diameter of approximately 300 µm (Stosiek et al. 2003). Localization of the labelled area in auditory cortex was verified post hoc in slices. Sulforhodamine 101 (0.2 mM in NRR) was bath-applied for 5 minutes to the exposed cortex to counterstain glial cells (Nimmerjahn et al. 2004). 2-photon calcium imaging was performed using a scanning microscope (Fluoview 1000MPE, Olympus, Volketswil, Switzerland) coupled to a femtosecond pulse infrared laser at 830 nm wavelength (MaiTai HP, Spectra-Physics). The beam was adjusted with a telescope to fill the back-focal plane of the 20x water-immersion objective (0.95 NA, Olympus, Volketswil, Switzerland). Average power in the back-focal plane was <120 mW. Frames were acquired (Fluoview software, Olympus, Volketswil, Switzerland) at 15 Hz at depths below the pia ranging from 20 to 400

µm. Time stacks were processed offline using ImageJ to extract ΔF/F for each neuron. Foot-shocks (1s, DC, 0.4 to 2 mA, 10-20 s interval) were delivered to the hindpaws. Frequency-modulated sweeps (logarithmically modulated between 5 and 15 kHz, 50 ms rise and fall) were generated (System3, Tucker Davis Technologies) and presented at 70 dB sound pressure level from an electrostatic speaker (ES1, TDT) mounted in front of the animal.

2-photon guided patch-clamp recordings were performed using patch pipettes of 4-10 MΩ resistance filled with NRR for cell-attached recordings, for whole cell current-clamp recordings with intracellular solution containing (in mM): 130 K-methanesulfonate, 6.3 KCl, 20 Na_2-phosphocreatine, 4 Mg-ATP, 0.3 Na-GTP, 10 HEPES, 290 mOsm, pH 7.3 with KOH and for whole cell voltage clamp recordings with intracellular solution containing (in mM): 125 Cs-methanesulfonate, 2 CsCl, 10 Na_2-phosphocreatine, 4 Mg-ATP, 0.3 Na-GTP, 5 TEACl, 3.5 QX-314, 0.5 EGTA, 10 HEPES, 290 mOsm, pH 7.3 with CsOH. Pipette solutions were supplemented with Alexa 488 or Alexa 594 (25-50µM, Invitrogen, Basel, Switzerland). Neurons of interest were identified by either marker expression (L1 interneurons in GAD67-GFP mice, PV^+-interneurons in PV-ires-Cre mice (Hippenmeyer et al. 2005) injected with AAV conditionally expressing venus), or by imaging the 'shadows' created by neurons in dye-filled extracellular space in wildtype mice (L1 interneurons, L2/3 pyramidal neurons). L1 interneurons were located between 30 and 70 µm below the pia. Cell-attached and whole-cell recordings were established using a 40x water-immersion objective (0.8 NA, Olympus, Volketswil, Switzerland) according to standard procedures (Margrie et al. 2003). Cell-attached recordings were performed in current-clamp or voltage-clamp configuration, and rejected when no action potential occurred during the entire length of the recording (≥ 5 minutes). Whole-cell current-clamp recordings were rejected when the series resistance exceeded 70 MΩ. Whole-cell voltage-clamp recordings were performed at 0-20 mV holding potential and rejected when the series resistance exceeded 50 MΩ. IPSCs were detected using a derivative threshold of 40 pA/ms. Pyramidal neurons were identified after dye filling by pyramidal morphology and spiny dendrites. Signals were recorded (Axopatch 200B, Molecular Devices or ELC-03XS, NPI), low-pass filtered at 20 kHz and digitized at 50 kHz (Digidata 1322A, Molecular Devices) using pClamp software (Molecular Devices). For cell-attached recordings, action potentials were detected based on amplitude. Peri-stimulus time histograms were computed across the entire population of recordings of a given type. Microstimulation of the basal forebrain (20 pulses of 0.2 ms duration delivered at 100 Hz) was performed using bipolar stimulating electrodes at the following coordinates: 0.8 mm posterior of bregma, 1.75 mm lateral of midline, 4.3 mm below cortical surface. Drugs were present in the bath solution throughout the experiment. Unless stated otherwise, chemicals were obtained from Sigma (Buchs, Switzerland).

Surgery and local drug injection

Mice were anesthetized with isoflurane (induction: 5%, maintenance: 1.5%) in oxygen-enriched air (Oxymat 3©, Weinmann, Hamburg, Germany) and fixed in a stereotaxic frame (Kopf Instruments, Tujunga, USA). Core body temperature was maintained at 36.5º C by a feed-back controlled heating pad (FHC, Bowdoinham, ME, USA). Analgesia was provided by local injection of ropivacain under the scalp (Naropin©, AstraZeneca, Switzerland) and systemic injection of meloxicam (100µl of 5mg/ml, i.p., Metacam©, Boehringer-Ingelheim, Ingelheim, Germany). Guide cannulae (26 gauge, with dummy screw caps, Plastics One, Roanoke, USA) were implanted bilaterally to inject at the following coordinates: 2.46 mm posterior of bregma, ±4.5 mm lateral of midline, 0.6 mm below cortical surface. Implants were fixed to the skull with cyanoacrylate glue (Ultra Gel©, Henkel, Düsseldorf, Germany) and dental cement (Paladur©, Heraeus, Hanau, Germany). Mice were then given one week to recover from surgery, during which time they were handled ≥ 5 times to habituate them to the injection procedure. Fifteen minutes before fear conditioning, 32 gauge stainless steel injectors attached to 25µl Hamilton syringes were inserted into the guide cannulae and an injection volume of 0.2 µl per hemisphere was delivered within 40 s using a microinfusion pump (Stoelting, Wood Dale, USA). Drug animals received bilateral injections of muscimol (100 ng per hemisphere) or a combination of mecamylamine (10 µg) and methyllycaconitine (10 µg) dissolved in NRR containing 1% fast green (Serva, Heidelberg, Germany). Control mice were injected with vehicle solution only. In a sub-set of mice, fluorescent muscimol bodipy (625 µM in NRR with 5% DMSO) was injected after fear memory retrieval to quantify spread of the drug. After completion of the experiment, mice were transcardially perfused with 4% paraformaldehyde in phosphate-buffered saline (PFA), their brains extracted and post-fixed in PFA overnight. For histological verification of the injection site, 60 µm coronal brain sections were made on a vibratome (Leica Microsystems, Heerbrugg, Switzerland) and imaged on a stereoscope (Leica Microsystems, Heerbrugg, Switzerland).

Virus injection and optogenetics

For labelling of PV^+-interneurons (**Fig. 3 a,b**), an adeno-associated virus (AAV, serotype 2/7 (venus) or 2/9 (tdTomato), Vector Core, University of Pennsylvania, Philadelphia, PA, USA) was employed to deliver a DNA construct for conditional, Cre-dependent expression of venus or tdTomato. AAV (approximately 0.5 µl per hemisphere) was injected from glass pipettes (tip diameter 10-20 µm) connected to a picospritzer (Parker Hannifin Corporation, Fairfield, NJ, USA) into auditory cortex of PV-ires-Cre mice (Hippenmeyer et al. 2005) at the following coordinates: 2.46 mm posterior of bregma, 4.5 mm lateral of midline, 0 to 1.1 mm below cortical surface. Experiments were performed after 0.5-3 months of expression time. For optical control of PV^+-interneurons (**Fig. 5 c-f**), the same methodology was used for bilateral injection of a conditional AAV coexpressing channelrhodopsin-2 (ChR-2) (Zhang et al. 2007) and venus (AAV2/7 EF1a::DIO-ChR2(H134R)-2A-NpHR-2A-Venus (Tang et al. 2009))

into auditory cortex of PV-ires-Cre mice. These animals were additionally implanted with custom-built connectors holding optic fibres (0.48 NA, 200 µm diameter, Thorlabs). Fibre ends were inserted 200 µm into neocortex at the injection site. Implants were fixed to the skull with cyanoacrylate glue (Ultra Gel©, Henkel, Düsseldorf, Germany) and dental cement (Paladur©, Heraeus, Hanau, Germany). After 2-4 months of expression time and habituation to handling, both implanted connectors were linked to a custom-built laser bench via optic fibres suspended over the conditioning context. This arrangement allowed the animals to move freely in the context. Stimulation of ChR-2 (< 20 mW per implanted fibre) was delivered during and for 5s after each foot-shock (**Fig. 5e**) using a custom-built laser bench (laser: MBL473, 473 nm wavelength, CNI Lasers, Changchun, China). One day after conditioning, mice were exposed to CS^+ and CS^- without optogenetic intervention in a neutral context. The optogenetic approach was previously validated *in vitro (Tang et al. 2009)* and *in vivo (Ciocchi et al. 2010)*, and here by acute extracellular recordings from auditory cortex of infected animals (**Supplementary Fig. 16**). Custom-built optrodes (optic fibre with 16-wire electrode attached) were lowered into the infected area. Laser pulses (300 ms) were delivered and recordings performed as detailed in section 'Extracellular recordings in freely-behaving mice'. After completion of the experiment, mice were transcardially perfused with 4% PFA. Brains were post-fixed in PFA overnight at 4°C, and cut into 80 µm thick coronal slices on a vibratome (Leica Microsystems, Heerbrugg, Switzerland). To improve the fluorescent signal, an immunostaining was performed. Slices were kept in blocking solution (3% BSA, 0.2% Triton in 0.1 M PBS) for 1 h at room temperature, before application of the primary antibody (Goat anti-GFP, Abcam plc, Cambridge, UK; 1:500 in blocking solution) and incubated at 4°C overnight. After washing, slices were incubated with secondary antibody (Alexa Fluor 488, donkey anti goat, Invitrogen, Basel, Switzerland; 1:1000 in PBS with 3% BSA) at 4°C over night. After a final wash, slices were mounted on cover slips and imaged.

Behavior

Fear conditioning and fear retrieval took place in two different contexts (context A and B). The conditioning and test boxes and the floor were cleaned before and after each session with 70% ethanol or 1% acetic acid, respectively. To score freezing behavior, an automatic infrared beam detection system placed on the bottom of the experimental chambers (Coulbourn Instruments, Whitehall, PA, USA) was used. Mice were considered to be freezing if no movement was detected for 2 s and the measure was expressed as a percentage of time spent freezing. To ensure that our automatic system scores freezing rather than just immobility, we previously compared the values obtained with those measured using a classical time-sampling procedure during which an experimenter blind to the experimental conditions determined the mice to be freezing or not freezing every 2 s (defined as the complete absence of movement except for respiratory movements). The values obtained

were 95% identical and the automatic detection system was therefore used throughout the experimental sessions. Conditioned stimuli for differential fear conditioning were 30 s trains of frequency-modulated sweeps (500 ms duration, logarithmically modulated between 5 and 15 kHz, 50 ms rise and fall) delivered at 1 Hz at a sound pressure level of 70 dB. The CS^+ (up-sweep) was paired with a foot-shock (1 s, 0.6 mA, 15 CS^+/foot-shock pairings; inter-trial interval: 20–180 s). The onset of the foot-shock coincided with the onset of the last sweep in the CS^+. The CS^- (down-sweep) was presented after each CS+/foot-shock association, but was never reinforced (15 CS^- presentations, inter-trial interval: 20–180 s). On the next day, conditioned mice were submitted to fear retrieval in context B, during which they received 4 and 4 presentations of the CS^- and the CS^+, respectively.

Extracellular recordings in freely-behaving mice

Surgical procedures are described above. Mice were secured in a stereotaxic frame and implanted with a pair of insulated silver wires (170 µm diameter) beneath the skin of each eyelid for delivery of periorbital shocks. In addition, mice were unilaterally implanted in auditory cortex with a multi-wire electrode aimed at the following coordinates: 2.46mm posterior of bregma, 4.5mm lateral to midline, and 0.6mm to 0.85mm below the cortical surface. Electrodes consisted of 16 individually insulated, gold-plated nichrome wires (13 µm inner diameter, impedance 30 to 100 kΩ, Sandvik) contained in a 26 gauge stainless steel guide cannula. The wires were attached to a connector (18 pin, Omnetics). The implant was secured using cyanoacrylate adhesive gel. After surgery mice were allowed to recover for 7 days. Analgesia was applied before, and during the 3 days after surgery (Metacam©, Boehringer-Ingelheim, Ingelheim, Germany). Electrodes were connected to a headstage (Plexon) containing sixteen unity-gain operational amplifiers. The headstage was connected to a 16-channel computer-controlled preamplifier (gain 100X, bandpass filter from 150 Hz to 9 kHz, Plexon). Neuronal activity was digitized at 40 kHz, bandpass filtered from 250 Hz to 8 kHz, and isolated by time–amplitude window discrimination and template matching using a multichannel acquisition processor system (Plexon). Single-unit spike sorting was performed using off-line spike sorter (OFSS, Plexon). Principal component scores were calculated for unsorted waveforms and plotted on three-dimensional principal component spaces, and clusters containing similar valid waveforms were manually defined. A group of waveforms was considered to originate from a single neuron if it defined a discrete cluster in principal component space that was distinct from clusters for other units, and if it displayed a clear refractory period (1 ms) in auto-correlograms. Template waveforms were then calculated for well-separated clusters and stored for further analysis. To avoid analysis of the same neuron recorded on different channels, we computed cross-correlation histograms. If a target neuron displayed a peak of activity at a time that the reference neuron fired, only one of the two neurons was considered for further analysis. To separate putative inhibitory interneurons from putative excitatory pyramidal neurons we used an unsupervised cluster

algorithm based on the Ward's method. Briefly, the Euclidian distance was calculated between all cell pairs based on the two-dimensional space defined by each cell's average spike width (measured from trough to peak) and baseline firing rate. An iterative agglomerative procedure was then used to combine cells into groups based on the matrix of distances such that the total number of groups was reduced to give the smallest possible increase in the within group sum of square deviation. Mice were presented with the same conditioned stimuli used for fear conditioning (see above). The CS^+ was paired with a periorbital shock (2 s, 2.5 mA, 15 CS^+/shock pairings, inter-trial interval 20-180 s), which was employed instead of foot-shocks to minimize electrical artefacts in the recording. The onset of the periorbital shock coincided with the onset of the last sweep. The CS^- was presented after each CS^+ (15 CS^- presentations, inter-trial interval: 20–180 s). CS^+ and CS^- were counterbalanced across animals. At the conclusion of the experiment, recording sites were marked with electrolytic lesions before perfusion, and electrode locations were reconstructed with standard histological techniques

***In vitro* electrophysiology**

300 µm thick coronal sections of auditory cortex were prepared from mice (6-8 weeks old) in ice-cold slicing artificial cerebro-spinal fluid (ACSF) containing (in mM): 124 NaCl, 1.25 NaH_2PO_4, 10 $MgSO_4$, 2.7 KCl, 26 $NaHCO_3$, 2 $CaCl_2$, 10 Glucose, 4 ascorbate (95% O_2/5% CO_2). Slices were incubated for 45 minutes at 37°C in an interface chamber, and then allowed to cool to room temperature. Recordings were performed at 34°C under infrared videomicroscopy in ACSF containing (in mM): 124 NaCl, 1.25 NaH_2PO_4, 1.3 $MgSO_4$, 2.7 KCl, 26 $NaHCO_3$, 2 $CaCl_2$, 10 Glucose, 4 ascorbate (95% O_2/5% CO_2). Whole-cell current clamp recordings were performed with patch pipettes (4-10 MΩ resistance) filled with intracellular solution containing (in mM): 130 methanesulfonate, 6.3 KCl, 20 Na_2-phosphocreatine, 0.3 Na-GTP, 4 Mg-ATP, 10 HEPES, 5 biocytin, 290 mOsm, pH 7.3 with KOH. Signals were recorded (Multiclamp 700B, Molecular Devices, Sunnyvale, CA, USA), low-pass filtered at 10 kHz and digitized at 20 kHz (Digidata 1322A, Molecular Devices) using pClamp9 software (Molecular Devices). Signals were analyzed using Igor Pro (Wavemetrics, Portland, OR, USA). The border between L1 and L2/3 was visually identified by the abrupt change in cell number, and only L1 interneurons at least 15 µm away from the border region were chosen for recordings. Local puff application of nicotine (100 µM, 20-80 ms duration) was delivered from a patch pipette connected to a picospritzer (Parker Hannifin Corporation, Fairfield, NJ, USA). For paired recordings between L1 interneurons and L2/3 PV^+-interneurons, PV-ires-Cre mice were injected with an AAV (Vector Core, University of Pennsylvania, Philadelphia, PA, USA) leading to Cre-dependent expression of tdTomato. Fluorescent PV^+-interneurons were identified using TillvisION (Till Photonics, Gräfelfing, Germany) and double recordings were performed as described above. Unless stated otherwise, drugs were bath-applied. To reveal morphology, slices were fixed in 4% PFA (4°C overnight), washed and stained for

biocytin (4 days at 4°C in PBS, 0.5-1% triton-x, 0.2% Alexa (488, 568 or 680) streptavidin conjugate, Invitrogen, Basel, Switzerland). To identify putative synaptic contacts, this was combined with immunostaining against venus (see above) or tdTomato (Rabbit anti-RFP, MBL, Nunningen, Switzerland; 1:500 in blocking solution; Alexa Fluor 568, donkey anti-rabbit, Invitrogen, Basel, Switzerland; 1:1000 in PBS with 3% BSA). Slices were mounted and confocal images acquired using an LSM700 confocal laser scanning microscope (Zeiss, Göttingen, Germany). Stacks were analysed using ImageJ (Rasband, W.S, NIH, Bethesda, Maryland, USA) and single representative cells were reconstructed using Neurolucida software (MBF Bioscience, Williston, VT, USA). Putative synaptic contacts were scored when the axon of a L1 interneuron was located within 1 µm from a PV^+ dendrite or soma.

3.2 Publication II

Distinct subtypes of amygdala interneurons control fear learning through stimulus-specific disinhibition

Steffen B.E. Wolff[1,2], Jan Gründemann[1,*], Philip Tovote[1,*], Gilad A. Jacobson[1], Christian Müller[1], Cyril Herry[3], Ingrid Ehrlich[4], Rainer W. Friedrich[1], Johannes J. Letzkus[1,∞,#], Andreas Lüthi[1,∞,#]

*[1]Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland [2]University of Basel, Switzerland. [3]INSERM U862, Neurocentre Magendie, 146 rue Leo Saignat, 33077 Bordeaux, France. [4]Hertie Institute for Clinical Brain Research, 72076 Tübingen, Germany. *,∞These authors contributed equally to this work. #To whom correspondence should be addressed. E-mail: johannes.letzkus@fmi.ch (J.J.L.); andreas.luthi@fmi.ch (A.L.).*

Nature (2014), accepted for publication

Learning is mediated by experience-dependent plasticity in neuronal circuits. Activity in neuronal circuits is tightly regulated by different subtypes of inhibitory interneurons, yet their role in learning is poorly understood. Using a combination of in vivo single unit recordings and optogenetic manipulations, we show that in the basolateral amygdala (BLA), parvalbumin (PV) and somatostatin (SOM) expressing interneurons bidirectionally control the acquisition of fear conditioning, a simple form of associative learning, through two distinct disinhibitory mechanisms. During the auditory cue, PV+ interneurons are excited and indirectly disinhibit the dendrites of BLA principal neurons via SOM+ interneurons, thereby enhancing auditory responses and promoting cue-shock associations. During the aversive footshock, however, both PV+ and SOM+ interneurons are inhibited, which boosts postsynaptic footshock responses and gates learning. These results demonstrate that associative learning is dynamically regulated by the stimulus-specific activation of distinct disinhibitory microcircuits through precise interactions between different subtypes of local interneurons.

INTRODUCTION

Fear conditioning is a powerful model system for investigating plasticity of neuronal circuits and the mechanisms of associative learning as the basic underlying circuitry has been thoroughly studied over the last decades (Davis 2000; LeDoux 2000; Maren and Quirk 2004; Fanselow and Poulos 2005). The basolateral amygdala (BLA) has been identified as a key brain area involved in auditory fear conditioning, where associative synaptic plasticity at glutamatergic sensory afferents is induced by the pairing of a neutral tone (conditioned stimulus (CS)) and a mild aversive footshock (unconditioned stimulus (US)) during conditioning (LeDoux 2000; Maren and Quirk 2004; Pape and Pare 2010). As a cortex-like structure (Waclaw et al. 2010), the BLA contains a variety of interneurons (INs) with marker expression, basic properties and connectivity very similar to neocortical circuits (Freund and Buzsaki 1996; Sah et al. 2003; Markram et al. 2004; Somogyi and Klausberger 2005; Ehrlich et al. 2009; Pape and Pare 2010; Fishell and Rudy 2011; Spampanato et al. 2011), and inhibition has been shown to play a role in fear conditioning (Ehrlich et al. 2009). Functionally, salient sensory stimulation leads to cell type specific responses in different IN subtypes in anaesthetized animals (Bienvenu et al. 2012). Moreover, fear conditioning and extinction have been shown to induce opposing changes in the strength of inhibitory transmission and in the expression of inhibition-related genes (Davis 1979; Harris and Westbrook 1998; Chhatwal et al. 2005; Heldt and Ressler 2007; Ehrlich et al. 2009). However, besides the general notion that inhibition is involved in fear learning, knowledge about the role of individual IN subtypes in fear learning is still lacking.

One of the major IN subclasses in cortical and cortex-like structures such as the BLA is defined by the expression of the calcium-binding protein parvalbumin (PV) (Kawaguchi and Kubota 1997; Markram et al. 2004; Somogyi and Klausberger 2005; Rudy et al. 2011; Atallah et al. 2012). Most PV^+ INs feature a fast-spiking phenotype with narrow spike widths and high firing rates (Rainnie et al. 2006; Woodruff and Sah 2007b). PV^+ INs preferentially synapse onto the perisomatic region of their target cells, thereby controlling neuronal activity and spike output (McDonald and Betette 2001; Somogyi and Klausberger 2005; Muller et al. 2006; Atallah et al. 2012). In addition to glutamatergic principal neurons (PNs), PV+ cells contact both other PV^+ INs and different IN types (Muller et al. 2005; Woodruff and Sah 2007b). In the hippocampus, PV^+ cells have been shown to interact with INs expressing somatostatin (SOM^+) to regulate pyramidal cell output (Lovett-Barron et al. 2012). In contrast to PV^+ cells, SOM^+ INs preferentially contact the distal dendrites (Muller et al. 2007a), which allows them to efficiently control the impact of inputs to their target cells (Gentet et al. 2012; Chiu et al. 2013). Here, we address the role of interneuron subtypes using a combination of optogenetic manipulations and single unit recordings from identified BLA INs during fear learning. Our results reveal that BLA PV^+ and SOM^+ INs exert bidirectional control over fear acquisition resulting in differential changes in inhibition along the somatodendritic axis of PNs. We identify inhibition of SOM^+ INs by PV^+ INs as a crucial mechanism for dendritic disinhibition of principal neurons. Our findings indicate that

exquisitely organized inhibitory and disinhibitory microcircuits comprised of defined subtypes of INs play important and distinct roles during associative learning.

RESULTS

PV$^+$ INs in the BLA control fear learning

To examine a possible role for PV$^+$ cells in fear learning, we injected recombinant adeno-associated viral vectors (rAAVs) for conditional expression of the excitatory opsin channelrhodopsin-2 (ChR2 (Zhang et al. 2007)) into the BLA of mice expressing CRE recombinase under the control of the PV promoter (Hippenmeyer et al. 2005), resulting in selective expression of ChR2 in PV$^+$ INs (**Fig. 1a-c**). For optical stimulation of the transduced cells, optic fibres aimed at the BLA were implanted bilaterally (**Fig. 1a; Extended Data Fig. 1a**).

Animals were conditioned to two distinct CSs (tones of different frequencies, counterbalanced between experiments). While PV$^+$ cell activity was manipulated during one entire CS-US pairing, the other CS-US pair served as a within-animal control (**Fig. 1d, left**). A fear memory retrieval test on the next day without optogenetic stimulation revealed that PV$^+$ IN activation during acquisition significantly attenuated the CS-induced freezing response when compared with the internal control CS (**Fig. 1d, middle; Ext. Data Table 1**). These results suggest that the level of PV$^+$ IN activity in the BLA can control the acquisition of conditioned fear responses.

Footshock-induced inhibition of PV$^+$ cells gates fear learning

Next, we asked whether optogenetic activation of PV$^+$ INs specifically during the US in the training phase would be sufficient to interfere with fear conditioning (**Fig. 2a**). When we tested fear memory retrieval the next day without optogenetic stimulation, we observed that learning was impaired to a similar extent as observed with PV$^+$ IN activation during the entire CS-US pairing (**Fig. 1d**). This demonstrates that PV$^+$ IN activity during the US is a crucial factor regulating fear learning. To address whether endogenous activity of PV$^+$ cells is necessary for fear learning, we used conditional expression of the inhibitory opsin ARCH (Chow et al. 2010) (**Fig. 2a**). In contrast to ChR2 manipulations, ARCH-mediated inhibition of PV$^+$ INs during the US resulted in increased freezing levels during fear memory retrieval (**Fig. 2a**). Together, these results demonstrate that fear learning is under bidirectional control of PV$^+$ INs in the BLA during the US.

To relate the behavioural changes induced by manipulation of PV$^+$ IN activity to the physiological role of PV$^+$ INs, we performed chronic single unit recordings from

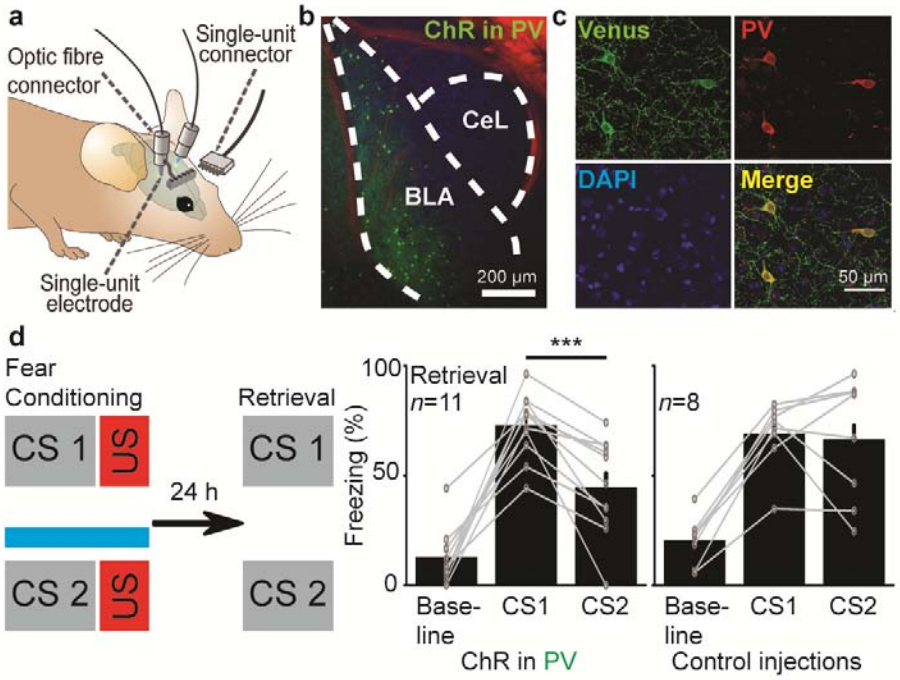


Figure 1. BLA PV$^+$ interneurons control fear learning.
a, Optogenetic manipulation and simultaneous single unit recordings using an optrode in freely behaving mice. **b,** Co-expression of ChR2 and Venus (green) in PV$^+$ INs in the BLA, using PV-CRE mice and conditional rAAVs. Counterstains: DAPI (blue) and NeuN (red). **c,** Restriction of ChR2 and Venus expression (green) to PV$^+$ INs is verified by PV IHC (red). **d,** Left, Fear conditioning paradigm to test for effects of PV$^+$ manipulation on fear learning. Animals were conditioned to two distinct auditory CSs. During conditioning, CS2 was paired with blue light for ChR2 activation during the entire CS-US pairing. During retrieval on the next day, both CSs were presented without light. CS order and pairing with blue light were counterbalanced. Right, animals displayed significantly less freezing to the light-paired CS2 during retrieval compared to CS1, indicating that PV$^+$ activation during the entire CS-US pairing impairs fear acquisition. Light stimulation had no effect on control animals expressing only GFP. Values are mean +/- s.e.m. ***P < 0.001. Statistical analysis in Methods.

optogenetically identified PV$^+$ cells (Lima et al. 2009). One of the implanted optic fibres served as an optrode, with an integrated 16-channel recording electrode, allowing for simultaneous optogenetic stimulation and single unit recordings in the BLA (**Fig. 1a; Ext. Data Fig. 1**). Units exhibiting significant, short latency activity changes upon illumination with blue or yellow light were identified as PV$^+$ INs (**Fig. 2b,c; Ext. Data Fig. 2a-e**; see Methods for identification criteria). In agreement with previous recordings from BLA PV$^+$ INs in slices (Woodruff and Sah 2007b, a), optogenetically identified PV$^+$ INs showed faster spike kinetics and higher levels of spontaneous activity compared with simultaneously recorded non-light responsive units (**Fig. 2d; Ext. Data Fig. 3**). Recordings from identified light-responsive units (n = 31) revealed that firing of the PV$^+$ IN population was reduced for the entire duration of the footshock (**Fig. 2e**), although a minority displayed excitation (7 out of 31, **Ext. Data Fig. 4a**). These data are consistent with the observed behavioural effects:

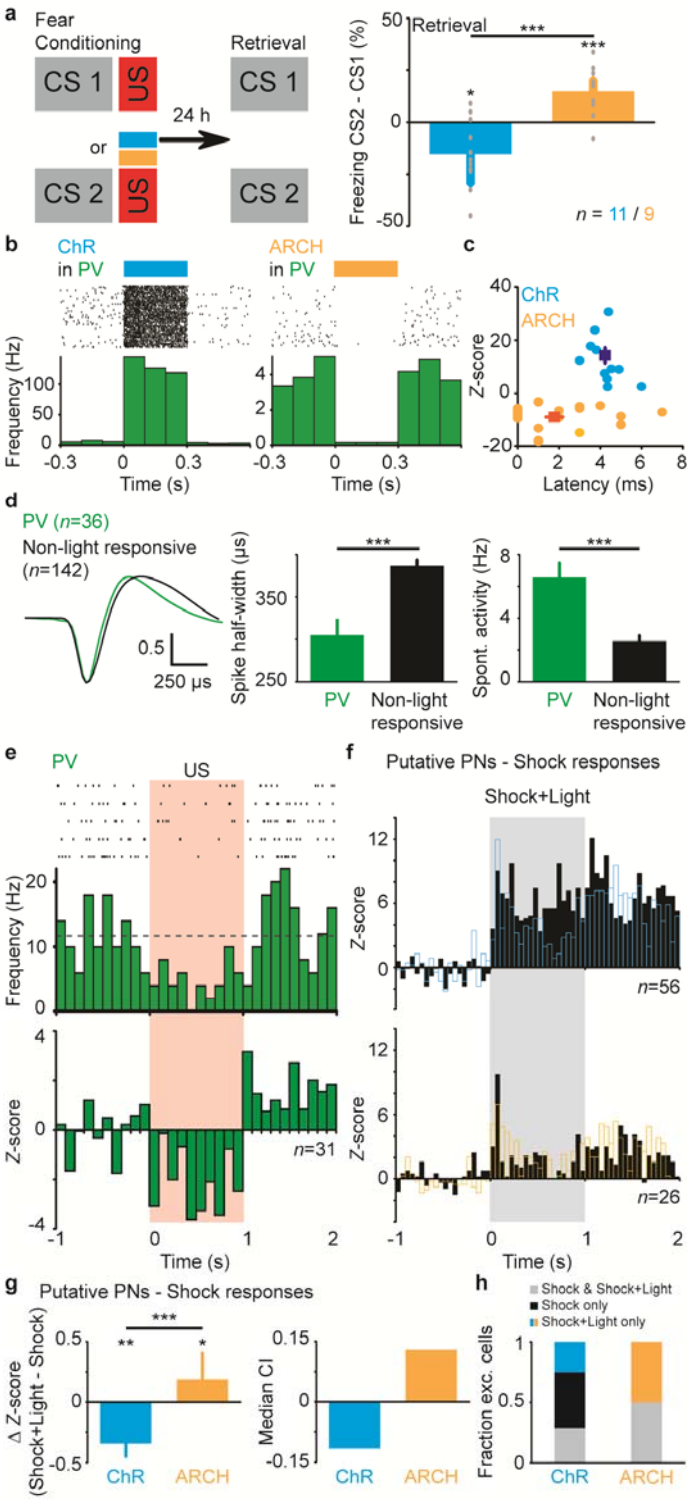


Figure 2. Footshock-induced PV$^+$ cell inhibition gates fear learning.
a, PV$^+$ activity during the US regulates fear learning. Left, animals were conditioned as in Fig. 1d, but illumination with blue or yellow light was restricted to the US period. Right, PV$^+$ activation and inhibition during the US had opposite effects on fear learning. Shown are differences in freezing between the two CSs. **b,** Optogenetic identification of PV$^+$ INs in the BLA of freely-behaving mice, based on light-induced changes in single unit activity (ChR2-mediated increase with blue light; ARCH-mediated decrease with yellow light). **c,** Light responsive units were defined as those with significant, short-latency light-induced changes in activity. **d,** Optogenetically identified PV$^+$ units differ in spike width and spontaneous firing rate from non-light responsive units. Non-light responsive units were simultaneously recorded in the same animals and mainly represent PNs. **e,** Footshocks inhibit PV$^+$ cell activity during conditioning (top, example response; bottom, *Z*-scored population response). **f,** Z-scored population responses for footshock-excited putative PNs with (blue or yellow) and without (black) optogenetic stimulation of PV$^+$ INs. The population includes all cells with a significant footshock response with or without light. Putative PNs are non-light responsive cells with baseline firing <3 Hz. **g,** Left, *Z*-score differences of individual footshock-excited putative PNs during the footshock induced by optogenetic manipulation of PV$^+$ INs. Right, Median Change Index for the footshock responses of putative PNs upon PV$^+$ cell manipulations. The change index is calculated as CI = (*Z*-score(Footshock + Light) - *Z*-score(Footshock)) / (*Z*-score(Footshock + Light) + *Z*-score(Footshock)) (all *Z*-scores are the averages of the footshock-response during the entire footshock). **h,** Fractions of the excited putative PNs, which are excited either exclusively during the footshock (black), exclusively during the footshock + light (blue or yellow) or in both conditions (grey). Values are mean unless stated otherwise, error bars indicate s.e.m. **P* < 0.05, ***P* < 0.01, ****P* < 0.001. Statistical analysis in Methods.

Optogenetic activation of PV^+ units counteracted the endogenous US-induced inhibition and decreased learning, while optogenetic inhibition enhanced the endogenous PV^+ inhibition resulting in enhanced fear learning (**Fig. 2a**).

Since PV^+ INs mediate strong perisomatic inhibition (Muller et al. 2006; Rainnie et al. 2006; Woodruff and Sah 2007a), we speculated that US-induced inhibition of PV^+ cells causes disinhibition of PNs. We therefore analysed the impact of PV^+ IN activity on excitatory footshock responses in putative PNs (non-light responsive units with spontaneous firing rates < 3 Hz; n = 82) (McDonald 1984; Sah et al. 2003; Likhtik et al. 2006; Pape and Pare 2010; Fishell and Rudy 2011). Indeed, we observed a significant decrease in excitatory footshock responses upon optogenetic activation of PV^+ INs (**Fig. 2f, g; Ext. Data Fig. 5**). In contrast, inhibition of PV^+ INs caused a significant increase of the PN responses (**Fig. 2f, g; Ext. Data Fig. 5**). In addition to affecting the strength of the footshock-induced responses, optogenetic manipulations of PV^+ IN activity also differentially changed the number of footshock-excited cells. While PV^+ activation led to a loss of excitation in almost half of the footshock-excited cells, PV^+ inhibition uncovered excitatory responses in a large fraction of putative PNs (**Fig. 2h; Ext. Data Fig. 5**). Together, these results indicate that footshock induced inhibition of PV^+ cells regulates activity in the BLA network in a manner consistent with our behavioural observations: PV^+ IN inhibition causes disinhibition and enhances US-induced activity in principal neurons, likely acting as a permissive gate for plasticity induction and fear learning.

Differential roles for PV^+ and SOM^+ INs during the CS

We next addressed the question whether inhibition of PV^+ INs and the consequent disinhibition of PNs also occurs during the auditory CS which precedes the footshock US. Unexpectedly, when optogenetically manipulating the activity of PV^+ INs exclusively during the CS, we found the opposite effects on fear learning compared to US-specific manipulations (**Fig. 3a**). This suggests fundamentally different roles of PV^+ INs in processing of CS and US.

In line with this notion, recordings showed that the PV^+ IN population exhibited CS-induced excitation (**Fig. 3b**), although a subset displayed inhibitory responses (14 out of 36, **Ext. Data Fig. 4c**). If PV^+ INs influenced PNs mostly via mono-synaptic perisomatic inhibition, we would expect optogenetic manipulations of PV^+ INs to have an effect on learning opposite to the one we observed.

A possible explanation for the paradoxical effect of PV^+ INs on fear conditioning could be that PV^+ INs interact not only with PNs, but in addition with other IN subtypes to cause disynaptic disinhibition of PNs. Because sensory afferents from auditory thalamus and

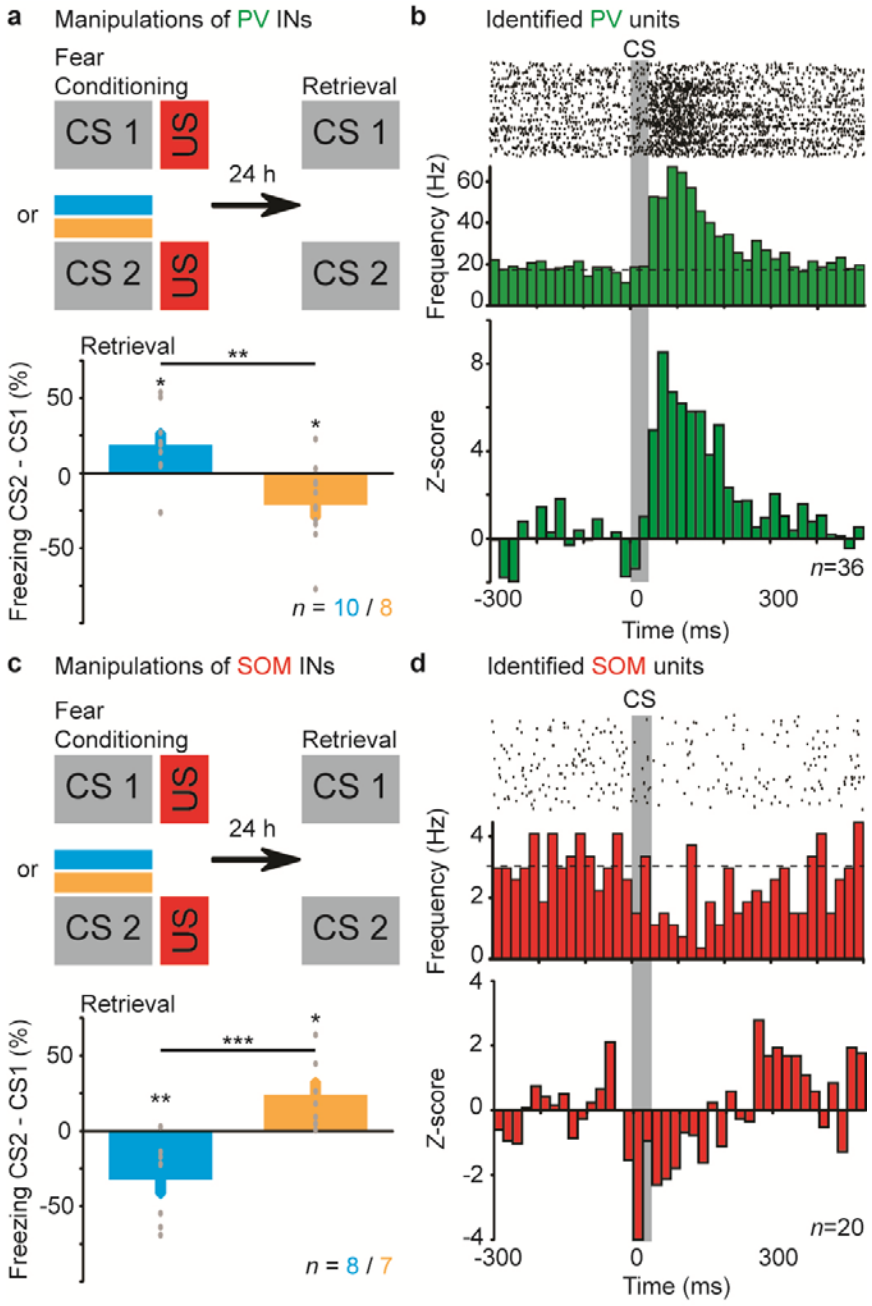


Figure 3. Differential roles of PV$^+$ and SOM$^+$ interneurons during the CS.
a, Top, animals were conditioned as described in Fig. 2a, but with optogenetic manipulation restricted to the CS. Bottom, Manipulations of PV$^+$ activity during the CS have bidirectional effects on learning, with an opposite sign compared to manipulations during the US (Fig. 2a). **b,** Recordings of optogenetically identified PV$^+$ cells show that they are excited by the CS during conditioning (top, example trace, dashed line indicates average baseline firing; bottom, population response). **c,** Top, animals expressing opsins in SOM$^+$ INs in the BLA were conditioned as in (a). Bottom, SOM$^+$ activity manipulations had bidirectional effects on learning, opposite compared to PV$^+$ manipulations. **d,** SOM$^+$ cells are inhibited by the CS during conditioning (top, example trace, dashed line indicates average baseline firing; bottom, population response). Values are mean +/- s.e.m. *P < 0.05, **P < 0.01, ***P < 0.001. Statistical analysis in Methods.

cortex make synapses onto the dendrites of BLA PNs (Pape and Pare 2010), we hypothesized that dendrite-targeting SOM^+ INs might mediate such an effect. To address this, we performed optogenetic manipulations and single-unit recordings from identified SOM^+ INs using the same strategy as described for PV^+ INs (**Ext. Data Fig. 1; 2; 3**) (Taniguchi et al. 2011). In agreement with the hypothesis that SOM^+ INs could be inhibited by PV^+ cells during the CS, we found that the behavioural consequences of optogenetic manipulations of SOM^+ cells during the CS were opposite to those induced by PV^+ manipulations (**Fig. 3c**), and that CS presentation induced inhibition of identified SOM^+ INs (**Fig. 3d; Ext. Data Fig. 4d**). Taken together, these results indicate that the CS causes subcellular reallocation of inhibition in PNs, with disinhibition of sensory dendritic inputs and a concomitant increase in perisomatic inhibition.

Dendritic disinhibition by SOM^+ INs

To directly address whether PV^+ INs indeed make synaptic contacts onto SOM^+ INs as has been demonstrated in the hippocampus (Lovett-Barron et al. 2012) (**Fig. 4a**, top left), we injected a rAAV containing floxed ChR2-mCherry into PV-Cre/GAD67-EGFP double knock-in mice. Whole-cell recordings from EGFP-expressing INs (n = 42) in acute brain slices were combined with optogenetic activation of PV^+ INs (**Ext. Data Fig. 6**). Post-hoc immunocytochemical identification revealed that 3 out of 4 recorded SOM^+ INs received reliable inhibitory input from PV^+ cells (**Fig. 4a,** top right and bottom; **Ext. Data Fig. 6**).

We next asked whether SOM^+ INs can control the processing of sensory information in PNs. To this end, optical stimulation of ChR2-mCherry expressing SOM^+ INs was combined with recordings from BLA PNs in acute brain slices. Electrical stimulation of thalamic sensory afferents evoked reliable EPSPs (Bissiere et al. 2003) (**Fig. 4b-e**), whereas direct optical stimulation of SOM^+ INs evoked IPSPs (**Fig. 4d,e**). Co-activation of excitation and inhibition caused a significant reduction of both the size and the decay time of the evoked EPSPs (**Fig. 4d,e,g**). In addition, spike output induced by supra-threshold thalamic stimulation was significantly reduced by simultaneous activation of SOM^+ INs (**Fig. 4f,g**). Together, these results demonstrate that SOM^+ INs can modulate excitatory inputs onto BLA PNs and thereby influence PN spiking output.

The emerging picture is that CS processing in BLA is under the control of a microcircuit in which PV-to-SOM inhibition disinhibits PN dendrites during tone presentation, while the PN perisomatic domain receives increased inhibition directly from PV^+ INs. In contrast, during the US both PV^+ and SOM^+ INs are inhibited (**Fig. 2e; Ext. Data Fig. 4a,b**), leading to a qualitatively different pattern of inhibition along the somato-dendritic axis of PNs (see Discussion).

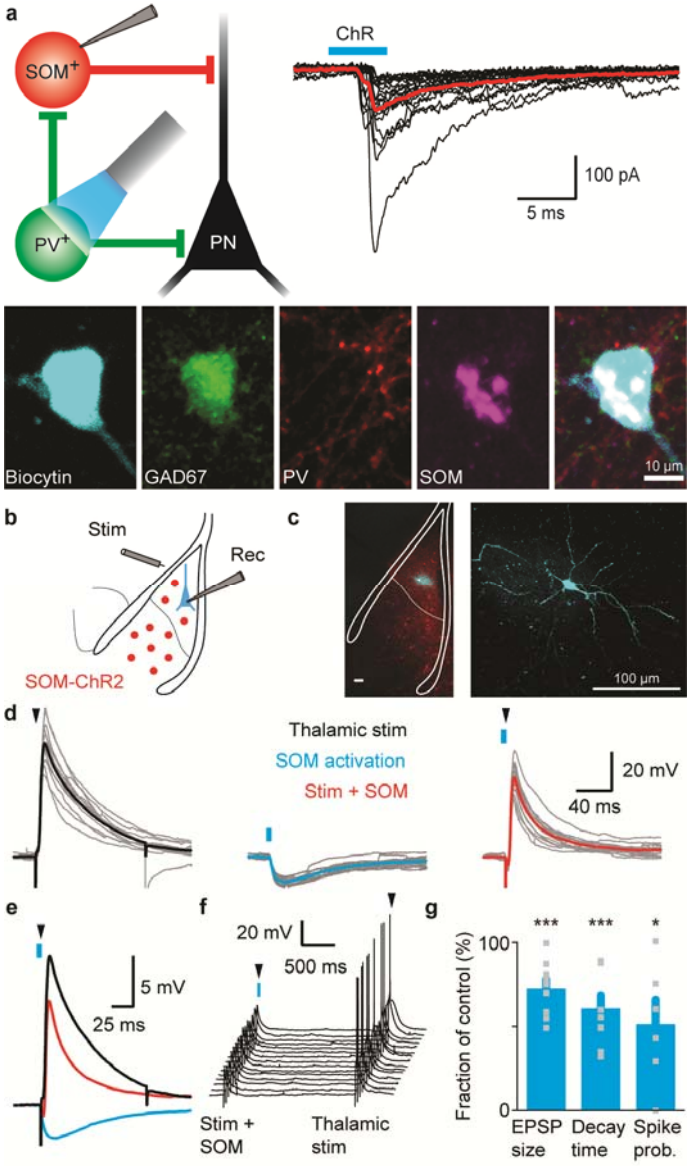


Figure 4. PV$^+$ and SOM$^+$ interneurons form a disinhibitory microcircuit in the BLA.
a, Top left, Scheme of proposed PV$^+$ and SOM$^+$ connectivity and experimental setup. Top right, Whole-cell recordings of EGFP$^+$ cells in GAD67::EGFP-PV::CRE mice, infected with an rAAV-FLEX-ChR2-mCherry. Inhibitory inputs in SOM$^+$ cells evoked by optogenetic activation of PV$^+$ INs. Bottom, Post-hoc IHC identification of SOM$^+$ cells (EGFP$^+$ cells n = 42, of those SOM$^+$ n = 4, SOM$^+$ and input n = 3, see Ext. Fig. 6d) **b-g,** SOM$^+$ INs modulate excitatory inputs and neuronal output of BLA pyramidal neurons in slices. **b,** Experimental design. **c,** ChR2-mCherry-positive SOM$^+$ INs in the BLA (red, left) and example of recorded pyramidal neuron (blue, left; right high magnification). **d,** Example traces of excitatory thalamic (left, electrical stimulation) and inhibitory SOM$^+$ cell evoked (middle, 5 ms blue light) postsynaptic potentials as well as coactivation of thalamic excitation and SOM$^+$ inhibition (right, electrical stimulation 5 ms after light onset). **e,** Averages of recordings in (d). **f,** Spike output of pyramidal neurons in response to thalamic afferent stimulation with and without SOM$^+$ cell activation. **g,** Normalized amplitude and decay time of thalamic EPSPs and of spike probability in response to thalamic EPSPs with simultaneous activation of SOM-inhibition. Values are mean +/- s.e.m. *P < 0.05, ***P < 0.001. Statistical analysis in Methods.

Impact of PV$^+$ and SOM$^+$ INs on BLA network activity

To evaluate how inhibition supplied by PV$^+$ and SOM$^+$ INs affects CS processing in the BLA, we compared tone-evoked excitatory responses in putative PNs with and without optogenetic manipulation of the INs. Activation of SOM$^+$ cells resulted in a strong reduction of tone-evoked activity in PNs, whereas inhibition of SOM$^+$ INs caused a slight enhancement (**Fig. 5a,b; Ext. Data Fig. 7**), supporting the idea that SOM$^+$ cells directly inhibit PN tone input and/or responses. Activation of PV$^+$ INs, on the other hand, caused a substantial enhancement of tone responses, whereas inhibition of PV$^+$ INs had little effect on tone responses (**Fig. 5a,b; Ext. Data Fig. 7**), suggesting that direct PV-PN inhibition is secondary to the indirect PV-SOM-PN disinhibitory pathway during CS processing. In addition to effects on the average CS response of putative PNs, PV$^+$ and SOM$^+$ INs also differentially affected the number of tone-excited PNs. While conditions impairing learning (SOM-ChR2 and PV-ARCH; **Fig. 3**) reduced the number of tone-excited cells, manipulations enhancing learning (PV-ChR2 and SOM-ARCH; **Fig. 3**) uncovered excitatory tone responses in large populations of putative PNs that were previously unresponsive to tone (**Fig. 5c; Ext. Data Fig. 7**). Together, these results demonstrate that PV$^+$ and SOM$^+$ INs have opposite effects on CS-responses that are consistent with the observed impact on learning (**Fig. 3**): Manipulations that increased fear learning caused stronger tone responses in more PNs, whereas conditions that impaired learning led to smaller responses in fewer PNs. In addition, these data suggest that the described disinhibitory microcircuit is an important factor governing tone processing in the BLA during learning. Interestingly, spontaneous activity of putative PNs was not differentially affected by manipulations of the two IN populations (**Fig. 5d; Ext. Data Fig. 8a,c**), suggesting that these effects are stimulus-specific and reflect the organization of sensory inputs onto PNs. Furthermore, the finding that manipulations of SOM+ INs had a significantly stronger impact on PN activity than manipulations of PV$^+$ INs is consistent with the interpretation that indirect effects via inhibition of SOM$^+$ cells may outweigh the impact of direct perisomatic inhibition supplied by PV$^+$ INs.

DISCUSSION

The level of activation of a PN in the BLA during fear learning is a key factor determining whether inputs to this PN will undergo synaptic plasticity and whether cellular plasticity is triggered that enables recruitment of the neuron into the memory trace (Han et al. 2007; Ehrlich et al. 2009; Zhou et al. 2009). Consistent with a central role of synaptic inhibition in this process, we observed strong effects on fear memory retrieval when we manipulate distinct IN types during fear memory acquisition. In addition, our data point to a multifaceted role of inhibition in acquisition of conditioned fear: We find that the effects of optogenetic manipulation of IN activity strongly depend on both stimulus identity (CS vs. US) and on IN type (PV$^+$ vs. SOM$^+$). Finally, our results identify IN-IN interactions as a novel and important factor governing auditory responses and learning.

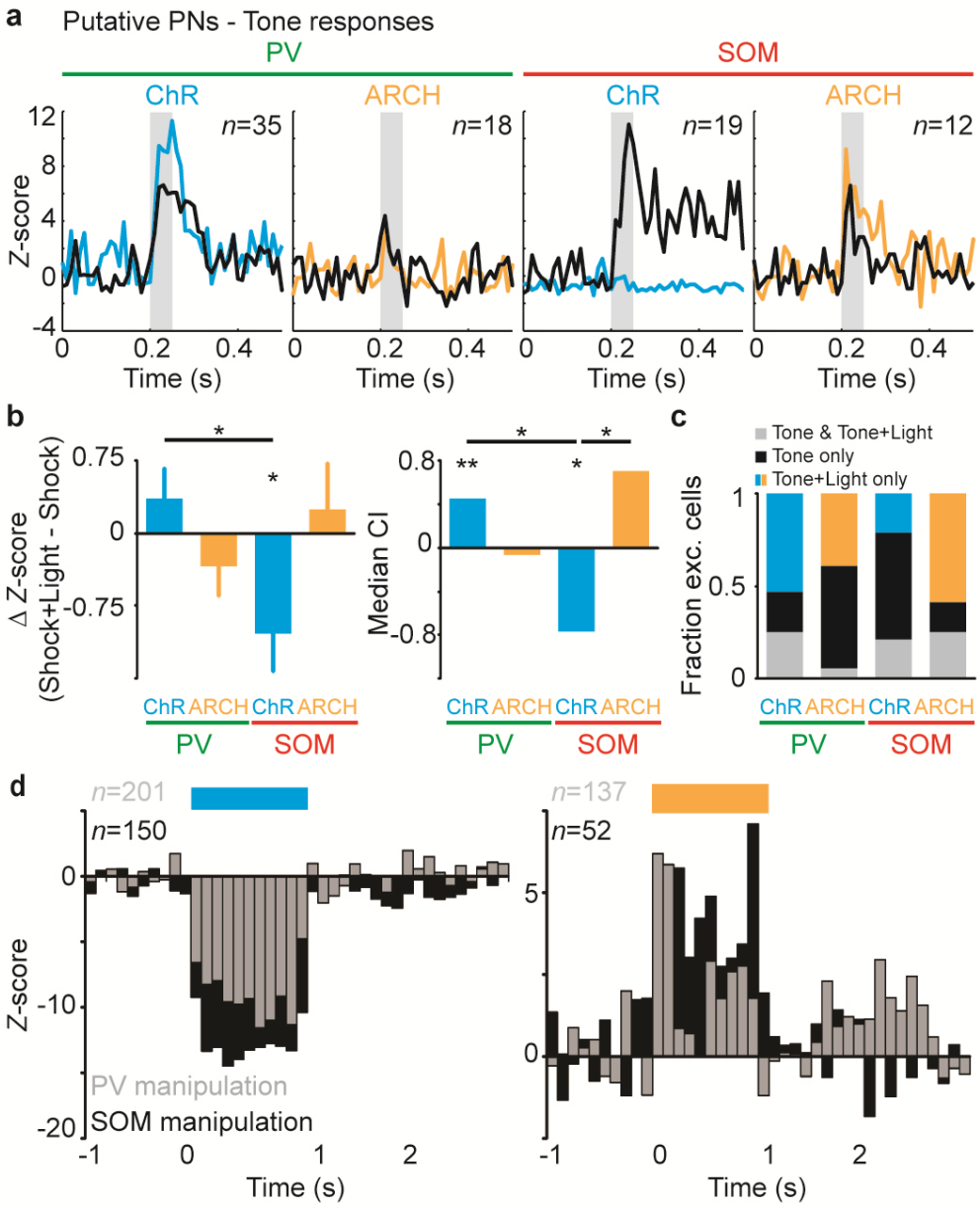


Figure 5. Distinct effects of PV$^+$ and SOM$^+$ interneurons on BLA network activity.
a, Tone-evoked population responses in tone-excited putative PNs with (blue or yellow) and without (black) optogenetic stimulation of PV$^+$ or SOM$^+$ INs. The populations include all cells with a significant tone response with or without light. **b,** Left, *Z*-score differences of individual tone-excited putative PNs during the tone induced by optogenetic manipulations of PV$^+$ or SOM$^+$ INs. Right, Median Change Index for the tone responses of putative PNs upon PV$^+$ or SOM$^+$ cell manipulations (See Methods). **c,** Fractions of the excited putative PNs, which are excited either exclusively during the tone (black), exclusively during the tone + light (blue or yellow) or in both conditions (grey). **d,** Bidirectional changes in spontaneous activity of putative PNs upon optogenetic manipulations of PV$^+$ and SOM$^+$ INs. Activation of both IN types reduces (left) and inhibition enhances (right) activity in the BLA network. Note the more pronounced effects of SOM$^+$ IN manipulations. Values are mean unless stated otherwise, error bars indicate s.e.m. **P* < 0.05, ***P* < 0.01, ****P* < 0.001. Statistical analysis in Methods.

During the presentation of an auditory cue, PV^+ INs are, on average, excited, most likely through direct sensory input from auditory thalamus and auditory cortex (Woodson et al. 2000; Sah et al. 2003). Similar to what has been described in other cortical structures (Markram et al. 2004; Somogyi and Klausberger 2005), this will lead to somatic feedforward inhibition in postsynaptic PNs, a process that would be expected to restrict their activation and, as a consequence, the induction of synaptic plasticity during learning. Unexpectedly, our results indicate that increasing PV^+ cell activity during the CS correlates with enhanced learning, and with stronger auditory responses in PNs. These observations can be explained in light of our finding that PV^+ INs not only contact PNs, but also supply strong inhibition to dendrite-targeting SOM^+ INs in the BLA. Moreover, activation of SOM^+ INs reduces learning and decreases auditory responses in BLA PNs. Together, these results strongly suggest that dendritic disinhibition in PNs overrides the effect of increased perisomatic inhibition, potentially through increased dendritic electrogenesis and associated burst firing (**Ext. Data Fig. 8b,d**) (Gentet et al. 2012) as has been demonstrated in the hippocampus (Lovett-Barron et al. 2012). Furthermore, it is conceivable that locally restricted regulation of dendritic inhibition adds a level of control by determining which inputs contribute to action potential firing and undergo plasticity upon association with the US.

In contrast, during the aversive US PV^+ cells were inhibited, and optogenetic augmentation of US-induced inhibition enhanced fear learning. This suggests that inhibition of PV^+ INs represents an important, permissive mechanism for fear learning, most likely by gating associative plasticity induction during the US, as has recently been shown in auditory cortex (Letzkus et al. 2011). Our data now extend and complement this view by demonstrating that SOM^+ cells are also inhibited during the US, suggesting disinhibition of the entire somatodendritic axis of PNs, thereby strongly increasing their activity. Since auditory inputs synapse onto dendrites (Farb and LeDoux 1997; McDonald 1998), dendritic disinhibition represents an important, novel mechanism in auditory fear learning, favouring the induction of synaptic plasticity during the aversive US. A likely source of US-induced inhibition are other IN subtypes that can contact both PV^+ and SOM^+ INs (Fishell and Rudy 2011; Letzkus et al. 2011; Jiang et al. 2013), and/or fast neuromodulatory processes.

Different types of INs exhibit distinct temporal dynamics during different network states that correlate with specific aspects of behavior (Klausberger and Somogyi 2008; Kvitsiani et al. 2013). It has been postulated that direct interactions between soma-targeting and dendrite-targeting INs are an important mechanism by which spatiotemporal patterns of inhibition might be regulated (Lovett-Barron et al. 2012). Here, we provide evidence that similar mechanisms can govern associative learning and the strength of the memory. In the identified microcircuit, differential modulation of PV^+ and SOM^+ INs may permit flexible regulation of learning according to the behavioural context and the animal's internal state.

METHODS SUMMARY

Virus injections and optogenetic stimulation

Animals were stereotaxically injected bilaterally into the BLA with recombinant AAVs (serotype 2/7 or 2/5) for conditional expression of ChR2 or ARCH. Optic fibres with a diameter of 200 µm were chronically implanted above the BLA. Animals were subjected to behavioural analysis at least 4 weeks after injection. Opsin expressing cells were stimulated with a laser (473 nm or 593 nm).

Behaviour

Injected and implanted PV-CRE or SOM-CRE mice were fear conditioned with two different CSs. While the CS1 served as a within-animal control, the CS2 was paired with optogenetic activation or inhibition of PV^+ or SOM^+ neurons either during CS and US, during CS only, or during US only. Freezing behaviour was quantified with an automatic infrared beam detection system (Herry et al. 2008). Mice in which optogenetically identified PV^+ or SOM^+ cells were recorded were subjected to discriminatory fear conditioning, in which the CS+, but not the CS^- was paired with the US. Additionally, we tested the effect of optogenetic manipulations on tone or footshock responses.

Single unit recordings and analysis

Custom-built optrodes were used to extracellularly record individual units in freely behaving mice. Spike sorting was based on time-amplitude window discrimination and template matching as previously described (Herry et al. 2008). Cluster separation and unit isolation were verified (Herry et al. 2008). Sweeps were aligned to CS or US onset and evoked responses were normalized to baseline activity using a Z-score transformation. PV^+ or SOM^+ neurons were optogenetically identified as units exhibiting significant, short-latency light-evoked responses or light-evoked inhibition.

Slice electrophysiology

Acute brain slices were prepared as described (Bissiere et al. 2003). To test connectivity, we optogenetically activated PV^+ cells in PV-CRE:GAD67-EGFP animals and patched $EGFP^+$ INs. Standard immunohistochemical methods were used for post-hoc identification of SOM^+ neurons. Optogenetic activation of SOM^+ cells was used to assess the effect on EPSPs and spike output evoked by electrical thalamic stimulation in PNs.

Acknowledgements

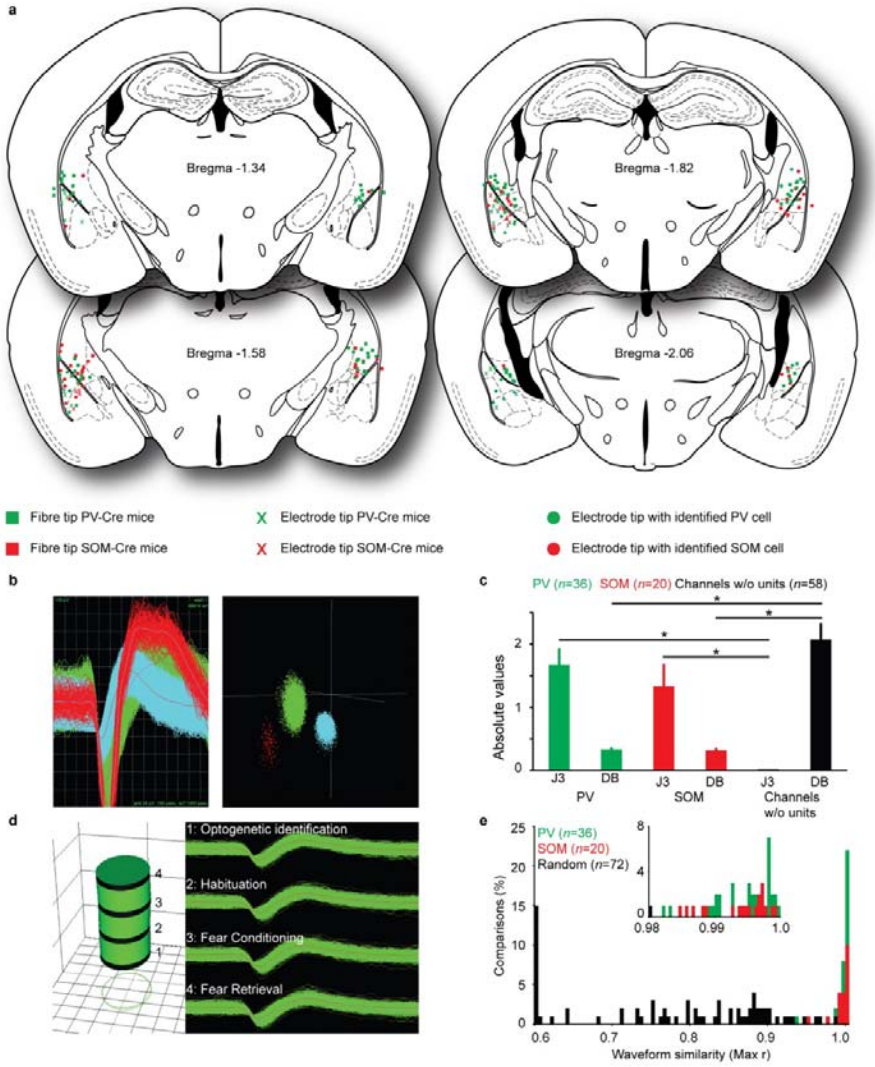
We thank all members of the Lüthi lab for discussions and critical comments on the manuscript. We would like to thank Julia Lüdke, Paul Argast and Peter Buchmann for excellent technical assistance. We also thank Karl Deisseroth, Ed Boyden, Josh Huang, Rolf Sprengel and Silvia Arber for generously sharing materials and mouse lines. This work was supported by the Novartis Research Foundation, by the National Center of Competences in

Research “SYNAPSY - The Synaptic Bases of Mental Diseases” financed by the Swiss National Science Foundation as well as a core grant to AL. SBEW is supported by a Schering Foundation Fellowship. JJL is supported by a Swiss National Science Foundation Ambizione Fellowship. JG and GAJ are supported by EMBO Long-Term Fellowships and Marie Curie Action Fellowships.

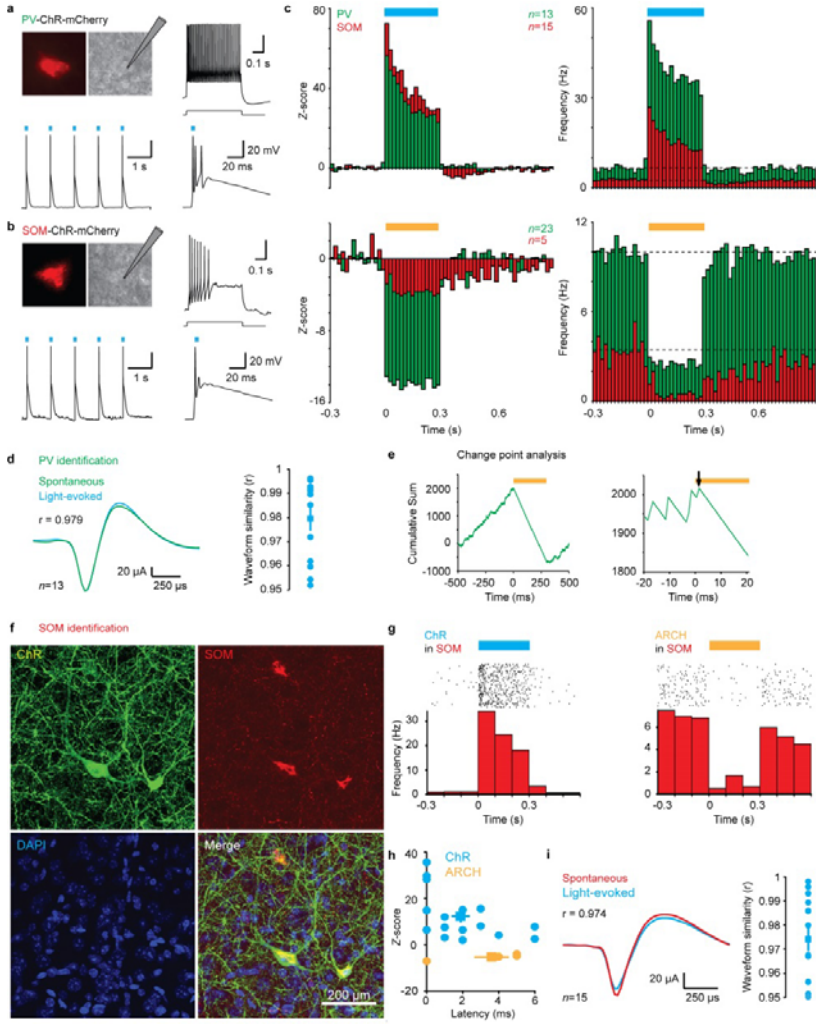
Author contributions

S.B.E.W. initiated the project and performed most experiments and data analysis. J.G. performed and analysed in vitro experiments. P.T. performed and analysed immunohistochemistry. G.A.J. and R.W.F. performed and helped with data analysis. C.H. and I.E. helped to establish optogenetic manipulations and single unit recordings. S.B.E.W., J.J.L. and A.L. conceived the project and wrote the manuscript. All authors contributed to the experimental design and interpretation, and commented on the manuscript.

EXTENDED DATA

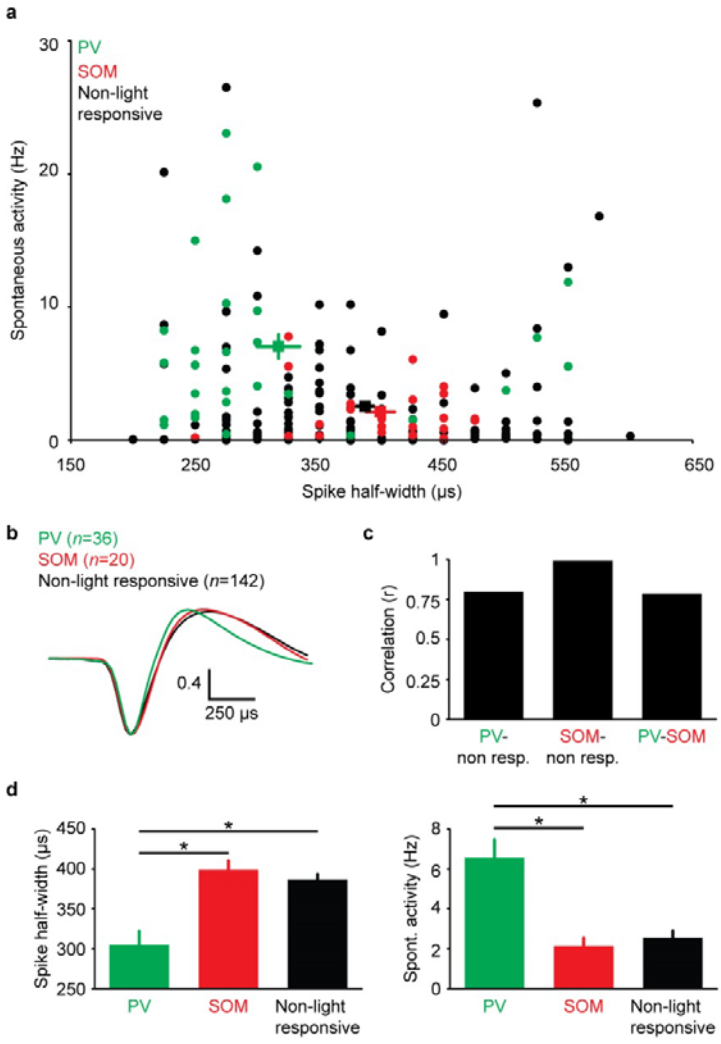


Extended Fig. 1: Fibre and electrode placement sites and single-unit recording quality.
a, Position of electrode tips for single unit recordings and of bilateral optic fibres for all PV-CRE and SOM-CRE animals used. **b,** Left, Superimposed waveforms of three single units recorded on the same channel in the BLA of a PV-CRE animal. Right, Spikes from individual units were sorted into clusters, using 3D principal component analysis. **c,** Recording quality was evaluated by calculation of J3 and Davies Bouldin validity index (DB) statistics for PV^+ and SOM^+ interneurons. To obtain control values two clusters were selected in the central noise cloud in channels without units. High J3 and low DB values indicate good isolation of single units. **d,** Left, Principal component (PC) space cylinders to determine the stability of clustered waveforms during long-term recordings. Straight cylinders indicate stable recordings of the same set of single units. Right, Superimposed waveforms recorded during the different sessions, used for the calculation of the PC space cylinders. **e,** For a quantitative evaluation of spike shape similarity over different recording days, we calculated linear correlation values for PV^+ and SOM^+ INs. As a control, we calculated the r values for waveforms of different random neurons. The maximum r value across recording sessions was used to quantify similarity (r = 1 indicates identical spike shapes). All PV^+ and SOM^+ INs had r values above 0.95, but only about 4% of the random comparisons. Values are mean +/- s.e.m. *P<0.05. Statistical analysis in Methods.



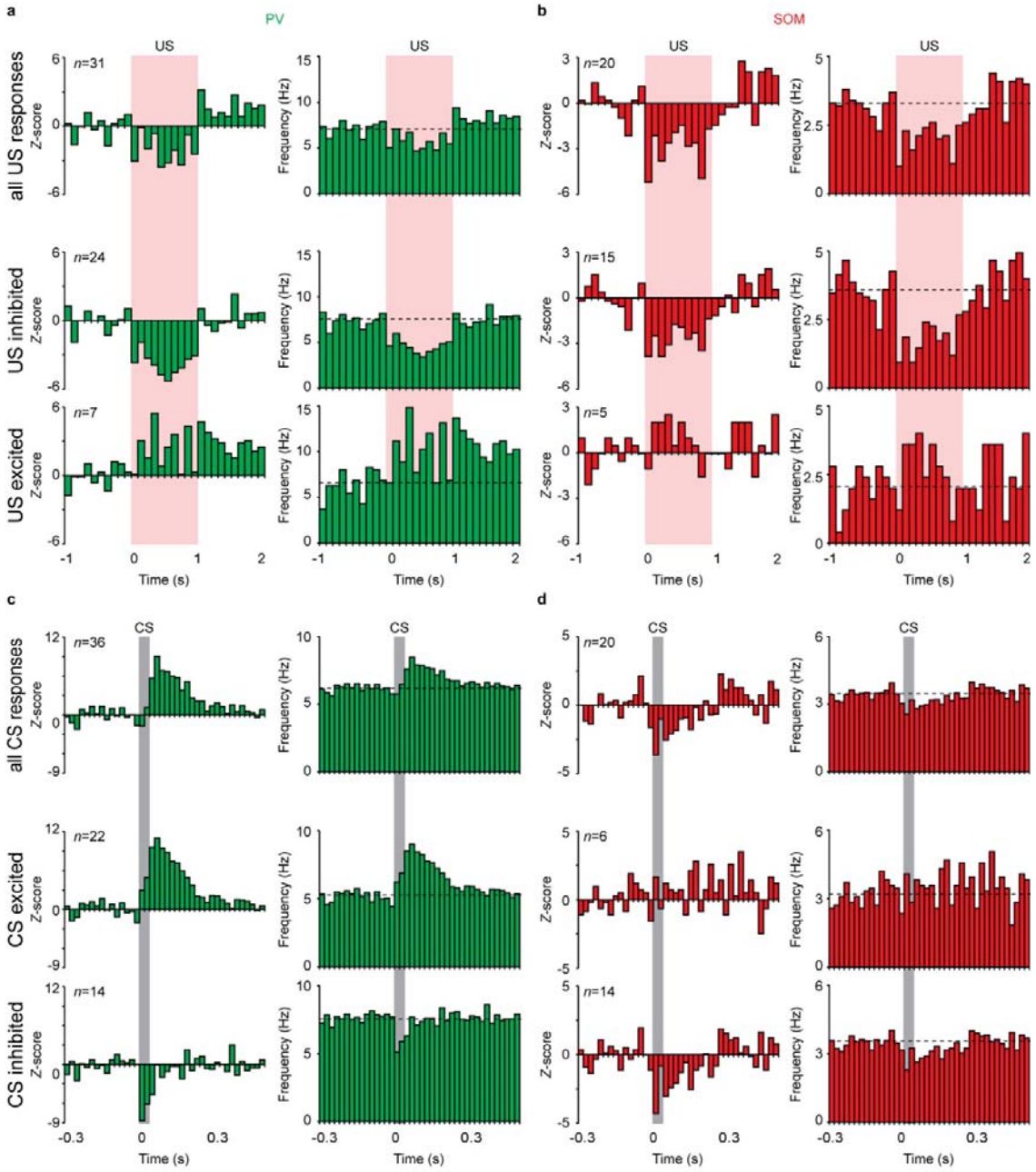
Extended Fig. 2: Optogenetic identification and manipulation of PV$^+$ and SOM$^+$ INs.

a,b, Top left: Ex vivo whole-cell patch clamp recordings of ChR2-mCherry-expressing PV$^+$ (a) and SOM$^+$ (b) BLA INs in amygdala slices. Top right: Distinct firing patterns of PV$^+$ (a) and SOM$^+$ (b) cells in response to depolarizing somatic current injection. Bottom: PV$^+$ (a) and SOM$^+$ (b) cells fire brief bursts of action potentials in response to blue light stimulation (468 nm, 5 ms, 10 mW). **c,** Single unit recordings of optogenetically identified PV$^+$ (green) and SOM$^+$ (red) interneurons upon 300 ms stimulation with blue (ChR2, top) or yellow (ARCH, bottom) light in behaving animals. Left: *Z*-scored activity. Right: Firing frequency. **d,** Left: Comparison of spontaneous and light-evoked superimposed average spike waveforms of all PV$^+$ INs identified by optogenetic activation. Right: Linear correlations between spontaneous and light-evoked spikes were calculated for individual optogenetically identified PV$^+$ INs. Only cells with r values above 0.95 were considered as directly light-activated. **e,** Change point analysis for determination of the latency of light-evoked activity changes. The cumulative sum of the activity of a neuron was calculated and the change point was determined. Shown is an example of ARCH-mediated inhibition for the entire illumination period (left) and at the light onset (right). The arrow indicates the change point. **f,** Expression of ChR2, co-expressed with Venus (green) is restricted to SOM$^+$ INs (red). **g,** Light-induced changes in activity of SOM$^+$ interneurons upon illumination with either blue or yellow light for activation of ChR2 or ARCH, respectively. **h,** Latencies and *Z*-scores of light-induced activity changes in optogenetically identified SOM$^+$ interneurons. **i,** Left, Comparison of average waveforms of spontaneous and light-evoked spikes in SOM$^+$ interneurons. Right, Linear correlations between spontaneous and light-evoked spikes in optogenetically identified SOM$^+$ INs.



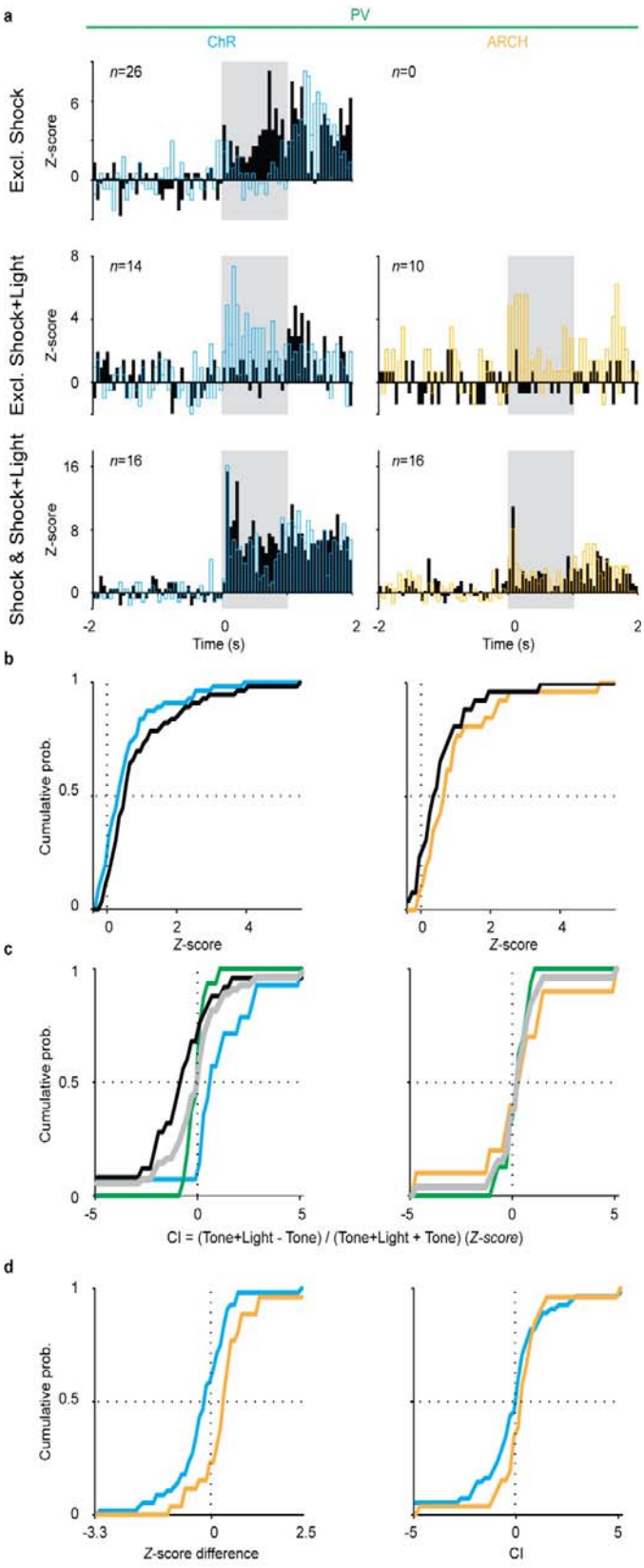
Extended Fig. 3: Physiological properties of PV$^+$ and SOM$^+$ interneurons and of non-light responsive cells.

a, Distribution of physiological properties for optogenetically identified PV$^+$ and SOM$^+$ INs and for non-light responsive cells. Spike half-width and spontaneous activity were compared for identified INs and for a population of simultaneously recorded non-light responsive cells, which is mainly comprised of principal cells, but could also include unidentified INs. Squares indicate average values for the three groups. SOM$^+$ INs cannot be separated from the non-light responsive cells while PV$^+$ cells are on average different from the other groups. **b,** Comparison of normalized spike waveforms of PV$^+$ and SOM$^+$ INs and non-light responsive cells. **c,** Linear correlations of the spike waveforms between the different groups. **d,** Comparison of spike half-width and spontaneous activity of PV$^+$ and SOM$^+$ INs and non-light responsive cells. Note the similarity of SOM$^+$ INs with non-light responsive cells. Values are mean +/- s.e.m. **P*<0.05. Statistical analysis in Methods.



Extended Fig. 4: US and CS responses in PV$^+$ and SOM$^+$ INs in the BLA.

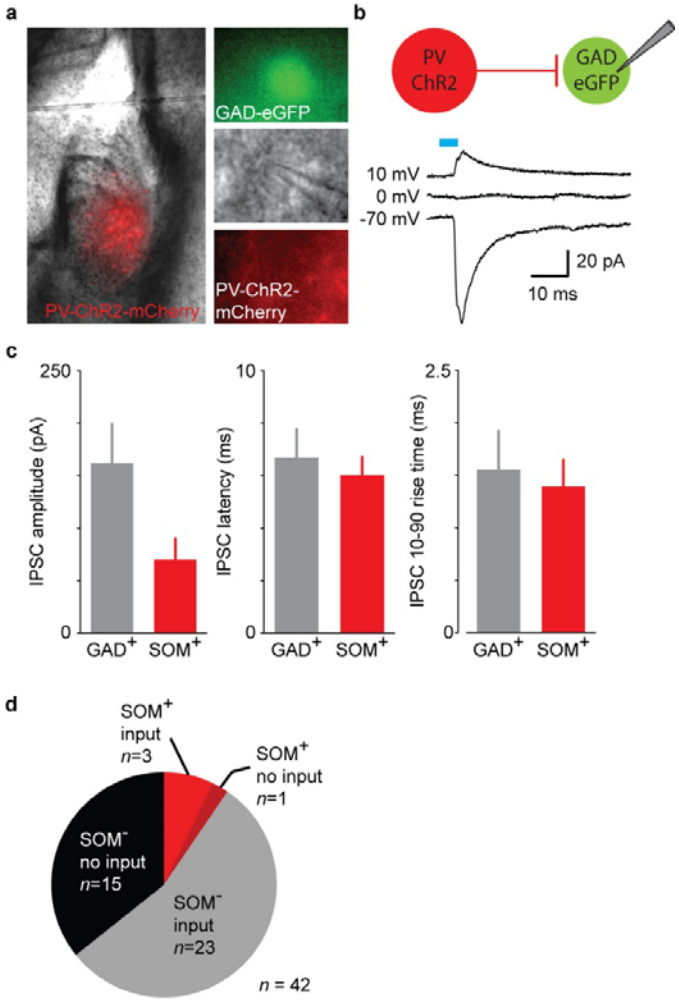
a-d, Z-scored population activity (left) and firing frequency (right) for identified PV$^+$ (green) and SOM$^+$ (red) cells in response to the US (a,b) or the CS (c,d). Shown are the population responses for all cells (top graphs). The graphs below show population responses that were separated according to whether the cells showed inhibitory or excitatory responses to the stimuli. Cells were included in the stimulus inhibited group or stimulus excited group if their average *Z*-scored stimulus response was < 0 or > 0, respectively.



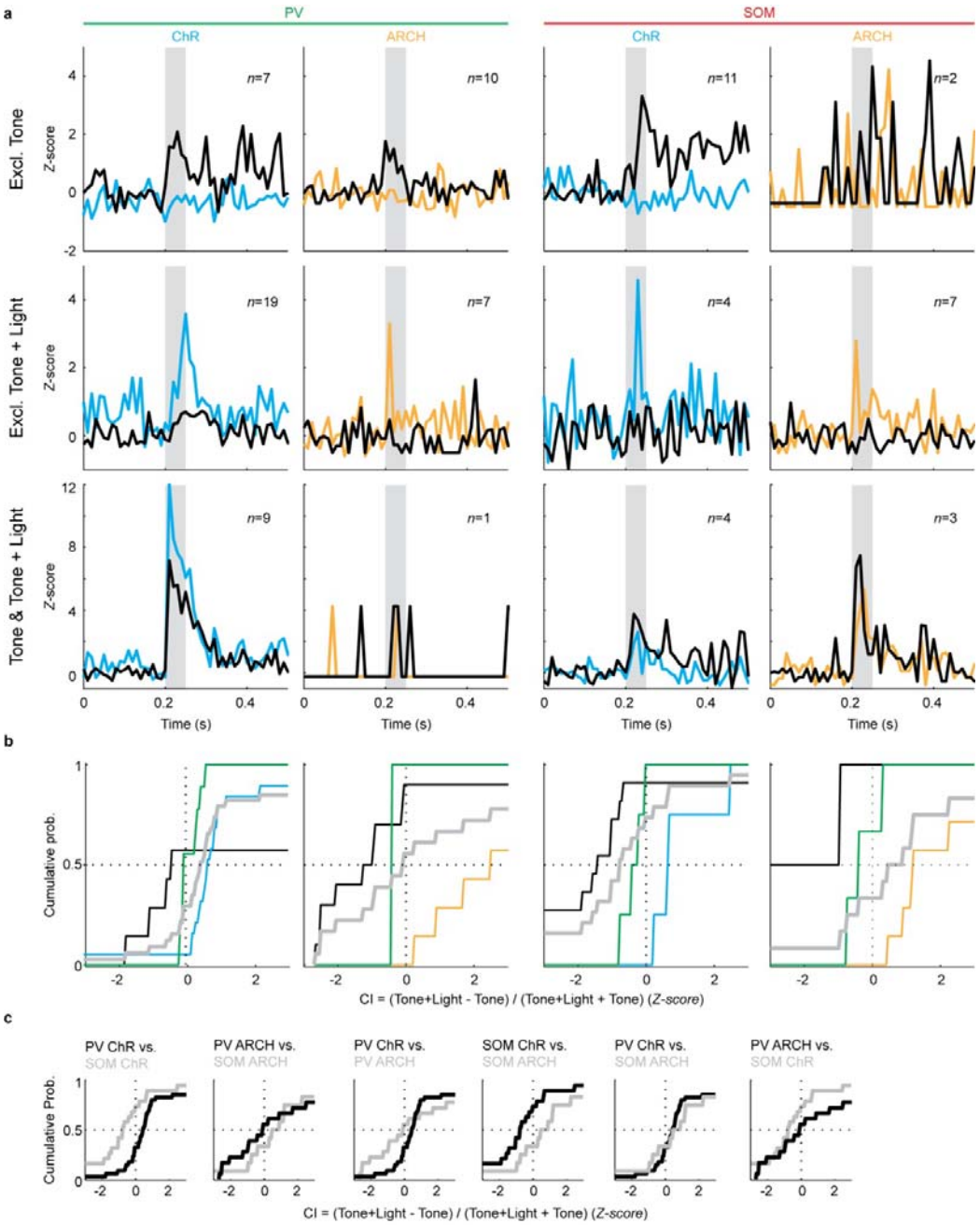
Extended Fig. 5: Effects of optogenetic manipulations of PV$^+$ IN activity on footshock responses of putative PNs.
a, Footshock-evoked responses in footshock-excited putative PNs with (blue or yellow) and without (black) optogenetic stimulation of PV$^+$ INs. Cells are split based on being excited either exclusively during the footshock alone, exclusively during the footshock+light or during conditions. **b,** Cumulative probability distributions of the average *Z*-score of individual putative PNs during the footshock without (black) and with optogenetic manipulations of PV$^+$ INs. **c,** Cumulative probability distributions of the Change Index (CI) as a measure for the light-induced changes in footshock responses for the different optogenetic IN manipulations. The change index is calculated as

CI=(Z-score(Footshock+Light)–
Z-score(Footshock)) /
Z-score(Footshock+Light)+
Z-score(Footshock))

(all Z-scores are the averages of the shock-response during the entire footshock). Shown are the distributions for the footshock-excited cells (black), for the footshock+light excited cells (blue or yellow), for cells which are excited in both conditions (green) and the weighted sum (grey). **d,** Left, Comparisons of cumulative probability distributions for the differences in average footshock *Z*-score during the entire footshock ((Footshock+Light) – Footshock) in putative PNs for PV$^+$ IN activation (blue) and inhibition (yellow). Right, Comparisons of cumulative probability distributions for the change index of the footshock responses of putative PNs upon optogenetic activation (blue) or inhibition (yellow) of PV$^+$ INs.

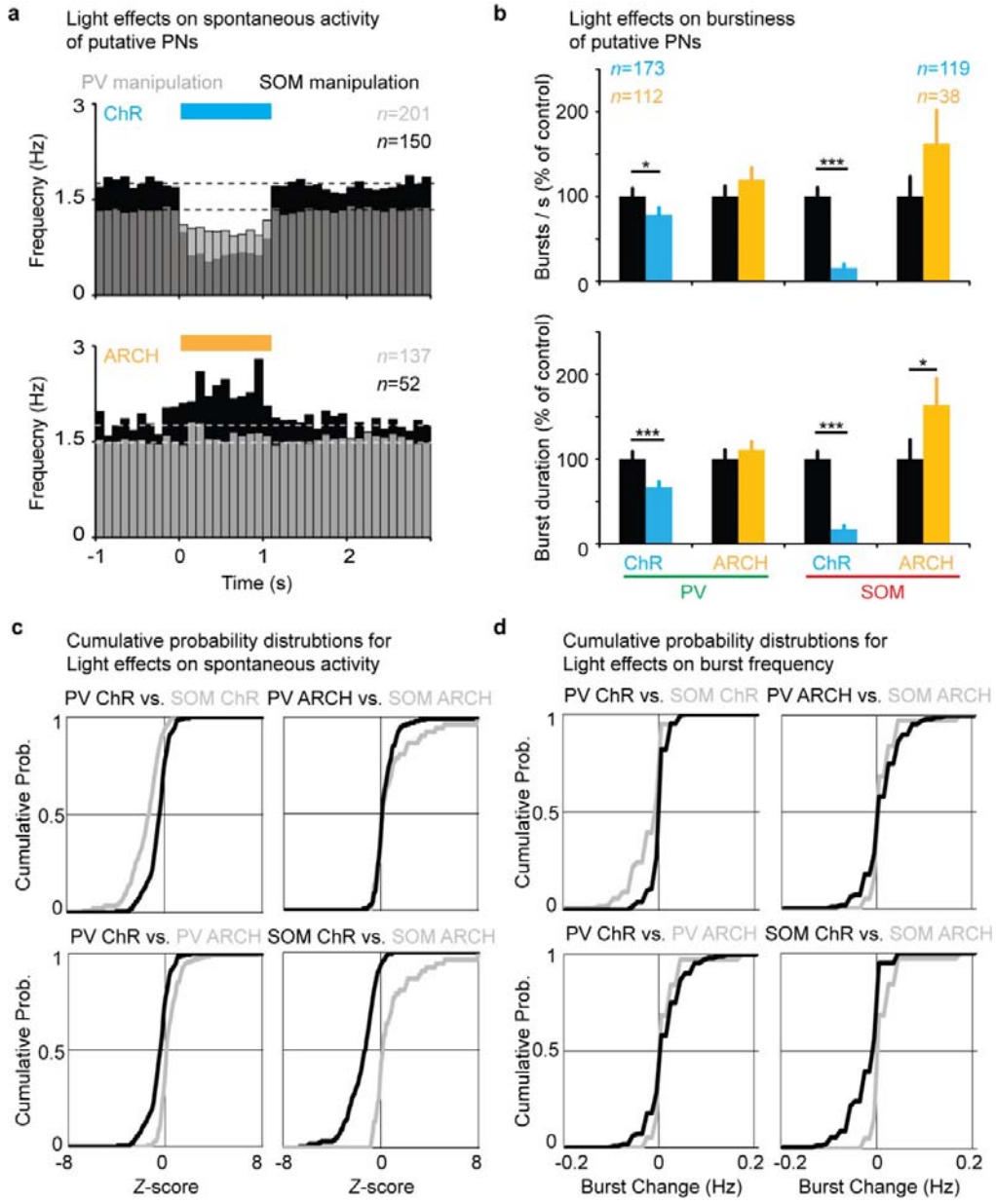


Extended Fig. 6: Inhibitory inputs from PV$^+$ to SOM$^+$ as well as unidentified GABAergic neurons. **a,** Slice patch clamp experiments. Left: BLA location of PV$^+$ interneurons expressing ChR2-mCherry in acute slices of PV-Cre::GAD-EGFP animals. Right: Whole-cell recording (middle) of neuron expressing GAD (top) but not ChR2 (bottom). **b,** Optogenetic stimulation of PV$^+$ neurons resulted in inhibitory postsynaptic currents (IPSCs) in GAD-EGFP$^+$ neurons (n = 26, 5 ms blue light). IPSCs reverse at circa 0 mV (Cl reversal potential, n = 4). **c,** IPSC amplitude, peak latency after blue light onset and 10-90 % rise time of PV inputs on all measured GABAergic (PV-GAD, n = 26) and somatostatin-positive (PV-SOM, n = 3) neurons, respectively. **d,** Distribution of recorded GFP$^+$ cells, which are SOM$^+$ or SOM$^-$, with and without inhibitory inputs upon optogenetic PV$^+$ IN activation.



Extended Fig. 7: Effects of optogenetic manipulations of PV$^+$ or SOM$^+$ IN activity on tone-evoked responses in putative PNs.

a, Tone-evoked responses in tone-excited putative PNs with (blue or yellow) and without (black) optogenetic stimulation of PV$^+$ or SOM$^+$ INs. Cells are split based on being tone excited either exclusively during the tone alone, exclusively during tone+light or in both conditions. **b,** Cumulative probability distributions of the Change Index (CI) as a measure for the light-induced changes in tone responses for the different optogenetic IN manipulations. The change index is calculated as CI=(Z-score(Tone+Light)-Z-score(Tone))/(Z-score(Tone+Light)+Z-score(Tone)) (all Z-scores are the averages of the tone-response during the pip). Shown are the distributions for the tone-excited cells (black), for the tone+light excited cells (blue or yellow), for cells which are excited in both conditions (green) and the weighted sum (grey). **c,** Comparisons of cumulative probability distributions for the change index of the tone responses of putative PNs upon optogenetic stimulation of PV$^+$ or SOM$^+$ cells.



Extended Fig. 8: Effects of optogenetic manipulations of PV$^+$ or SOM$^+$ INs on spontaneous activity and burstiness of putative PNs.

a, Effects of optogenetic activation (top) or inhibition (bottom) of PV$^+$ (grey) or SOM$^+$ (black) cells on the spontaneous activity of putative PNs. **b,** Effects of optogenetic manipulations of PV$^+$ or SOM$^+$ cells on the burst frequency (top) and burst duration (bottom) in putative PNs. **c,** Comparisons of cumulative probability distributions for light-induced changes in spontaneous activity in putative PNs upon optogenetic stimulation of PV$^+$ or SOM$^+$ INs. **d,** Comparisons of cumulative probability distributions for light-induced changes in burst rate in putative PNs upon optogenetic stimulation of PV$^+$ or SOM$^+$ INs. Values are mean +/- s.e.m. **P*<0.05, ****P*<0.001. Statistical analysis in Methods.

	Control	PV					SOM	
	B during CS+US	B during CS+US	B during US	Y during US	B during CS	Y during CS	B during CS	Y during CS
Baseline	20.4 +/- 3.8	12.8 +/- 3.7	14.2 +/- 3.0	29.9 +/- 7.7	15.4 +/- 4.3	19.6 +/- 4.9	16.9 +/- 4.6	15.7 +/- 5.0
CS1 (no light)	68.9 +/- 5.3	72.8 +/- 4.4	51.6 +/- 4.2	51.0 +/- 4.9	46.2 +/- 7.5	69.3 +/- 8.0	61.6 +/- 9.6	41.3 +/- 11.2
CS2 (light)	66.3 +/- 9.9	44.7 +/- 6.6	36.2 +/- 5.7	65.9 +/- 6.3	63.5 +/- 4.9	48.3 +/- 9.3	29.0 +/- 10.2	64.8 +/- 7.7
n	8	11	11	9	10	8	8	7

Extended Table 1: Freezing values for all behavioural groups.
Average freezing values in % of total time for animals in all groups in which optogenetic manipulations were performed during behavioural training. Responses are shown for both CSs. B = blue light stimulation, Y = yellow light stimulation. Control: GFP infected only. All other groups were expressing effective opsins. Baseline: Baseline freezing during 2 min before presentation of the first CS.

METHODS

Animals. Male PV-Cre (Hippenmeyer et al. 2005) or SOM-Cre (Taniguchi et al. 2011) mice (2 – 4 months) were individually housed in a 12 h light/dark cycle while the experiments were running. Before behavioural experiments, animals were habituated to the experimenter by handling for ≥3 times. Food and water was available *ad libitum*. Behavioural experiments were performed during the light cycle. All animal procedures were performed in accordance with institutional guidelines and were approved by the Veterinary Department of the Canton of Basel-Stadt.

Virus injection and optogenetics. Mice were anesthetized with isoflurane (Attane, Provet; induction 5%, maintenance 1.5%) in oxygen-enriched air (Oxymat 3, Weinmann) and fixed in a stereotactic frame (Kopf Instruments). Local injections of ropivacain (Naropin, AstraZeneca) under the scalp and systemic injections of meloxicam (60 µl of 0.5 mg/ml, intraperitoneal, Metacam, Boehringer Ingelheim) were provided for analgesia. A feedback-controlled heating pad (FHC) assured maintenance of the core body temperature at 36.5 °C.

For selective expression of opsins, conditional Cre-dependent recombinant adeno-associated viruses (rAAV) were injected into the basolateral amygdala (BLA) of PV-Cre or SOM-Cre mice, respectively. For expression of the excitatory channelrhodopsin-2 (Zhang et al. 2007) we used a custom rAAV 2/7 EF1a::DIO-ChR2(H134R)-2A-NpHR-2A-Venus (Tang et al. 2009; Letzkus et al. 2011) or rAAV 2/5 EF1a::DIO-ChR2(H134R)mCherry and for expression of the inhibitory proton pump ARCH (Chow et al. 2010), we used an rAAV 2/5 CBA::flex-ARCH-GFP (all Vector Core, University of Pennsylvania). rAAVs (~0.5 µl per hemisphere) were delivered into the BLA using glass pipettes (tip diameter 10-20 µm) connected to a picospritzer (Parker Hannifin Corporation) at the following coordinates: 1.7 mm posterior to bregma, ±3.3 mm lateral to the midline and 4.1 - 4.2 mm below the cortical surface. For optogenetic manipulations during behaviour without electrophysiological recordings, animals were additionally implanted with bilateral custom-built optic fibre connectors (fibre: 0.48 numerical aperture, 200 µm diameter, Thorlabs). Fibre tips were lowered to 300 µm above the BLA at the injection site (**Ext. Data Fig. 1a**). Implants were fixed to the skull with skull screws, cyanoacrylate glue (Ultra Gel, Henkel) and dental cement (Paladur, Heraeus). After surgery, animals were allowed to recover for at least 4 weeks before behavioural training to ensure sufficient expression of the opsins. For optogenetic manipulations during behaviour, the implanted fibres were connected to a custom-built laser bench using an AOTF (AA Opto-Electronic) to control laser intensity (lasers: MBL473, 473 nm wavelength and MGL593.5, 593.5 nm wavelength, CNI Lasers). To ensure that animals could move freely, the connecting fibres were suspended over the behavioural context.

Optrode recordings and optogenetic identification of interneurons. In the majority of animals one optical connector was replaced by a custom-built optrode to allow

simultaneous optogenetic stimulations and single unit recordings during behaviour. Optrodes consisted of an optic fibre connector with an electrode attached directly to the fibre, with the tip protruding approximately 300-500 µm beyond the fibre. Electrodes were made of 16 individually insulated, gold-plated nichrome wires (13 µm inner diameter, impedance 30 to 100 kΩ, Sandvik), attached to a connector (18 pin, Omnetics) (Letzkus et al. 2011). The whole implant was fixed to the skull as described above. Single unit recordings were performed as described below.

For optogenetic identification of PV^+ or SOM^+ INs, we tested all recorded units in all animals for light responsiveness (Lima et al. 2009). We presented pulses of either blue or yellow light to activate ChR2 or ARCH, respectively (300 ms, 60 pulses, 2 s inter pulse interval, 15 – 20 mW at fibre tip). Units were considered as light responsive if they showed significant, time-locked (< 7 ms) changes in neuronal activity upon illumination (**Figure 2c; Ext. Data Fig. 2h**). The onset of optogenetic excitation was determined as the average latency of the first light-evoked spike in each trial. To determine the onset of inhibition reliably in neurons with low baseline firing rates, we applied a change-point analysis (Gallistel et al. 2004) (Change Point Analyzer 2.0, Taylor Enterprises Inc.). Change point analysis identifies the time-point exhibiting a significant change in neuronal activity relative to the preceding time-points. Change points are graphically represented by a change in the slope of a plot showing the cumulative sums of the firing frequency averaged over the 60 light trials (**Ext. Data. Fig. 2e**).

To ensure that light-evoked and spontaneous spikes of identified interneurons were indeed originating from the same neuron, we calculated linear correlation (r) values for spontaneous and light-evoked spikes to quantitatively determine the similarity of their waveform shapes. Only groups of spikes with $r > 0.95$ were considered as originating from the same cell (**Ext. Data Fig. 2d,i**)

Behaviour and optogenetic manipulation of behaviour. Before behavioural training, light-responsive interneurons were identified in opsin expressing PV-Cre or SOM-Cre animals, as described above (**Fig. 2b-d; Ext. Data Fig. 2**). Depending on the presence of identified interneurons, mice were submitted to different auditory fear conditioning paradigms. To determine the physiological activity of identified interneurons during fear learning, we used a discriminative fear conditioning paradigm without optogenetic manipulations (Herry et al. 2008). On training day 1 (Habituation), animals were presented with 4 repetitions of a CS+ and a CS^- (total CS duration of 30 s, consisting of 50 ms pips repeated at 0.9 Hz, 2 ms rise and fall; pip frequency: 7.5 kHz or white noise (WN), counterbalanced for randomization, 75 dB sound pressure level). On day 2 (Fear Conditioning) animals were conditioned by pairing the CS+ with the US (1 s footshock, 0.75 mA DC, 5 CS/US pairings; inter-trial interval: 20 - 180 s; US applied at the time of next expected pip occurrence, ~1.1 s after last pip). The CS^- was intermingled with the CS+, but never reinforced by a US. During Fear Retrieval on the next day, 4 repetitions of first the CS^- and then of the CS+ were presented to test for fear learning. Single unit activity was recorded during all behavioural sessions. Light-

responsiveness of the identified units was tested before habituation and after the last behavioural session.

Animals without identified interneurons were subjected to a fear conditioning paradigm combined with optogenetic manipulations. In animals with implanted optrodes, single unit activity of non-light responsive units was recorded during all behavioural sessions. On day 1 (Habituation), two different CSs (CS1 and CS2) were presented 4 times to the animal - (total CS duration of 10 s, consisting of 50 ms pips repeated at 0.9 Hz, 2 ms rise and fall; pip frequency: 3 kHz or 12 kHz, counterbalanced for randomization, 75 dB sound pressure level). In addition, animals were presented with 4 repetitions of the light stimuli, which were used in the subsequent conditioning, without a CS. On day 2 (Fear Conditioning), animals were fear conditioned by pairing both CS1 and CS2 with a US (1 s foot shock, 0.75 mA DC, 1 CS/US pairing each, interval between CSs: 180s; US applied at the time of next expected pip occurrence, ~1.1 s after last pip). While one of the two CSs served as an internal control, the other CS was paired with light stimulation. Light was applied either only during the CS (blue light: 300 ms pulses starting 50 ms before each pip; yellow light: continuous illumination for 10 s starting with the onset of first CS pip), only during the US (blue or yellow light: 1 s continuous illumination, coinciding with the US), or during CS and US (combination of the described stimulation patterns). The order of CS1 and CS2 and the pairing with light stimulation were counterbalanced within behavioural groups. On day 3 (Retrieval), animals were tested for fear memory retrieval. Both CS1 and CS2 were presented 4 times each without light stimulation or reinforcement. Freezing induced by the control CS and the light-paired CS was compared to determine the effect of the optogenetic manipulation in each animal (**Ext. Data Table 1**). The order of CS1 and CS2 during the test was counterbalanced within behavioural groups. Since the behavioural effects of optogenetic manipulations could not be pre-specified before the experiments, we chose sample sizes with enough repetitions to allow for identification of outliers and for validation of experimental reproducibility (n = 7-11).

To test for the effects of optogenetic manipulations of interneuron activity on footshock- or tone-induced activity in the BLA network, we performed additional behavioural sessions in animals with implanted optrodes. We analysed footshock and tone responses in unidentified cells. This population consists mostly of principal cells with a minority of interneurons (McDonald 1984; Sah et al. 2003; Likhtik et al. 2006; Pape and Pare 2010; Fishell and Rudy 2011). To exclude fast-spiking interneurons, we restricted the analysis to cells with a baseline firing rate < 3 Hz. To determine effects on footshock-induced activity, we delivered 20 footshocks (1 s, 0.75 mA DC) to the animals, alternately with and without coincident stimulation with blue or yellow light. To study the effects of interneuron activity manipulations on tone processing, we used 8 presentations of tones (White Noise, 30 s total duration, consisting of 50 ms pips repeated at 0.9 Hz, 75 dB sound pressure level) alternately with and without coincident light stimulation. For activation of ChR2, we used a train of blue light stimuli (300 ms pulses starting 50 ms before each pip) and for ARCH we

used continuous illumination with yellow light (30 s), as for the manipulations during fear learning (see above). For the analysis of the light-induced changes in shock or tone responses see the corresponding sections below.

Fear conditioning was performed in a context (context A) different from that used for all other behavioural sessions (context B). Context A was cleaned with 70 % ethanol and context B with 1 % acetic acid. Freezing behaviour was quantified using an automatic infrared beam detection system (Coulbourn Instruments) as described previously (Herry et al. 2008). The animals were considered freezing if no movement was detected for 2 s.

***In vivo* electrophysiology and analysis.** Single-unit spike sorting was performed using Off-Line Spike Sorter (OFSS, Plexon) as previously described (Herry et al. 2008). In summary, principal component scores were calculated for unsorted waveforms and plotted on three-dimensional principal component spaces. Clusters containing similar waveforms were manually defined (**Ext. Data Fig. 1b-e**). A group of waveforms was considered to be originating from a single neuron if it defined a discrete cluster in principal component space that was distinct from clusters for other units (determined by MANOVA) and if it displayed a clear refractory period (> 1 ms) in the auto-correlogram. In addition, we used two parameters to quantify the overall separation between identified clusters in an individual channel. The J3 statistic corresponds to the ratio of between-cluster to within-cluster scatter, and the Davies–Bouldin validity index (DB) reflects the ratio of the sum of within-cluster scatter to between-cluster separation. To obtain control values for these statistics, two clusters from the central cloud of points in the principal component space from channels without detectable units were artificially defined. High values for the J3 and low values for the DB compared to the control values indicate good single-unit isolation (**Ext. Data Fig. 1c**). Template waveforms were then calculated for well-separated clusters and stored for further analysis. Clusters of identified neurons were analysed offline for each recording session using principal component analysis and a template-matching algorithm. Only stable clusters of single units recorded over the time course of the entire behavioural training were considered. Long-term single-unit stability was evaluated using Wavetracker software (Plexon) to calculate principal component space-cylinders from a 2 min recording of spontaneous activity before each training session. Straight cylinders suggest stability of the unit isolation during all behavioural sessions (**Ext. Data Fig. 1d**). In addition, we determined single unit stability by quantitative evaluation of the similarity of waveform shapes by calculating linear correlation (*r*) values between average waveforms obtained over training days (Herry et al. 2008). As a control, we computed the *r* values from average waveforms of different neurons (**Ext. Data Fig. 1e**).

To exclude multiple recordings of the same neuron on different channels, we computed cross-correlation histograms. If a target neuron presented a peak of activity at the time when the reference neuron fired, only one of the two neurons was considered for further analysis.

Neuronal activity is reported either as frequency or as *Z*-score value. *Z*-score values were calculated by generating a peri-stimulus time histogram (PSTH) of the binned neuronal data and summation of all spikes of all cells in each individual bin. The average baseline count preceding stimulus onset was subtracted from this population spike count. This difference was divided by the baseline standard deviation to obtain the *Z*-score.

To analyse bursting activity we used the Poisson surprise method (Legendy and Salcman 1985). Bursts were compared between baseline activity and illumination periods (blue or yellow light; 1 s illumination and 1 s baseline before light onset). Briefly, this method detects bursts by finding at least 3 spikes with consecutive interspike intervals (ISI) shorter than half the mean ISI, and calculates the "unlikeliness" or "surprise" that a given ISI would be expected if the spike train was generated by a Poisson process. The Poisson surprise method is a rigorous detector of bursts, because it is insensitive to fluctuations in average firing rate. We chose a surprise value of 3, corresponding to a probability of 0.05 of finding a burst of spikes within a random sequence.

Analysis of optogenetic modulation of shock responses in putative PNs. Shock responses and their modulation by optogenetic activation or silencing of PV^+ cells were studied in putative PNs as described above. Analysis was restricted to putative PNs, defined as units whose mean firing rate during a 100 s baseline period was < 3 Hz (Likhtik et al. 2006). A peri-stimulus time histogram for the shock stimulus – $PSTH^S(t)$ – was calculated with 50 ms resolution between t = - 10 s to t = 2 s relative to shock onset for each cell. From this the mean (μ_B^S) and standard deviation (σ_B^S) of the firing rate preceding each shock (t = - 10 s to t = 0 s) were calculated and used to generate the *Z*-score normalised shock-triggered PSTH:

$$Z^S(t) = \frac{PSTH^S(t) - \mu_B^S}{\sigma_B^S}$$

Units were defined as shock responsive (S) if the response in one of the 20 time bins between t = 0 s and t = 1 s exceeded a *Z* value of 3.72 (corresponding to $p<0.001$, corrected for multiple comparisons). Analysis was performed for the shock responses in the presence of light in the same way and all units showing excitatory shock responses in at least one of the two conditions were pooled. Response magnitudes for the S and SL conditions were defined as:

$$R^S = \langle Z^S(t) \rangle_{t=0-1\,s}$$

$$R^{SL} = \langle Z^{SL}(t) \rangle_{t=0-1\,s}$$

correspondingly. To quantify the effect of light on the shock responses, a change index (CI) was calculated for each unit:

$$CI = \frac{R^{SL} - R^{S}}{R^{SL} + R^{S}}$$

For purely excitatory responses, CI values can range between +1 (units that respond to shock only in the presence of light) and -1 (units that respond to shock only in the absence of light). CI values outside the range (-1,1) indicate inhibitory responses to one of the conditions. A value of 0 indicates no effect of light. For both animal groups (PV-ChR2, PV-ARCH), several measures were used to quantify the effect of light on shock responses at the population level. First, the average differences of the responses ($R^{SL} - R^{S}$) were calculated (**Fig. 2g left**). Next, the median CI for the shock responsive units in each group was calculated (**Fig. 2g right**). Finally, the fraction of S only, SL only and S&SL units from all responsive units was also calculated for both animal groups (**Fig. 2h**). To obtain the population activity of all shock-excited units, all spikes of all excited units were summed in each individual bin of the PSTH. From this the mean and standard deviation of the population baseline (t =–10 s to t = 0 s) were determined and used to generate the *Z*-score normalised shock-triggered PSTH for the population of shock-excited units (**Fig. 2f**).

Analysis of optogenetic modulation of tone responses in putative PNs. Tone responses and their modulation by optogenetic activation or silencing of PV^{+} or SOM^{+} cells were studied in putative PNs as described above. Tones were 50 ms white noise bursts presented at a rate of 0.9 Hz. Each block contained 27 tone pips, but only the first 9 were analysed, to match the behavioural paradigm.

Analysis was restricted to putative PNs, defined as units whose mean firing rate during a 100 s baseline period was < 3 Hz (Likhtik et al. 2006). A peri-stimulus time histogram for the tone stimulus – $PSTH^{T}(t)$ – was calculated with 10 ms resolution between t = - 200 ms to t = 300 ms relative to tone onset. The mean (μ_B^T) and standard deviation (σ_B^T) of the firing rate preceding each tone pip (t = - 200 ms to t = 0 ms) were calculated and used to generate *Z*-score normalised tone-triggered PSTH:

$$Z^{T}(t) = \frac{PSTH^{T}(t) - \mu_B^T}{\sigma_B^T}$$

Units were defined as tone responsive (T) if the response in one of the 6 time bins between t = 0 ms and t = 60 ms exceeded a *Z* value of 3.59 (corresponding to $p<0.001$, corrected for multiple comparisons). Analysis was similarly performed for the tone responses in the presence of light. Cells were defined as responsive to tone in the presence of light (TL) in a similar fashion, the only modification being that for ChR2 stimulation, which was non-continuous and consisted of 300 ms light pulses preceding the tone by 50 ms, only the time period between t = -50 ms and t = 0 ms was used for calculating μ_B^{TL} and σ_B^{TL}. This was done to avoid confounding tone responses with firing rate changes induced indirectly by the light, which could change the baseline and as a result increase the standard deviation estimate.

All units showing excitatory tone responses in at least one of the two conditions were pooled. Response magnitudes for the T and TL conditions were defined as:

$$R^T = \langle Z^T(t) \rangle_{t=0-60\ ms}$$

$$R^{TL} = \langle Z^{TL}(t) \rangle_{t=0-60\ ms}$$

correspondingly. To quantify the effect of light on the tone responses, a change index (CI) was calculated for each unit:

$$CI = \frac{R^{TL} - R^T}{R^{TL} + R^T}$$

For purely excitatory responses, CI values can range between +1 (units that respond to tone only in the presence of light) and -1 (units that respond to tone only in the absence of light). CI values outside the range (-1,1) indicate inhibitory responses to one of the conditions. A value of 0 indicates no effect of light. For each animal group (PV-ChR2, PV-ARCH, SOM-ChR2, SOM-ARCH), several measures were used to quantify the effect of light on tone responses at the population level. First, the average differences of the responses ($R^{TL} - R^T$) were calculated (**Fig. 5b left**). Next, the median CI for the tone responsive units in each group was calculated (**Fig. 5b right**). Finally, the fraction of T only, TL only and T&TL units from all responsive units was also calculated for each animal group (**Fig. 5c**). To obtain the population activity of all tone-excited units, all spikes of all excited units were summed in each individual bin of the PSTH. From this the mean and standard deviation of the population baseline (t = - 200 ms to t = 0 ms) were determined and used to generate the *Z*-score normalised tone-triggered PSTH for the population of tone-excited units for each animal group (**Fig. 5a**).

Slice electrophysiology. Coronal brain slices were prepared from 3 – 6 month old PV-Cre::GAD-EGFP or SOM-Cre::GAD-EGFP mice at least 4 weeks after virus injection. Animals were anesthetized with isoflurane and decapitated in ice-cold artificial cerebrospinal fluid (ACSF in mM: NaCl 124, NaH2PO4 1.25, KCl 2.7, NaHCO3 26, ascorbic acid 2.25, CaCl2 2, MgCl2 8.7, glucose 18). 300 µm thick amygdala slices were cut with a sapphire blade (vibrating microtom: Microm HM 650 V), transferred to an interface chamber and incubated for 35 min at 37 °C in ACSF (1.3 mM MgCl2). Slices were next transferred to a recording chamber on an upright microscope (Olympus, BX61W1) and superfused with ACSF (as above, except for 1.3 mM MgCl2, 2.5 mM CaCl2, 10 µM CNQX, 10 µM CPP, 35 ± 1 °C). Interneurons were identified as EGFP-positive, mCherry-negative cells using epifluorescence (Lumen Dynamics, X-Cite 120 Hg-Lamp, Q-Imaging, RETIGA EXi camera, **Ext. Data Fig. 6a**). Electrophysiological recordings from basolateral amygdala interneurons and pyramidal cells were acquired in whole-cell mode using 3-5 MΩ glass pipettes. Internal solutions were as follows: voltage clamp (in mM): CsCl 110, K-Gluconate 30, EGTA 1.1, HEPES 10, CaCl2 0.1,

Mg-ATP 4, Na-GTP 0.3, 0.4% biocytin, QX 314 chloride 4; current clamp (in mM): K-Methylsulfate 106, KCl 40, HEPES 10, Na-Phosphocreatine 20, Mg-ATP 4, Na-GTP 0.3, 0.4% biocytin. ChR2-expressing SOM^+ or PV^+ INs were optogenetically stimulated with a blue light LED (465 nm, Plexon, LED-driver LD-1) or a blue light laser (473 nm, see above) coupled to an optical fibre, which was positioned above the slice (stimulation intensities: 10-20 mW). Excitatory postsynaptic potentials (EPSPs) were elicited with a bipolar stimulation electrode (twisted platinum / iridium wire) placed in the slice medial to the LA to stimulate thalamic afferents. For spike probability experiments, stimulus intensity was adjusted to yield EPSPs close to spike threshold, such that control EPSP-evoked spike probability was less than 1. Current clamp recordings were performed without excitatory neurotransmitter blockers and Vm was set at -60 mV via somatic current injection. Data were acquired with a Multiclamp 700B amplifier, Digidata 1322A A/D converter and pClamp 9 software (all Axon instruments) at 50 kHz and filtered at 4 or 10 kHz in voltage or current-clamp mode, respectively. Series resistance was monitored continuously and experiments were typically discarded when Rs exceeded 30 MΩ. Outside-out patches were pulled at the end of the experiment and cells were fixed for 1 h in 4% paraformaldehyde before immunohistochemistry.

Histological verification of optrode placement and virus expression. After completion of experiments, mice were deeply anesthetized with Avertin (0.3 g/kg) and an electrolytic lesion was made at the electrode tip by applying 0.1 µA for 15 s to two of the electrode wires. Mice were then transcardially perfused with phosphate-buffered saline (PBS) followed by freshly made, ice-cold 4% paraformaldehyde (PFA). Brains were extracted and kept in PFA overnight (o/n) for post-fixation. Coronal, 60 µm thick brain slices were then cut with a vibratome (VT1000 S, Leica) and stored in PBS containing 0.05% sodium azide. To visualize virus expression, standard immunostaining procedures were performed on free-floating brain sections: O/n incubation at 4 °C to chicken anti-GFP antibody (1:1000, catalogue # A10262, Invitrogen), together with mouse anti-NeuN (catalogue # MAB377, Invitrogen), 2 hr incubation with anti-chicken Alexa 488, and anti-mouse Alexa 594 antibodies, as well as 5 min exposure to 4´,6Diamidin-2-phenylindol (DAPI, 1:10000) as a counterstain. For PV and SOM staining, mice were transcardially perfused with 4 % PFA additionally containing 1.5 % picric acid and 0.05 % glutaraldeyde, and post-fixation was omitted. Before application of the primary antibody, slices were incubated 30 min at room temperature with iFX-enhancer (Invitrogen) to achieve a better signal-to-noise ratio. Slices were then incubated at 4 °C for 48 hrs with either rat anti-SOM antibody (1:500, catalogue # MAB354, Millipore) or anti-PV antibody raised in guinea pig (1:500, catalogue # 195004, Synaptic Systems), followed by o/n incubation at 4°C with secondary anti-rat or anti-guinea pig Alexa 647 antibodies (1:1000, both Invitrogen). Slices were mounted on gelatin-coated glass slides, coverslipped and imaged using a stereomicroscope (Leica) with a monochrome camera (Qimaging) and LED (Lumencor) light source. Using the merged epifluorescence images of virus expression and counterstain as well as the reflected light image of the same sections, we determined viral expression as well as placement of the electrode and optical

fibres by assessing the electrolytic and mechanical lesion produced by the electrode tip or fibre, respectively. Mice were included into the analysis if they showed virus expression bilaterally within BLA, the electrode tip was placed inside of the BLA and fibre tip placement was not more than approximately 400 µm away from BLA. Inclusion of animals was done blind to the behavioural group and the behavioural results.

Histological verification of cell identity in patched slices. After completion of *in vitro* electrophysiological experiments, 300 µm thick slices were post-fixed for 1 hour in fresh 4% PFA, then washed and stored in PBS. Staining for GFP, PV and SOM was performed as stated above, and for visualization of the biotin-filled patched cell, streptavidin conjugated to Alexa 405 (Invitrogen) was added together with the secondary antibodies. Staining for the mCherry tag of the injected virus was omitted due to sufficient fluorescence of the native protein. Slices were imaged using a confocal laser scanning microscope (LSM700, LSM710; Zeiss) to determine quadruple co-localization of PV or SOM, respectively, with GFP, mCherry and Biotin.

Statistical Analysis. All distributions passed tests for normality (Kolmogorov-Smirnov) and for equal variance (Levene Median), unless noted differently. Statistical tests were performed for specific experiments as follows.

Figure 1d: Effects of optogenetic PV^+ activation during CS and US on acquisition of fear memory. Comparison of freezing to two different CSs (control and with PV^+ cell manipulation during conditioning) in each mouse (n = 11). Due to a non-normal distribution, the Wilcoxon Signed Rank test was used and revealed a significant difference ($P<0.001$).

Figure 2a: Comparison of differences between CS1 (control) and CS2 (with optogenetic manipulation during conditioning) for PV^+ cell activation (n = 11) or inhibition (n = 9) during the US. Two-tailed, unpaired Student's t-test revealed a significant difference ($P<0.001$). In addition, freezing levels to the two CSs were compared separately for PV^+ cell activation and inhibition. Two-tailed, paired Student's t-tests revealed significant differences ($P<0.05$ for PV^+ cell activation; $P<0.001$ for PV+ cell inhibition).

Figure 2d: Comparisons of spike half-width and spontaneous activity between optogenetically identified PV+ cells (n = 36) and putative PNs from the same recordings (n = 142). Due to a non-normal distribution, the Mann-Whitney Rank Sum test was used and revealed significant differences (for both: $P<0.001$).

Figure 2g: Left: Differences in average Z-scores of putative PNs during the footshock with and without optogenetic activation (n = 56) or inhibition (n = 26) of PV^+ INs. Due to non-normal distributions, the Mann-Whitney Rank Sum test was used and revealed a significant difference between the effects of PV^+ activation and inhibition ($P<0.001$). In addition, the Wilcoxon Signed Rank test was used to determine statistical significance of the differences in each individual group and revealed significant differences between the footshock responses with and without light for both PV^+ IN manipulations (PV+ IN activation $P<0.01$; PV^+ IN inhibition $P<0.05$). Right: No significant differences were found for the median change index in putative PNs for PV^+ IN activity manipulations.

Figure 3a: Comparison of differences between CS1 (control) and CS2 (with optogenetic manipulation during conditioning) for PV^+ cell activation (n = 10) or inhibition (n = 8) during the CS. Two-tailed, unpaired Student's t-test revealed a significant difference (P<0.01). In addition, freezing levels to the two CSs were compared separately for PV^+ cell activation and inhibition. Two-tailed, paired Student's t-tests revealed significant differences (P<0.05 for PV^+ cell activation and inhibition).

Figure 3c: Comparison of differences between CS1 (control) and CS2 (with optogenetic manipulation during conditioning) for SOM^+ cell activation (n = 8) or inhibition (n = 7) during the CS. Two-tailed, unpaired Student's t-test revealed a significant difference (P<0.001). In addition, freezing levels to the two CSs were compared separately for SOM^+ cell activation and inhibition. Two-tailed, paired Student's t-tests revealed significant differences (P<0.01 for SOM^+ cell activation and P<0.05 for SOM+ cell inhibition).

Figure 4g: Comparisons of EPSP size (n = 9 cells), EPSP decay time (n = 9) and of spike probability (n = 6) between electrical stimulation and electrical stimulation with optogenetic activation of SOM^+ cells. Data were normalized to control conditions. Two-tailed, paired Student's t-tests revealed significant differences (P<0.05 for spike probability; P<0.001 for EPSP size and decay time).

Figure 5b: Left, Comparisons of the differences of average tone responses with and without optogenetic manipulations between the different animal groups (PV-ChR2, n = 36 cells; PV-ARCH, n = 18; SOM-ChR2, n = 19; SOM-ARCH, n = 12). One way ANOVA indicated significant main effects between animal groups: $F(3,80) = 3.166$, P<0.05. Post-hoc pairwise Tukey's tests revealed a significant difference between PV-ChR2 and SOM-ChR2 (P<0.05). In addition, two-tailed, paired Student's t-test was used to determine statistical significance of the differences in each individual group and revealed an effect in the SOM-ChR2 group (P<0.05). Right, Comparisons of the median change index for the tone responses with and without light between the different animal groups (as in the left panel). Due to a non-normal distribution, the Kruskal-Wallis One Way ANOVA on Ranks test was used and revealed significant differences between the different animal groups ($H = 12.975$, 3 d.f., P<0.01). Pairwise multiple comparisons (Dunn's method) revealed significant differences between PV-ChR2 and SOM-ChR2 (P<0.05) and between SOM-ChR2 and SOM-ARCH (P<0.05). In addition, the Wilcoxon signed rank test was used to determine statistical significance of the median CIs for the individual groups and revealed effects for PV-ChR2 (P<0.01) and for SOM-ChR2 (P<0.05).

Ext. Data Fig. 1c: Comparisons of the J3 and DB values for single unit cluster quality between optogenetically identified PV^+ cells (n = 36) or SOM^+ cells (n = 20) and random noise clusters in channels without units (n = 58), respectively. Due to a non-normal distribution, the Kruskal-Wallis One Way ANOVA on Ranks test was used and revealed significant differences between the different cell groups, both for the J3 ($H = 86.491$, 2 d.f., P<0.001) and the DB ($H = 86.275$, 2 d.f., P<0.001) values. Pairwise multiple comparisons (Dunn's method) revealed significant differences between PV^+ cells and control and between SOM^+ cells and control for both J3 and DB (all P<0.05).

Ext. Data Fig. 1e: Linear correlations were calculated for optogenetically identified PV^+ cells (n = 36), SOM^+ cells (n = 20) and between waveforms from different and random non-light responsive units (n = 72) over different behavioural sessions. Plotted are the maximum r values.

Ext. Data Fig. 3d: Comparisons of the spike half-width and the spontaneous activity of optogenetically identified PV^+ cells (n = 36), SOM^+ cells (n = 20) and non-light responsive units (n = 142). Due to a non-normal distribution, a Kruskal-Wallis One Way ANOVAs on ranks was used and indicated statistically significant differences between the groups for both parameters (Spike half-width: $H = 26.347$, 2 d.f., $P<0.001$; spontaneous activity: $H = 34.518$, 2 d.f., $P<0.001$). Pairwise multiple comparisons (Dunn's Method) revealed significant differences for both parameters between PV+ and non-light responsive cells, as well as between PV^+ and SOM^+ cells (all $P<0.05$).

Ext. Data Fig. 8b: Comparisons of the burst rate and the burst duration in putative PNs with and without optogenetic manipulation of PV^+ or SOM^+ cells. Bursts were detected using the Poisson surprise method (see above). Data was normalized to control conditions without light. Two-tailed, paired Student's t-tests revealed significant differences (for Burst rate: $P<0.05$ for PVChR2; $P<0.001$ for SOM-ChR2; for Burst duration: $P<0.05$ for SOM-ARCH; $P<0.001$ for PVChR2 and SOM-ChR2).

3.3 Publication III

Long-range connectivity defines behavioral specificity of amygdala neurons

Verena Senn[1,2,3,*], Steffen B.E. Wolff[1,2,*], Cyril Herry[1,4], François Grenier[1], Ingrid Ehrlich[1,5], Jan Gründemann[1], Jonathan P. Fadok[1], Christian Müller[1], Johannes J. Letzkus[1] and Andreas Lüthi[1]

[1]Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland. [2]University of Basel, Switzerland. [3]Present address: Ernst Strüngmann Institute, 60528 Frankfurt, Germany. [4]Present address: INSERM U862, Neurocentre Magendie, 146 Rue Léo-Saignat, 33077 Bordeaux, France. [5]Present address: Hertie Institute for Clinical Brain Research, 72076 Tübingen, Germany. [*]These authors contributed equally to this work.

Neuron (2014) **81**(2):428-37

Memories are acquired and encoded within large-scale neuronal networks spanning different brain areas. The anatomical and functional specificity of such long-range interactions and their role in learning is poorly understood. The amygdala and the medial prefrontal cortex (mPFC) are interconnected brain structures involved in the extinction of conditioned fear. Here, we show that a defined subpopulation of basal amygdala (BA) projection neurons targeting the prelimbic (PL) subdivision of mPFC is active during states of high fear, whereas BA neurons targeting the infralimbic (IL) subdivision are recruited, and exhibit cell-type specific plasticity, during fear extinction. Selective optogenetic manipulation of IL- or PL-projecting BA neurons demonstrated that the balance between these pathways is causally involved in fear extinction. Together, our findings demonstrate that, although intermingled locally, long-range connectivity defines distinct subpopulations of amygdala projection neurons, and indicate that the formation of long-term extinction memories depends on the balance of activity between two defined amygdala-prefrontal pathways.

INTRODUCTION

Pairing an initially neutral stimulus (conditioned stimulus; CS), with an aversive stimulus (unconditioned stimulus; US) leads to the formation of a robust and long-lasting fear memory. Inhibition of conditioned fear responses can be achieved by repeated exposure to the CS in the absence of the US, a process called fear extinction (Sotres-Bayon et al. 2006; Ji and Maren 2007; Myers and Davis 2007; Quirk and Mueller 2008; Herry et al. 2010). Fear and fear extinction memories are thought to be acquired and stored within distinct but interacting neuronal networks, because animals can retrieve and express both conditioned fear and extinction behavior in a context- or internal state-dependent manner (Sotres-Bayon et al. 2006; Ji and Maren 2007; Myers and Davis 2007; Quirk and Mueller 2008; Herry et al. 2010).

The basolateral amygdala (BLA), comprising the lateral (LA), basal (BA), and basomedial (BMA) subnuclei, is involved in the acquisition of both fear and extinction memories (Myers and Davis 2007; Herry et al. 2010; Amano et al. 2011). In addition, the formation of context-dependent long-term extinction memories depends on the hippocampus and on the medial prefrontal cortex (mPFC)(Quirk and Mueller 2008; Amano et al. 2011). The infralimbic (IL) subdivision of the mPFC plays a central role in extinction memory consolidation and retrieval, whereas the adjacent prelimbic (PL) subdivision has been implicated in sustained fear expression and resistance to extinction (Quirk and Mueller 2008; Burgos-Robles et al. 2009; Sierra-Mercado et al. 2011). Given that amygdala neurons projecting to the hippocampus and to the mPFC are predominantly located in BA (Hoover and Vertes 2007), and that the BA contains functionally distinct classes of neurons responding selectively to conditioned or extinguished CSs (i.e. fear- and extinction neurons)(Herry et al. 2008), we sought to investigate the anatomical and functional specificity of BA projection pathways at the level of identified neurons.

RESULTS

To address the question whether BA projection neurons exhibit target specificity and whether neurons projecting to distinct targets are differentially activated during fear conditioning and extinction, we first combined retrograde tracing with immunohistochemical analysis of activity-dependent expression of the immediate early gene product FOS. Mice (n = 41) were injected with retrogradely transported latex beads (retrobeads) into two prominent targets of BA neurons, the mPFC and the ventral hippocampus (vHC). Retrogradely labeled neurons were found to be predominantly located in the magnocellular (anterior) subdivision of the BA (**Figure 1A**). In some animals which were injected with retrobeads of two distinct colors into mPFC and vHC, we found two largely non-overlapping cell populations (co-labeled neurons: 6.0% of HC-projecting neurons; 5.9% of mPFC-projecting neurons; n = 1836 HC-projecting neurons and 1877 mPFC projecting neurons from 17 double-injected animals) retrogradely labeled from these two

brain structures, suggesting that BA neurons project out of the amygdala in a target specific manner.

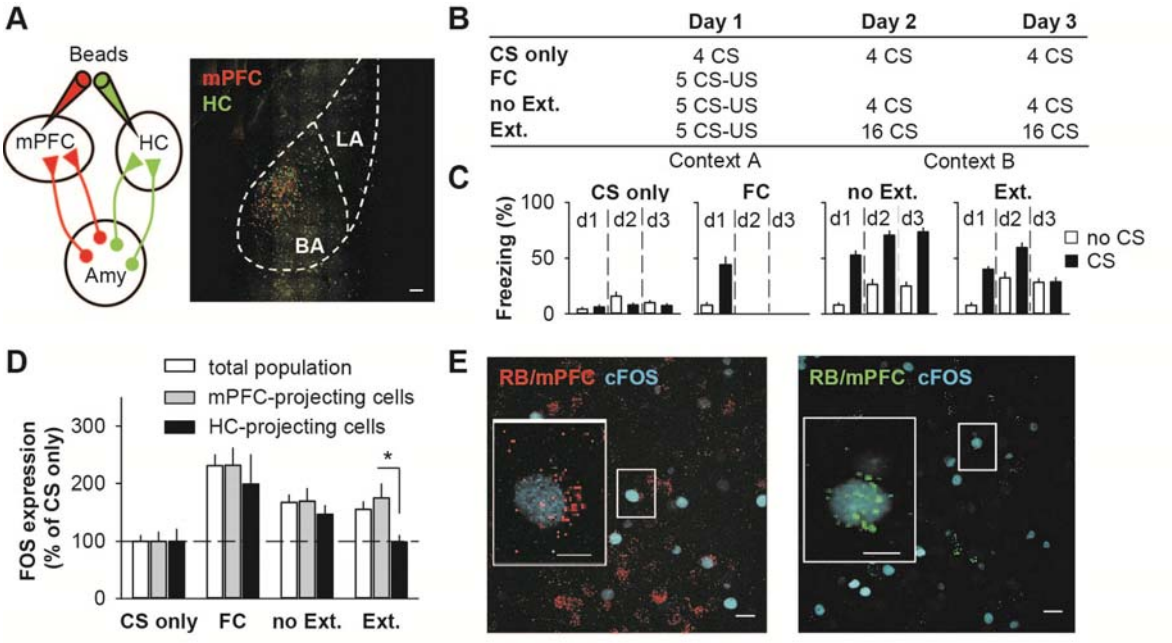


Figure 1. Extinction induces FOS expression in BA neurons projecting to the mPFC.
(A) Experimental scheme and epifluorescent image illustrating backlabeled BA neurons projecting to medial prefrontal cortex (mPFC) or hippocampus (HC). BA, basal amygdala; LA, lateral amygdala; scale bar: 50 µm. (B) Experimental groups: FC, fear conditioning; Ext., extinction; no Ext., no extinction. (C) Behavioral results: When exposed to the CS, the CS-only (n = 21) and extinction (Ext.; *n* = 17) groups exhibit low levels of freezing, whereas freezing levels are high in the fear conditioned (FC; n = 12; $p < 0.0001$ vs. pre-CS) and no extinction (no Ext.; n = 22; $p < 0.0001$ vs. pre-CS) groups. (D) After CS presentations eliciting high fear (FC and no Ext.), the total population of BA neurons expresses high levels of FOS (white bars), similar to the two subpopulations projecting to the mPFC (grey bars) and the HC (black bars). In contrast, extinction training caused FOS induction in the total population and in mPFC-projecting neurons, but not in BA neurons projecting to HC. (Total population: CS only, n_{FOS} = 2530 neurons, n_{NeuN} = 73182; FC, n_{FOS} = 4023, n_{NeuN} = 50340; no Ext., n_{FOS} = 3151, n_{NeuN} = 54328; Ext., n_{FOS} = 2630, n_{NeuN} = 48704; n = 62 mice; mPFC-projecting: CS only, n_{FOS} = 691, n_{beads} = 1589, $n_{overlap}$ = 83; FC, n_{FOS} = 937, n_{beads} = 932, $n_{overlap}$ = 112; no Ext., n_{FOS} = 1321, n_{beads} = 1947, $n_{overlap}$ = 188; Ext., n_{FOS} = 692, n_{beads} = 1166, $n_{overlap}$ = 96; n = 37 mice; HC-projecting: CS only, n_{FOS} = 400, n_{beads} = 1042, $n_{overlap}$ = 49; FC, n_{FOS} = 389, n_{beads} = 373, $n_{overlap}$ = 33; no Ext., n_{FOS} = 830, n_{beads} = 893, $n_{overlap}$ = 59; Ext., n_{FOS} = 332, n_{beads} = 749, $n_{overlap}$ = 33; n = 24 mice). (E) Confocal images illustrating FOS-positive, mPFC projecting (left) and HC projecting (right) BA neurons. Insets illustrate an mPFC- (left) and a HC-projecting (right) FOS-positive BA neuron. Scale bars: 20 µm; inset: 10 µm. * $p < 0.05$.

Five days after retrobead injection, mice were subjected to one of four different behavioral paradigms: the acute fear conditioning group (FC; n = 12) was analyzed immediately after conditioning to investigate immediate FOS expression (**Figures 1B and 1C**). The extinction group (Ext.; n = 17) was subjected to two extinction sessions in a different context, which

completely abolished CS-induced freezing responses (**Figures 1B and 1C**). In addition, two control groups were used to assess the specificity of fear conditioning- and extinction-induced FOS expression. The no extinction group (no Ext.; n = 22) was fear conditioned on day 1, but instead of being subjected to extinction training on days 2 and 3, these animals only received brief CS reminders that did not result in fear extinction. Extinction and no-extinction animals exhibited completely different freezing behavior when exposed to a CS and were analyzed at the same time point on day 3 (2 h after the last CS exposure)(**Figures 1B and 1C**). Finally, a group of animals was exposed to the two contexts and to the same number of CSs as the no-extinction animals, but did not receive any CS-US pairings on day 1 (CS only group; n = 21).

To measure learning-induced FOS expression, animals were sacrificed 2 h after training, when FOS expression is at its peak (Herry and Mons, 2004). Naive animals that were only exposed to their home cage displayed virtually no FOS-expression in the BLA (**Figure S1**). Both fear conditioning and extinction resulted in significant FOS induction as compared to the CS control group in the entire basolateral complex, including the lateral amygdala, BA and basomedial amygdala (**Figure 1D; Figure S1**). Notably, even though freezing behavior was completely different in extinction and no-extinction animals, the percentage of FOS-positive BA neurons was almost identical between the two groups (**Figure 1D**; white bars; **Figure S1**).

To address whether different neuronal populations express FOS after fear conditioning versus extinction in a projection-specific manner, we employed retrograde labeling from mPFC and vHC (**Figure 1E**). Relative to CS control animals, both mPFC-projecting and vHC-projecting BA neurons showed increased FOS expression in the two groups of animals with high fear levels (FC and no Ext.). After extinction, however, FOS expression was elevated exclusively in mPFC-projecting neurons (**Figure 1D; Table S1**). Importantly, we ruled out the possibility that this result was biased by unequal numbers of retrogradely labeled neurons between groups (Figure S2). These findings support the notion that mPFC- and vHC-projecting neurons represent two functionally distinct neuronal subpopulations, but also raise the question why mPFC-projecting neurons exhibit similar levels of FOS expression after fear conditioning and extinction.

One possible explanation is that BA neurons projecting to the mPFC comprise a heterogeneous population of distinct neurons specifically involved in fear conditioning and extinction. Consistent with this notion, previous reports have attributed opposite roles in the expression and extinction of conditioned fear to the prelimbic (PL) and infralimbic (IL) subdivisions of the mPFC (Quirk and Mueller 2008; Burgos-Robles et al. 2009; Sierra-Mercado et al. 2011). In agreement with this hypothesis, analysis of tissue obtained from animals with selective PL- or IL-targeted retrobead injections (**Figures 2A and 2B**) revealed that fear conditioning caused FOS expression predominantly in PL-projecting BA neurons,

whereas extinction led to FOS expression selectively in the IL-projecting population (**Figure 2C; Figures S2 and S3; Table S1**).

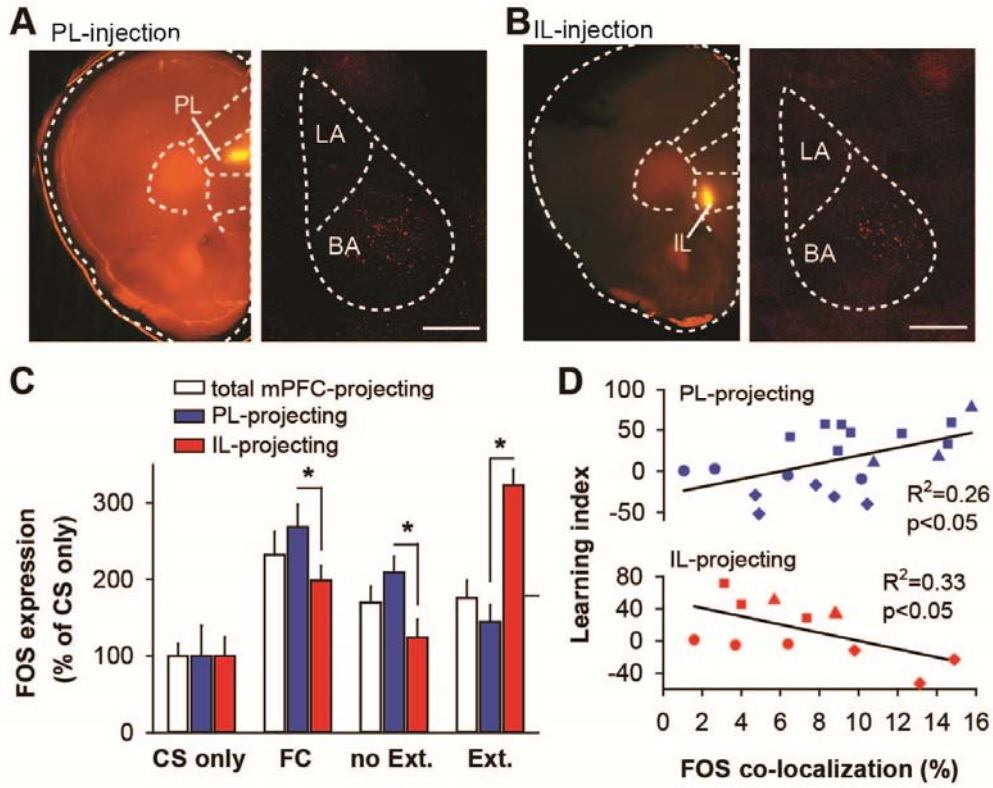


Figure 2. Projection to mPFC subdivision determines activation of BA neurons in either fear or extinction learning.

(A,B) Examples of RB injection sites and location of projection neurons in the BA. (A) Injection in the prelimbic (PL) and (B) infralimbic (IL; subdivisions of the mPFC. Scale bars: 700 µm. PL- and IL-projecting neurons are located in the BA. LA: lateral amygdala; BA: basal amygdala. Scale bars: 700 µm. (C) Whereas the total population of mPFC-projecting BA neurons exhibits high FOS levels in both the no Ext. and Ext. groups, neurons projecting to the IL express FOS specifically after extinction and neurons projecting to the PL express FOS after CS presentations eliciting high fear (FC and no Ext.). (Total population, same data as shown in Fig. 1: CS only, n_{FOS} = 691, n_{beads} = 1589, $n_{overlap}$ = 83; FC, n_{FOS} = 937, n_{beads} = 932, $n_{overlap}$ = 112; no Ext., n_{FOS} = 1321, n_{beads} = 1947, $n_{overlap}$ = 188; Ext. n_{FOS} = 692, n_{beads} = 1166, $n_{overlap}$ = 96; n = 37 mice; IL-projecting, subset of the total population: CS only, n_{FOS} = 119, n_{beads} = 279, $n_{overlap}$ = 9; FC, n_{FOS} = 301, n_{beads} = 373, $n_{overlap}$ = 28; no Ext., n_{FOS} = 326, n_{beads} = 412, $n_{overlap}$ = 22; Ext., n_{FOS} = 286, n_{beads} = 278, $n_{overlap}$ = 33; n = 12 mice; PL-projecting, subset of the total population: CS only, n_{FOS} = 212, n_{beads} = 449, $n_{overlap}$ = 23; FC, n_{FOS} = 463, n_{beads} = 434, $n_{overlap}$ = 62; no Ext., n_{FOS} = 898, n_{beads} = 1414, $n_{overlap}$ = 157; Ext., n_{FOS} = 406, n_{beads} = 888, $n_{overlap}$ = 63; n = 20 mice). (D) Opposite correlations between the magnitude of learning-induced changes in freezing behavior and the percentage of FOS-expressing neurons in PL-projecting neurons (top panel; circles: CS only, n = 4 animals; triangles: FC, n = 3; squares: no-Ext, n = 7; diamonds: Ext., n = 5)), and IL-projecting neurons (bottom panel; circles: CS only, n = 3 animals; triangles: FC, n = 3; squares: no-Ext, n = 3.; diamonds: Ext., n = 3). Changes in freezing values (Δ freezing) were calculated as differences between CS-induced freezing after learning vs. CS-induced freezing before learning (for Ext. and FC groups) or CS-induced freezing vs. pre-CS baseline freezing (for CS-only and no-Ext. groups). * $p < 0.05$.

Notably, when correlating the percentage of FOS-expressing neurons with the magnitude of learning-induced behavioral changes in all animals, we found significant, but opposite, correlations for both PL- and IL-projecting neurons (**Figure 2D**). These complementary changes in FOS expression indicate that acquisition and extinction of fear cause a shift in the balance of activity between two distinct subpopulations of BA projection neurons.

We next addressed the question whether the balance between BA→IL and BA→PL projections could play a causal role in the formation of extinction memories. In order to specifically manipulate the activity of IL- and PL-projecting BA neurons with optogenetics, we used an intersectional approach involving multiple viral vectors. We injected a mixture of two retrogradely transported viruses expressing Cre-recombinase (HSV-1 and CAV-2)(Hnasko et al. 2006; Lima et al. 2009) into either IL or PL (**Figures 3A and 3B; Figures S4 and S5**). Animals were then injected locally into the BA with a conditional adeno-associated viral vector (AAV) co-expressing channelrhodopsin-2 (ChR2) and Halorhodopsin (NpHR) in a Cre-dependent manner (**Figures 3A and 3B**)(Tang et al. 2009), such that opsin expression was restricted to IL- or PL-projecting BA neurons, respectively (**Figure 3B**). In a subset of animals, optrodes for simultaneous single unit recordings and optical stimulation were chronically implanted. Illumination with blue light (to activate ChR2) or with yellow light (to activate NpHR) induced bi-directional changes in activity in 24% of all recorded neurons (including animals with injections not restricted to IL or PL), confirming the validity of the optogenetic approach (**Figure 3C, Figure S13**). Next, we chronically implanted animals bilaterally with optical fibers just above the BA (**Figure S6**) and fear conditioned them using two different tones as CSs. Subsequently, both of these CSs were extinguished. During extinction, one CS was always delivered in combination with blue-light stimulation, whereas the other was presented together with yellow light. Strikingly, animals exhibited weaker long-term extinction memory (reflected by higher freezing levels) for the CS that was extinguished in conjunction with inhibition of IL-projecting BA neurons (yellow light stimulation, **Figure 3D**). When selectively targeting PL-projecting BA neurons, we observed the opposite outcome, that is inhibition of the BA→PL pathway resulted in increased extinction memory (**Figure 3D**). None of the optogenetic manipulations had an effect on within-session extinction (**Figure S7**), indicating that acute within session extinction, and the formation of long-term extinction memories represent independent biological processes. Together, these data provide strong evidence that the balance of activity between the BA→IL and BA→PL pathways is an important factor causally involved in the acquisition of long-term extinction memories.

Consistent with the increase in FOS expression in IL-projecting neurons after extinction and in keeping with the behavioral data, we found one third of the identified IL-projecting neurons to be selectively activated by an extinguished CS (5 out of 14; **Figure S8**). None of the IL-projecting neurons were activated exclusively by a CS eliciting a high fear response (**Figure S8**; $p < 0.05$; Fisher's Exact Test). Conversely, we found that a subpopulation of PL-projecting neurons was selectively activated by a conditioned, non-extinguished CS (2 out of

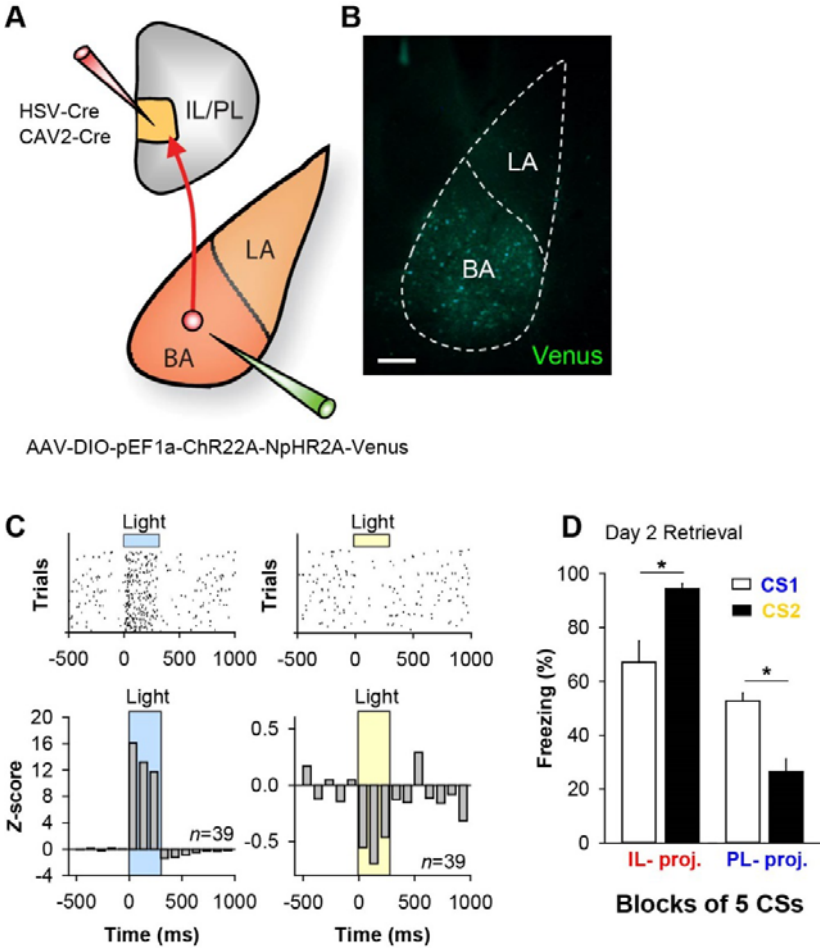


Figure 3. Balance of activity between IL- and PL-projecting BA cells is causally involved in fear extinction.

(A,B) Intersectional approach for specific optogenetic identification and manipulation of IL- or PL-projecting BA cells. A mixture of HSV-1 and CAV-2, both expressing Cre recombinase, was injected into either subdivision of the mPFC to retrogradely infect BA projection neurons. A Cre-dependent AAV was injected into the BA to specifically express ChR2, NpHR and Venus in Cre-expressing BA projection neuron. (C) Raster plots (top) and z-scored population peri-stimulus time histograms (bottom) of opsin-mediated light-responses (ChR2-mediated activation, left; NpHR-mediated inhibition, right) in mPFC-projecting BA cells (n = 39 units from 9 animals), recorded with chronically implanted optrodes for simultaneous single unit recordings and illumination. Light responses show validity of the intersectional optogenetic approach. (D) Optogenetic manipulation of IL- or PL-projecting BA neuron activity oppositely affects long-term extinction memory acquisition. Animals (n = 5 each) expressing ChR2 and NpHR specifically in IL-projecting or PL-projecting BA cells were subjected to fear conditioning and extinction. Two distinct auditory CSs were both paired with foot-shocks. During extinction training the BA was illuminated via chronically implanted optic fibers. CS1 was paired with blue, and CS2 with yellow light (Figure S7). During retrieval on the next day, CSs were presented without light. Statistical analysis of the behavioral outcome of optogenetic manipulations of BA→IL and BA→PL pathways revealed a significant interaction of stimulation light (blue/yellow) x pathway (Two-way ANOVA, $F_{1,16} = 29.97$, $p < 0.001$; post-hoc comparison blue light (IL vs. PL): $p > 0.05$; yellow light (IL vs. PL): $P < 0.05$). *$p < 0.05$.

7) with none of the PL-projecting neurons responding to an extinguished CS (**Figure S8**). Together, this demonstrates that functionally identified “extinction neurons” and “fear neurons” overlap with two distinct anatomically defined subpopulations of BA neurons projecting to IL or PL, respectively.

Given the differential contribution of PL- and IL-projecting BA neurons to fear extinction, we next asked whether these populations also display differential plasticity upon learning. We prepared acute coronal slices from animals injected with retrobeads and obtained whole-cell recordings from 132 retrogradely labeled neurons (**Figure 4A**). Spontaneous excitatory synaptic input and the somatic action potential input-output function of IL- and PL-projecting neurons was not affected by behavioral training (**Figures S3 and S9**), even though PL-projecting neurons exhibited a lower input resistance and were more depolarized in slices obtained from ‘no extinction’ animals compared to cells recorded from animals subjected to extinction (**Table S2**). In contrast, several other biophysical properties of PL- and IL-projecting BA neurons were differentially modified by fear conditioning and extinction (summarized in **Tables S2 and S3**).

First, we found that PL-projecting BA neurons showed increased firing of action potential bursts in response to a depolarizing current injection in the high fear ‘no extinction’ group, but not in the low fear ‘extinction’ group (**Figure 4B**). Strikingly, IL-projecting neurons exhibited opposite changes, with increased bursting selectively after extinction (**Figure 4C**). Importantly, changes in burst firing were also observed *in vivo* when recording from functionally defined fear and extinction neurons in behaving animals (Herry et al. 2008) (**Figures 4B and 4C**). There were no significant changes in the overall spontaneous firing rates of fear and extinction neurons, which could account for changes in burst firing (**Figure S10**). Thus, fear conditioning and extinction resulted in cell type-specific changes in action potential firing dynamics towards more burst firing. Because at many synapses action potential bursts transmit information more reliably than single spikes (Lisman 1997), learning may have changed the impact of PL- and IL-projection neurons on their targets by altering firing dynamics.

In addition, acquisition of extinction caused plastic changes of the action potential waveform exclusively in IL-projecting neurons. One effect was a reduction of the fast afterhyperpolarization (AHP_{fast})(**Figure 4D**). Moreover, IL-projecting neurons displayed a selective increase in action potential half-width with extinction learning (**Figure 4E**). To address whether plasticity of action potential kinetics also occurs *in vivo*, we analyzed extracellular single unit recordings from identified BA fear and extinction neurons (see above). In keeping with the results obtained from *ex vivo* whole-cell recordings, this analysis revealed that the width of extracellularly recorded spikes increased with fear extinction specifically in identified extinction neurons (**Figure 4F**). In contrast to extinction neurons, fear neurons showed no changes in spike waveforms (**Figure 4F**). Thus, converging evidence from slice experiments and recordings in freely moving animals during behavior implicates

plasticity of action potential firing dynamics and action potential waveforms in distinct subpopulations of BA projection neurons in fear conditioning and extinction.

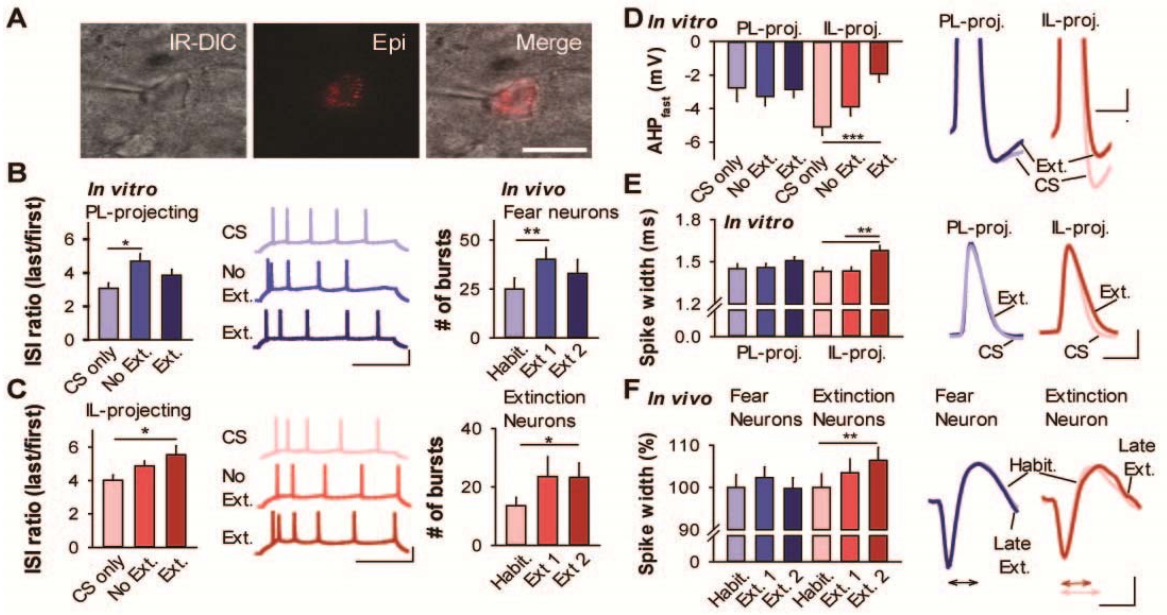


Figure 4. Fear conditioning and extinction induce pathway-specific changes in action potential firing dynamics and action potential kinetics.
(A) Infrared DIC and fluorescent images of a retrobead-labeled BA neuron in an acute brain slice. Scale bar: 20 µm. (B) Burst firing of PL-projecting neurons in response to a depolarizing current injection recorded ex vivo is selectively enhanced in slices of the no Ext. group (left, CS only, n = 16; no Ext., n = 22; Ext., n = 27; scale bars: 100 ms, 40 mV). Similarly, fear neurons recorded in vivo show increased numbers of bursts (right, n = 32 neurons from 15 animals). (C) IL-projecting neurons recorded ex vivo show increased burst firing after extinction (left, CS only, n = 26; no Ext., n = 20; Ext., n = 21; scale bars: 100 ms, 40 mV). Likewise, extinction neurons recorded in vivo exhibit increased numbers of bursts upon extinction (right). (D,E) Reduced AHP_{fast} and broadened spike width in IL-projecting neurons recorded ex vivo after extinction (IL-projecting neurons: CS only, n = 26; no Ext., n = 20; Ext., n = 21; PL-projecting: CS only, n = 16; no Ext., n = 22; Ext. n = 27). Scale bars in (D): 2 ms, 2 mV; in (E): 1 ms, 20 mV. (F) Extinction neurons recorded in vivo exhibit increased spike width after extinction. (n = 26 neurons from 13 animals). Arrows between dashed lines indicate measured spike width. Scale bars: 200 µs, 20 µV. *P < 0.05, **P < 0.01, ***P < 0.001.

DISCUSSION

Together, our findings indicate that acquisition of fear extinction is mediated, at least in part, by a shift in the balance between two separate pathways from the BA to distinct subdivisions of the mPFC. The largely opposing roles of PL and IL in fear and extinction behavior (Quirk and Mueller 2008) thus reflect functional and target specificity of BA principal neurons and suggests that these specific subpopulations of amygdala projection neurons are behaviorally relevant drivers of IL and PL function.

During extinction learning, the balance between the two amygdala-mPFC pathways may be shifted by cell type-specific synaptic and intrinsic plasticity mechanisms. Although we could not find any change in the overall glutamatergic synaptic input onto IL- or PL-projecting BA neurons, this does not exclude the possibility that specific inputs may be altered upon extinction learning. Learning-induced intrinsic plasticity of neuronal excitability has been found in different brain structures (Disterhoft et al. 1986; Saar et al. 1998; Zhang and Linden 2003; Santini et al. 2008). In many cases, changes in currents underlying the AHP eventually resulting in altered neuronal firing dynamics have been described (Disterhoft et al. 1986; Saar et al. 1998; Zhang and Linden 2003; Santini et al. 2008). While these changes often occurred in diverse neuronal populations (Disterhoft et al. 1986; Saar et al. 1998; Santini et al. 2008), we demonstrate here that intrinsic plasticity can be expressed in an exquisitely cell type-specific fashion. In addition, our results provide the first description of learning-induced changes in action potential half-width in a defined neuronal circuit. Mechanistically, changes in the properties or location of voltage- or calcium-dependent ionic conductances could be involved (Faber and Sah 2002; Grubb and Burrone 2010; Kuba et al. 2010)(Figure S11). In principle, changes in action potential kinetics or firing dynamics represent powerful mechanisms by which the impact of neuronal firing on synaptic transmission, or on dendritic integration, could be altered. It remains to be tested whether and how such changes are propagated along the axon and/or into the dendritic tree. Moreover, it will be interesting to identify the precise neuronal targets in IL and PL, and how BA input onto these targets affects function and plasticity of mPFC circuits.

In addition to inputs from the BA, the mPFC acts as a site of convergence integrating inputs from various brain structures including hippocampal inputs conveying contextual information relevant for the expression of extinction memories (Hobin et al. 2006; Sotres-Bayon et al. 2012). The relative strength of fear and extinction memory expression, in turn, may involve differential projections from IL and PL back to various nuclei of the amygdala, to the hypothalamus and/or to downstream structures in the brainstem (Sesack et al. 1989; Vertes 2004; Gabbott et al. 2005; Pinard et al. 2012).

Local connectivity and functional properties of projection neurons in other brain areas can depend on their long-range targets (Morishima and Kawaguchi 2006; Hattox and Nelson 2007; Le Be et al. 2007; Brown and Hestrin 2009; Varga et al. 2010; Lammel et al. 2011). Our present data indicate that target-specificity of long-range glutamatergic projections in the forebrain is likely an important principle underlying specificity at the behavioral level. The fact that PL- and IL-projecting BA neurons are intermingled in a salt and pepper like manner, rather than being anatomically segregated, may help to establish local interactions important for rapid behavioral switching (Herry et al. 2008). Indeed, previous findings indicate that long-range target specificity is associated with exquisite specificity of local excitatory and inhibitory circuits (Brown and Hestrin 2009; Varga et al. 2010). It remains to be determined how defined populations of amygdala projection neurons are wired and integrated into the local circuitry and which cell types they contact in their long-range target

structures (Gabbott et al. 2006; Floresco and Tse 2007). Interestingly, anatomical reconstructions of biocytin-filled PL- and IL-projecting neurons revealed that they likely differ with regard to their local connectivity. Whereas all sufficiently recovered IL-projecting neurons sent collaterals to the central nucleus of the amygdala (CEA), PL-projecting neurons did not target the CEA (**Figure S12**). Eventually, an integrated understanding of functional circuit organization at both the local and global scale will be required to determine how complex behaviors emerge, and how they are modified by experience.

SUPPLEMENTARY MATERIAL

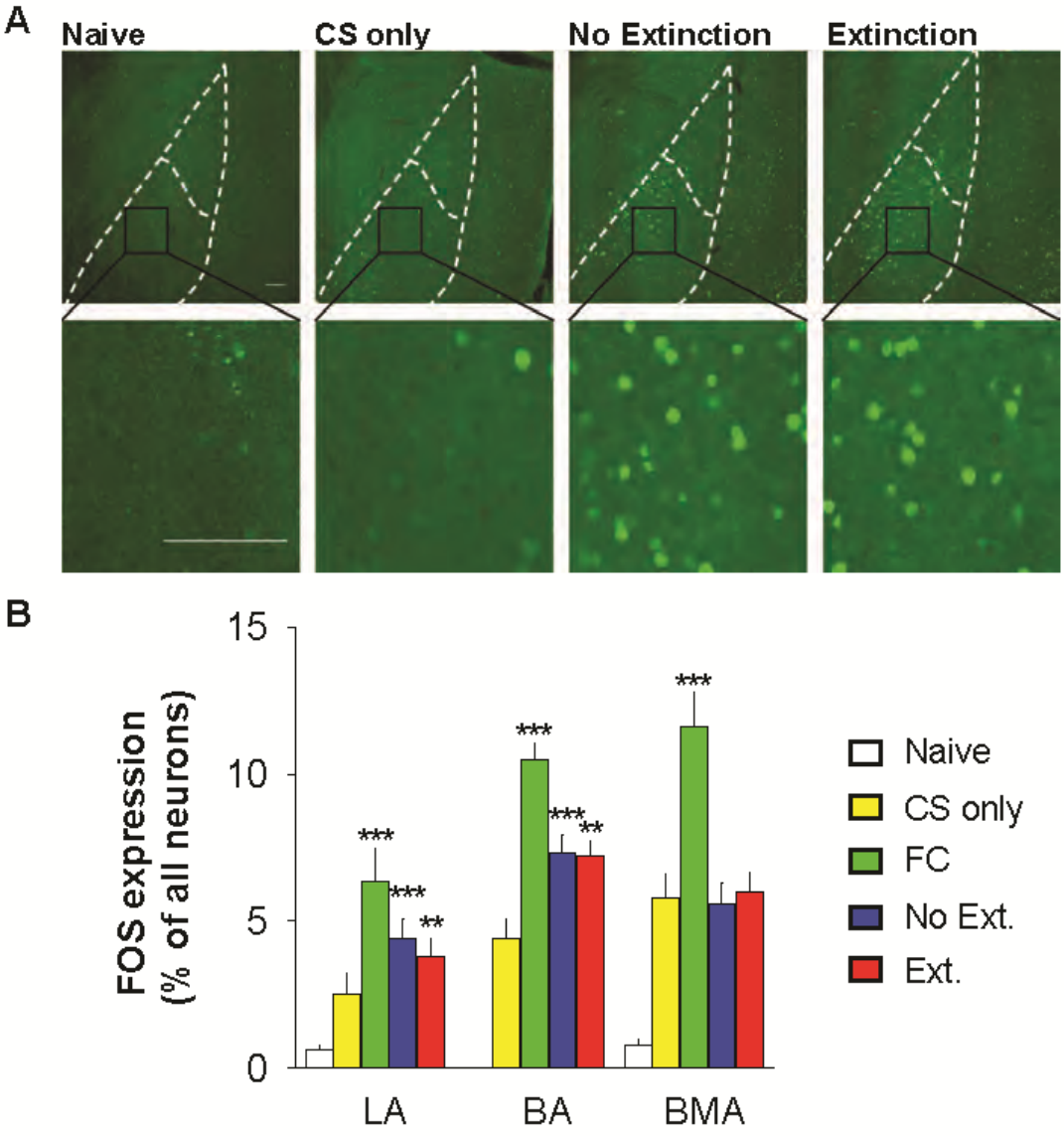


Figure S1. Analysis of FOS expression in different subnuclei of the amygdala.
(A) Example images illustrating FOS immunolabeling in amygdala in different behavioral groups. Scale bars: 50 µm. (B) Quantification of FOS expression in NeuN-labeled neurons in slices obtained from different behavioral groups reveals different patterns in distinct amygdala subnuclei. LA, lateral amygdala; BA, basal amygdala; BMA, basomedial amygdala. **p < 0.01, ***p < 0.001.

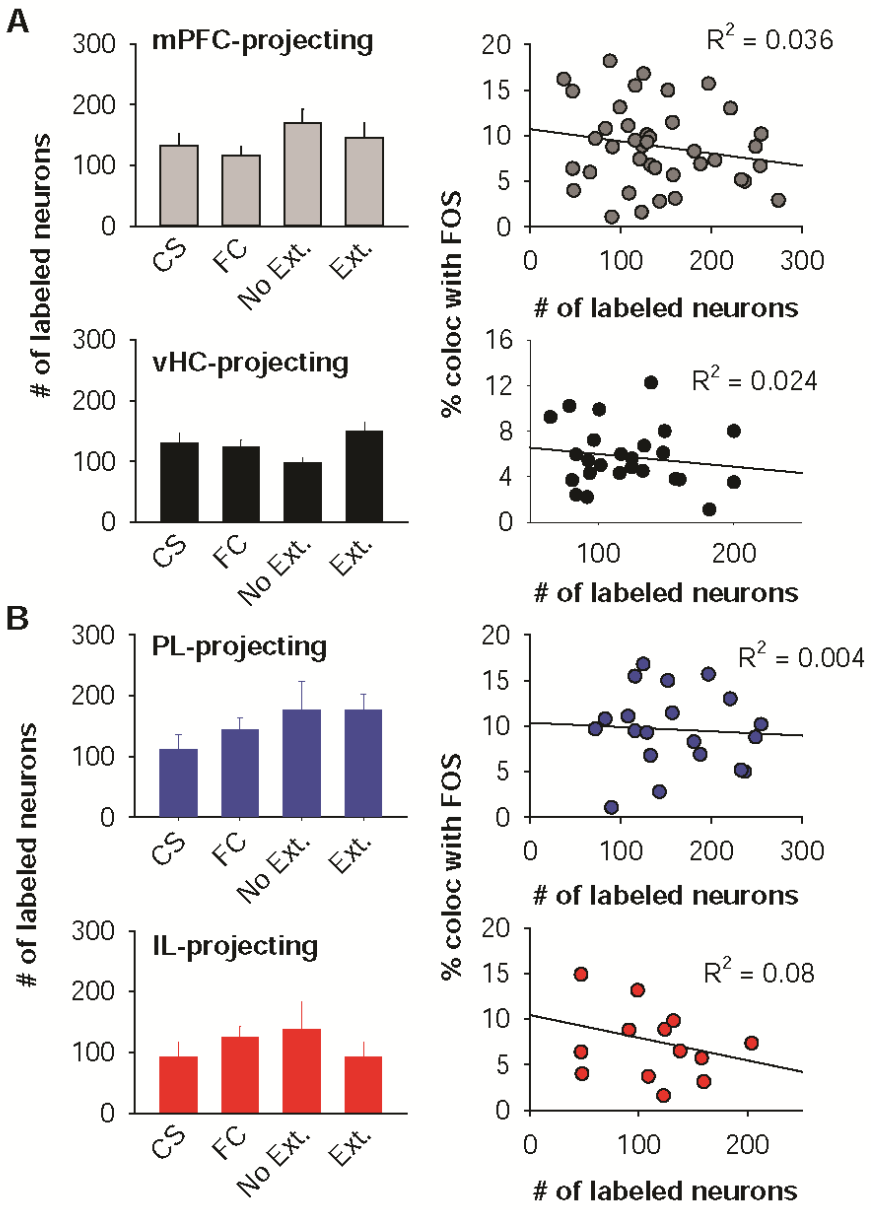


Figure S2. Colocalisation of retrograde tracer and FOS does not depend on the number of retrogradely labeled cells.

Plots show that the number of retrogradely labeled cells does not significantly differ between behavioral groups (left panels; ANOVA, post-hoc Tukey test). No correlation between the number of retrogradely labeled neurons and the number of FOS-positive neurons can be found (right panels; $r_{mPFC} = 0.19$, $p = 0.24$; $r_{HC} = 0.16$, $p = 0.46$; $r_{IL} = 0.27$, $p = 0.40$; $r_{PL} = 0.06$, $p = 0.79$).

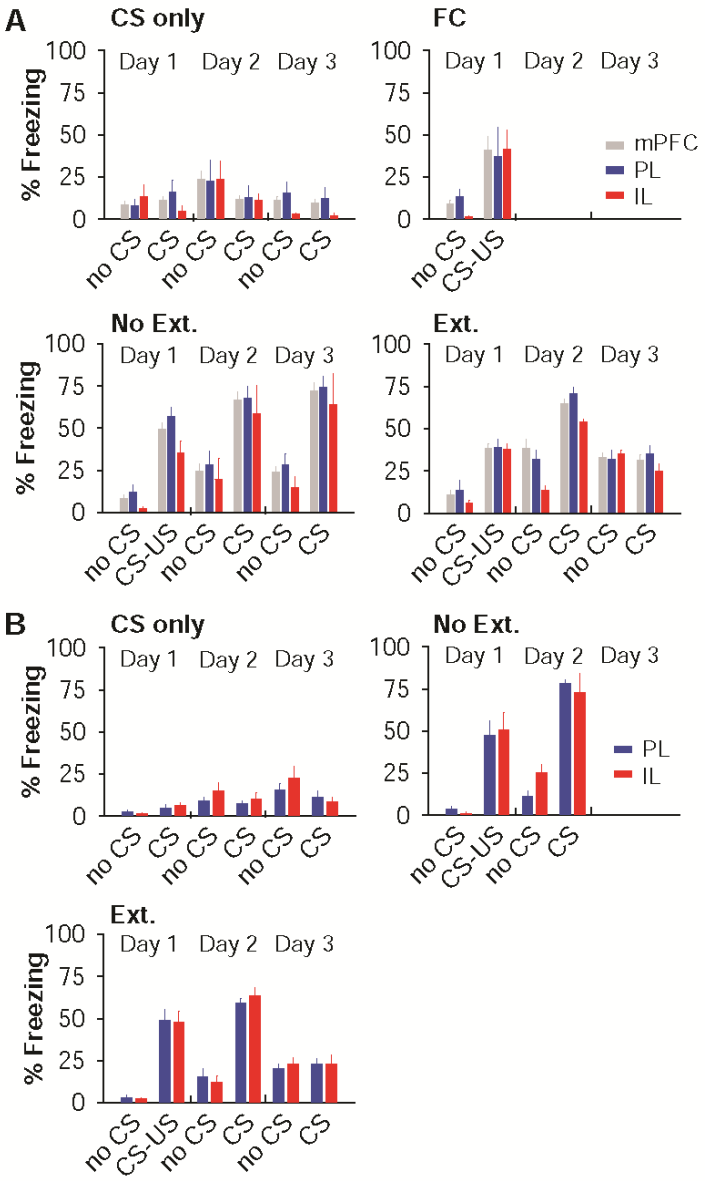


Figure S3. Fear and extinction behaviour in mice used for ex vivo analysis.
(A) Behavior of mice used for ex vivo immunohistochemical analysis shown in Figure 2c (mPFC injected mice: CS-only, n = 10; FC, n = 8; no Ext., n = 11; Ext., n = 8; PL-injected mice: CS-only, n = 4; FC, n = 3; no Ext., n = 8; Ext., n = 5; IL-injected mice: CS-only, n = 3; FC, n = 3; no Ext., n = 3; Ext., n = 3). (B) Behavior of mice used for ex vivo electrophysiological analysis shown in Figure 4 (PL-injected mice: CS-only, n = 6; no Ext., n = 6; Ext., n = 16; IL-injected mice: CS-only, n = 10; no Ext., n = 5; Ext., n = 14). There were no significant differences between mPFC-, PL-, or IL-injected animals (ANOVA, post-hoc Tukey test).

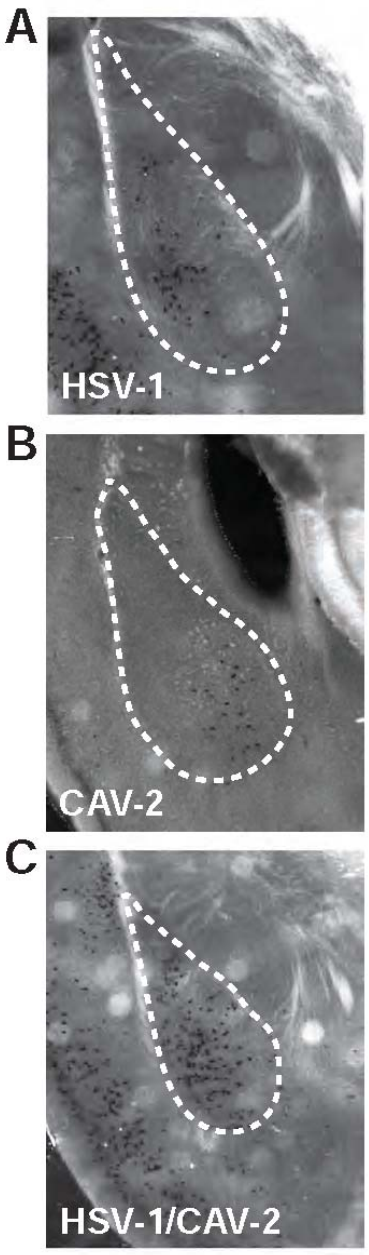


Figure S4. Differential infection patterns of retrograde viruses.
Retrograde viruses expressing Cre recombinase were injected into the IL of LacZ Cre reporter mice. X-Gal staining was used to reveal Cre expression in the BA. While HSV-1 shows preferential infection of a lateral subpopulation of IL-projecting BA cells (A), CAV-2 infected neurons are located more medially (B). (C) Combined injection of HSV-1 and CAV-2 labels a population comparable with the one labeled by retrobeads (c.f. Figures 1A, 2A and 2B).

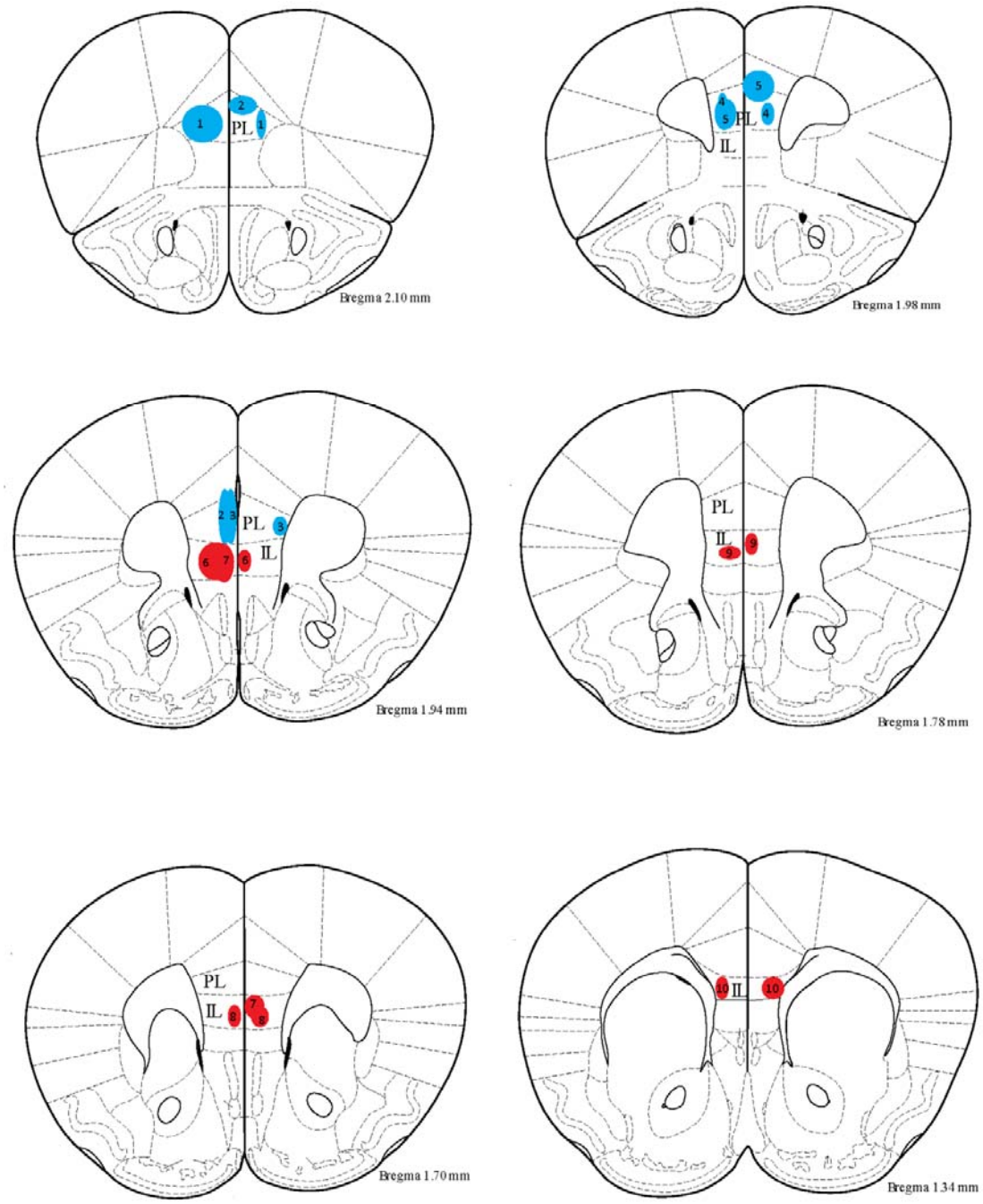


Figure S5. Spread of viral injections into the mPFC.
Spread of the injections of the HSV-CRE/CAV-CRE virus mix into either the PL (blue) or the IL (red) for all animals used for optogenetic manipulations during behavior. Shown is the maximal extent of the injection site for each animal, as determined by the spread of the co-injected blue beads. Injection sites with the same number are from the same animal.

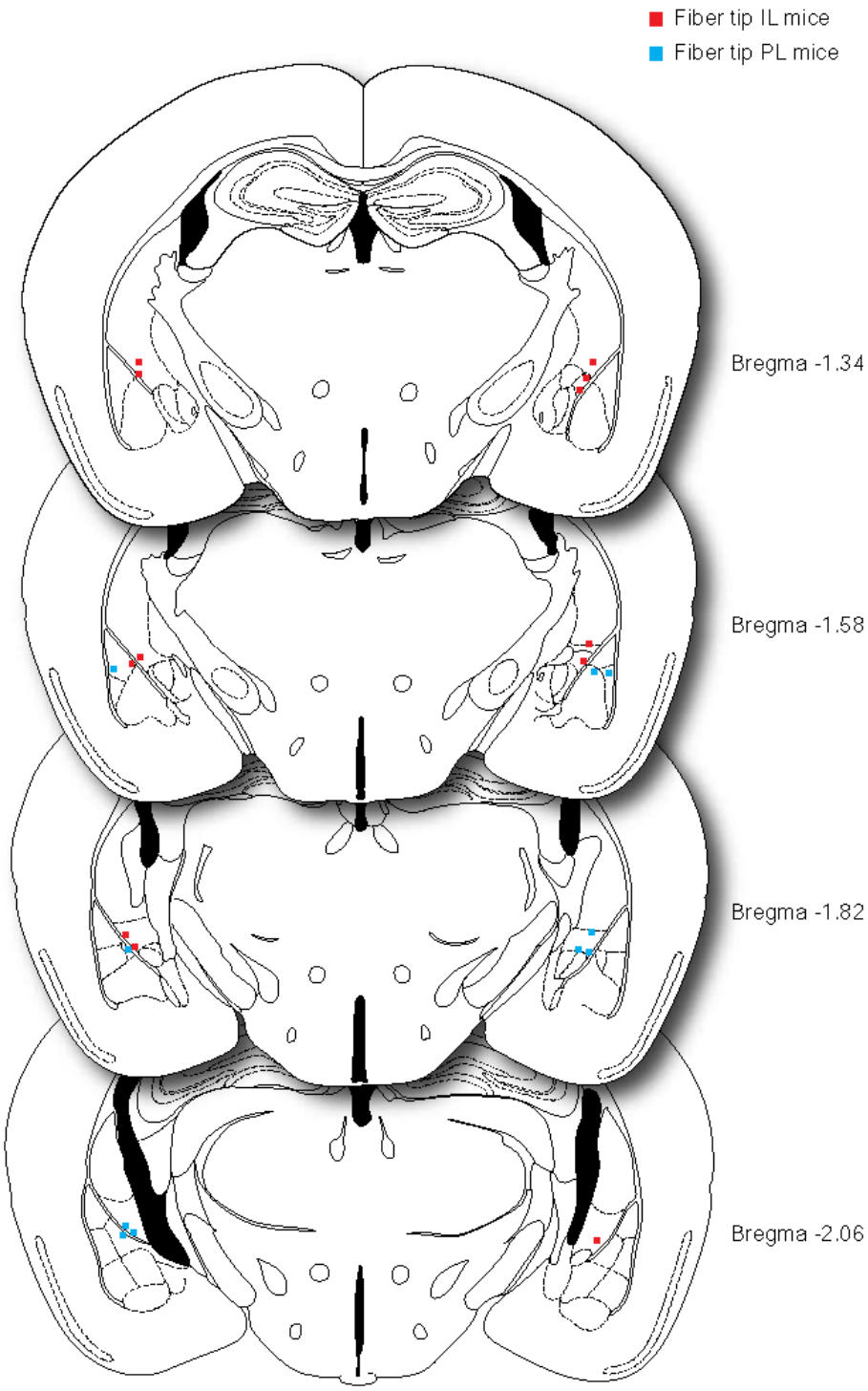


Figure S6. Placement of optical fibers.
Position of bilateral optic fibres for all IL and PL mice used for behavior with optogenetic manipulations.

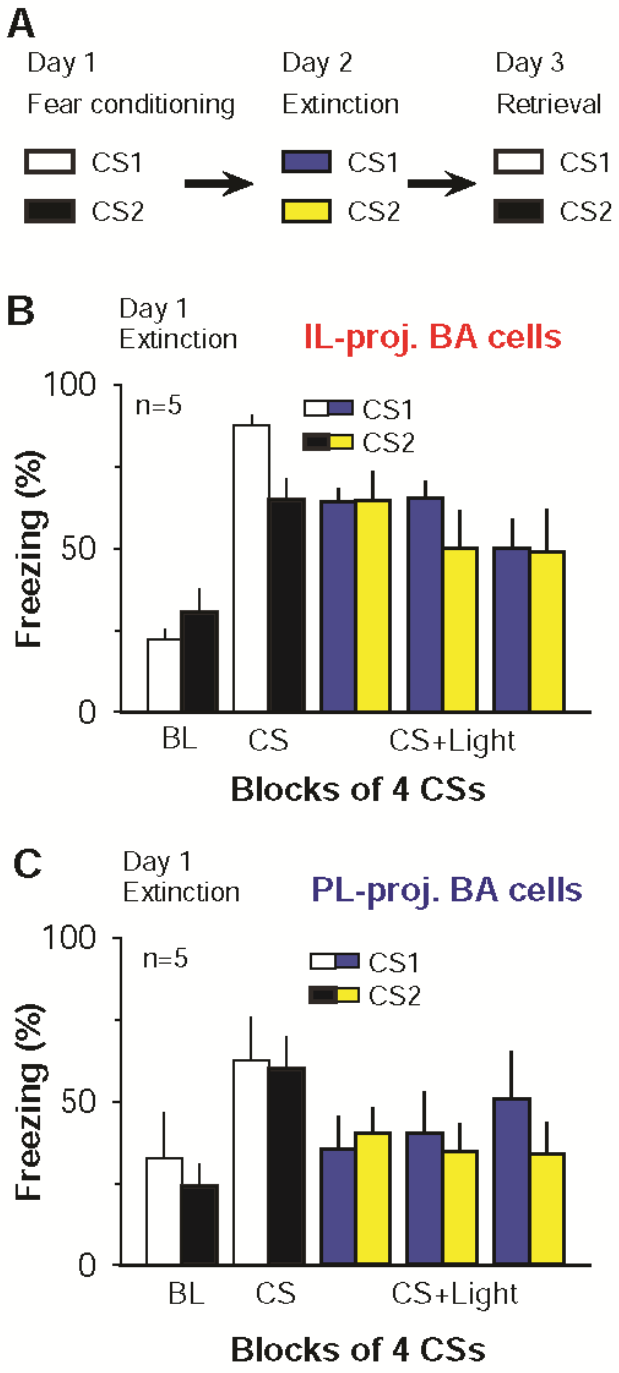


Figure S7. Within-session extinction is not altered.
Optogenetic activation or inhibition (A) of the BA→IL (B) and BA→PL (C) pathways had no significant effect on within-session extinction (*n* = 5 each).

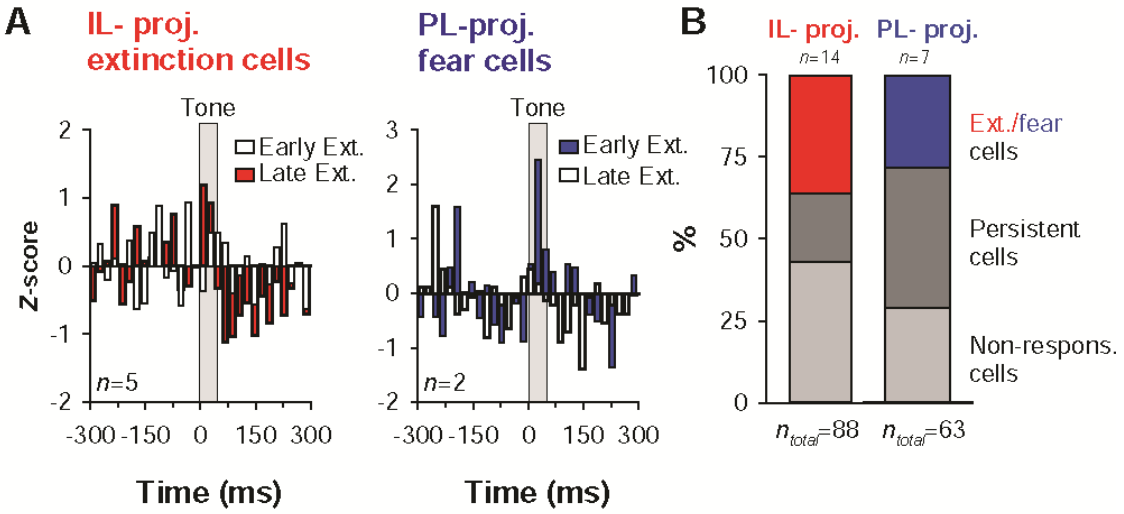


Figure S8. Functional characterization of optogenetically identified IL- and PL-projecting neurons. (A) Chronic optrode recordings allow functional characterization of optogenetically identified IL-projecting and PL-projecting BA cells during discriminative fear conditioning and extinction training. Histograms illustrate responses from IL-projecting extinction neurons (n = 5) and PL-projecting fear neurons (n = 2). Individual units were classified as fear neurons if they exhibited significant responses after FC but not after extinction and vice versa for extinction neurons (see methods). (B) A subpopulation of 36% (5 out of 14 units from 4 animals) of IL-projecting BA neurons shows specific activation to an extinguished CS (extinction cells). Persistent cells show significant responses to both conditioned and extinguished CSs (Herry et al., 2008). A subpopulation of 28% of PL-projecting BA neurons (2 out of 7 units from 2 animals) shows specific activation to a fear conditioned, non-extinguished CS (fear cells). Persistent cells show significant responses to both conditioned and extinguished CSs (Herry et al., 2008).

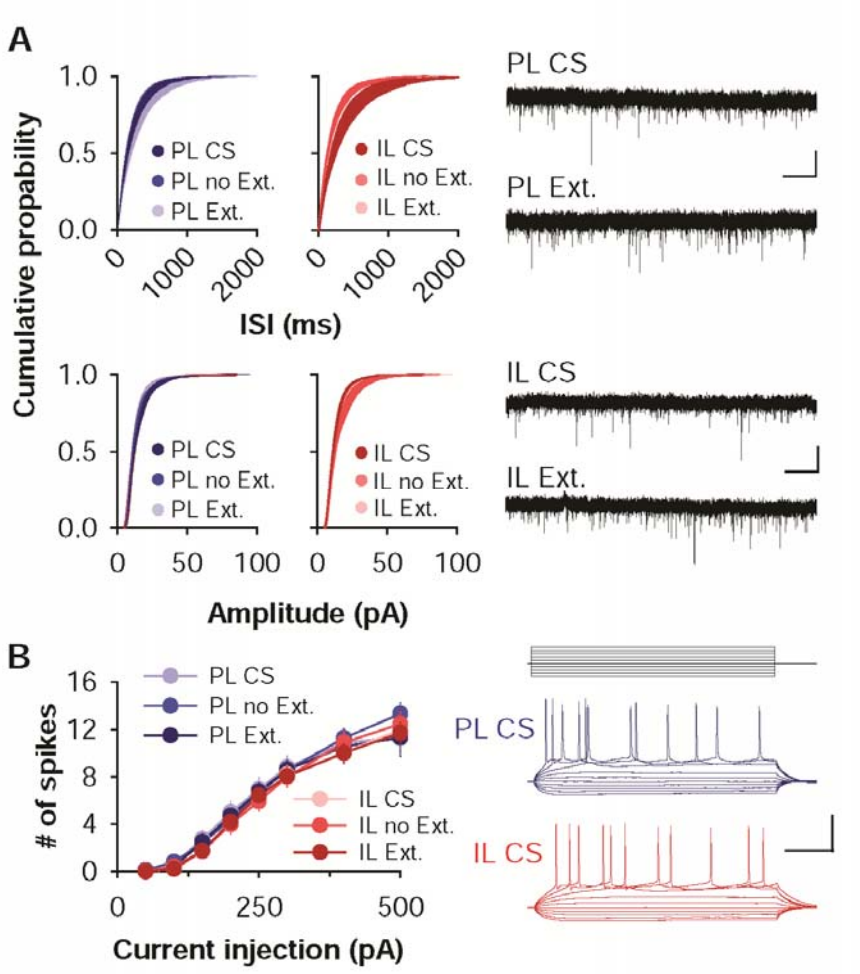


Figure S9. *Ex vivo* analysis of synaptic input and global intrinsic excitability of PL- and IL-projecting BA neurons.

(A) To address whether the overall synaptic drive onto PL- or IL-projecting neurons was altered by fear conditioning or extinction, we analyzed spontaneous excitatory synaptic currents (sEPSCs). This analysis did not reveal any differences in sEPSC amplitude, frequency or kinetics between cell types nor between different behavioral groups (PL-projecting neurons: CS only, n = 14; no Ext., n = 14; Ext., n = 20; IL-projecting neurons: CS only, n = 15; no Ext., n = 13; Ext., n = 10). Because sEPSCs reflect mixed synaptic input from different afferents and local connections, our data do not exclude the possibility that fear conditioning or extinction could have induced input-specific changes in synaptic transmission which we were unable to detect amongst sEPSCs originating from many different inputs. Plots are averages (with horizontal error bars, s.e.m.) of binned cumulative distributions of individual cells. Scale bars: 50 pA, 1 s. (B) Fear conditioning and extinction did not lead to changes in the overall excitability of PL- and IL-projecting neurons as measured by depolarizing somatic current steps (PL-projecting neurons: CS only, n = 16; no Ext., n = 22; Ext., n = 27; IL-projecting neurons: CS only, n = 27; no Ext., n = 20; Ext., n = 21). Scale bar 100 ms, 40 mV.

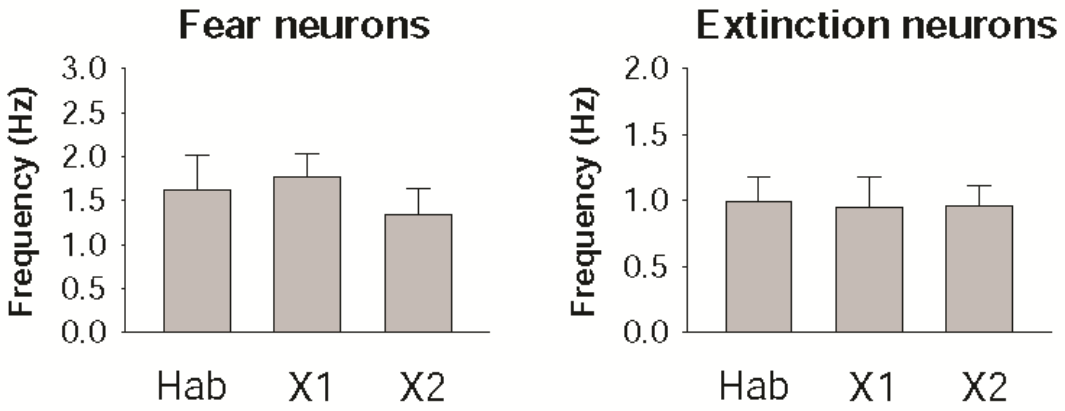


Figure S10. Fear and extinction neurons do not exhibit any changes in spontaneous firing rates across behavioral sessions.
Fear neurons: n = 32; extinction neurons: n = 26; Hab: habituation; X1: early extinction; X2: late extinction.

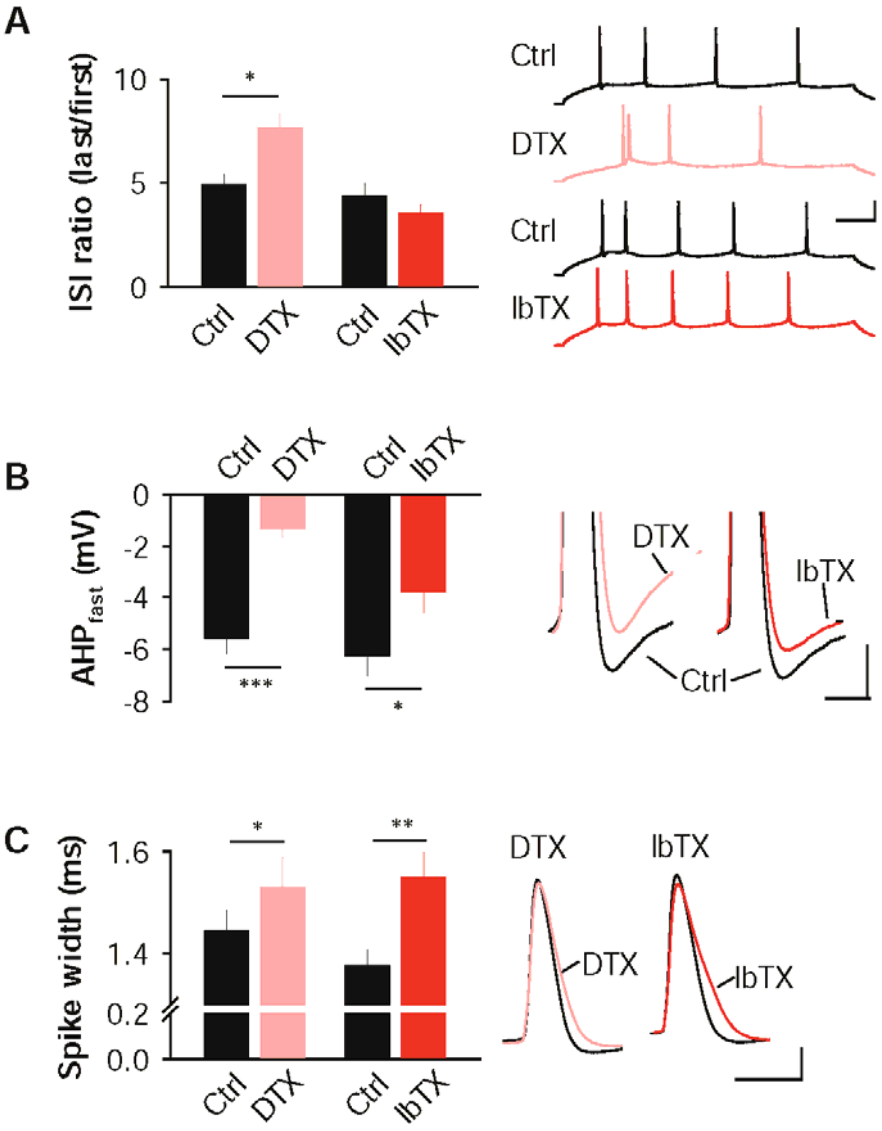


Figure S11. Contribution of voltage-gated and large conductance Ca^{2+}-dependent potassium channels to intrinsic properties of IL-projecting BA neurons.
(A) Burst firing of IL-projecting BA neurons is selectively increased upon blockade of VDPCs (α-DTX, 100 nM; n = x), but not after blockade of BK-channels (IbTX, 50 nM; n = x). (B, C) Currents mediated by voltage-dependent potassium channels (VDPCs) and large conductance potassium channels (BK-channels) are involved in AHP_{fast} and in action potential repolarization in IL-projecting BA neurons. Specific blockers of VDPCs (α-DTX, 100 nM) and BK-channels (IbTX, 50 nM) block the AHP_{fast} (B)(n = x and x) and increase spike width (C)(n = x and x). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

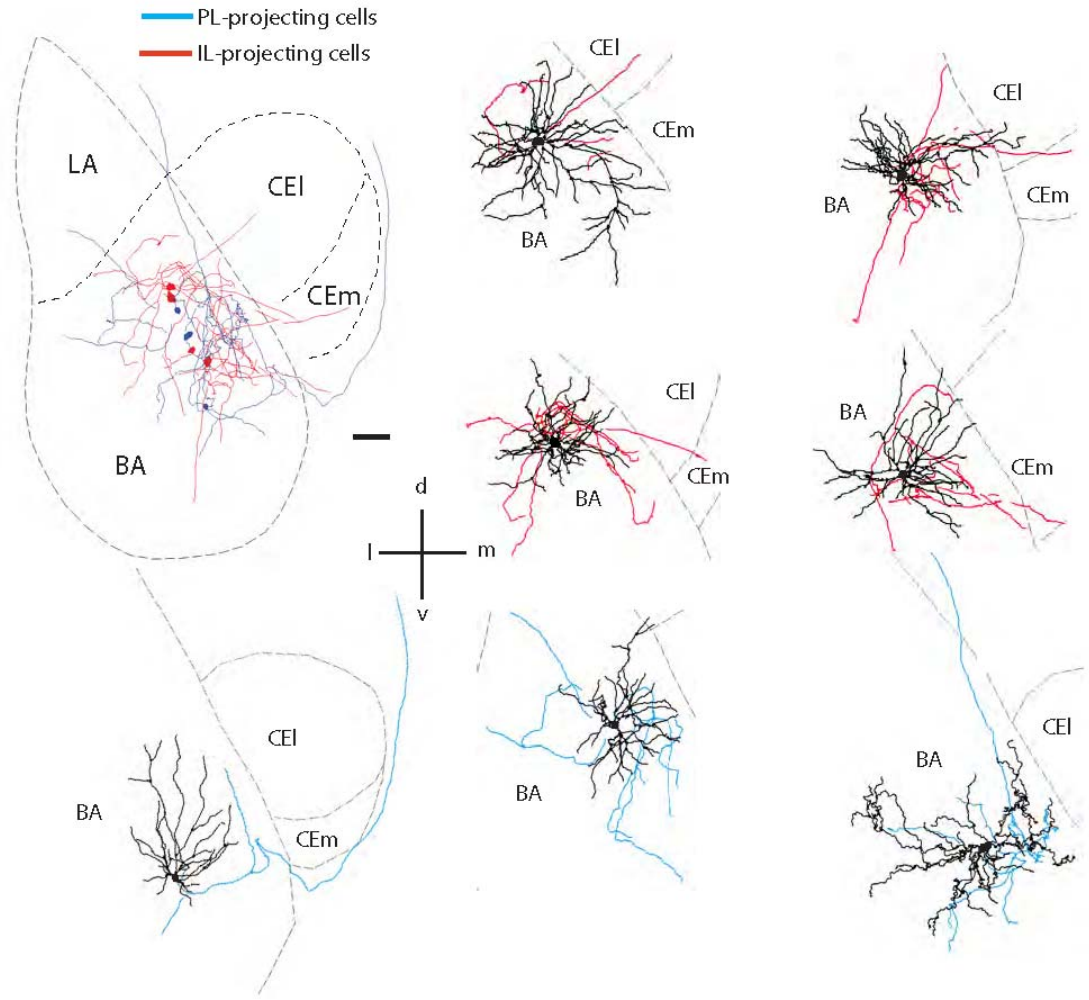


Figure S12. Morphological reconstructions of biocytin-filled PL- and IL-projecting BA neurons.
Top left panel shows a superposition of PL-projecting neurons (n = 3, blue) and IL-projecting neurons (n = 4, red) filled with biocytin during whole-cell recordings in acute brain slices. Only axonal arbors are shown. Whereas all four IL-projecting neurons (shown on the top right) exhibit axon collaterals projecting to the lateral or medial subdivision of the central amygdala (CEl, CEm), no central amygdala-projecting collaterals were found in any of the three PL-projecting neurons (bottom). For individual neurons, the somato-dendritic domain is shown in black and axons in blue (PL-projecting neurons) or red (IL-projecting neurons). Scale bar: 100 µm.

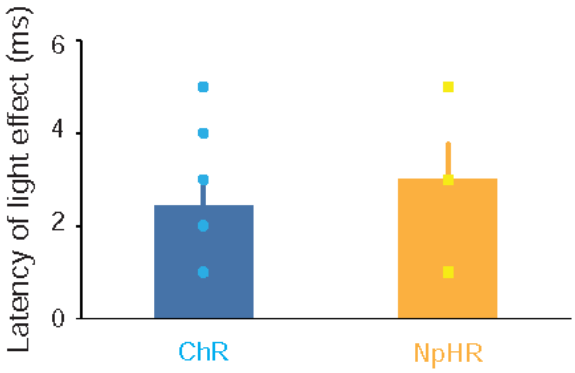


Figure S13. Latencies of light-induced activity changes.
Latencies of light-induced activity changes of optogenetically identified PL- or IL-projecting BA neurons, determined by a change point analysis of neuronal activity.

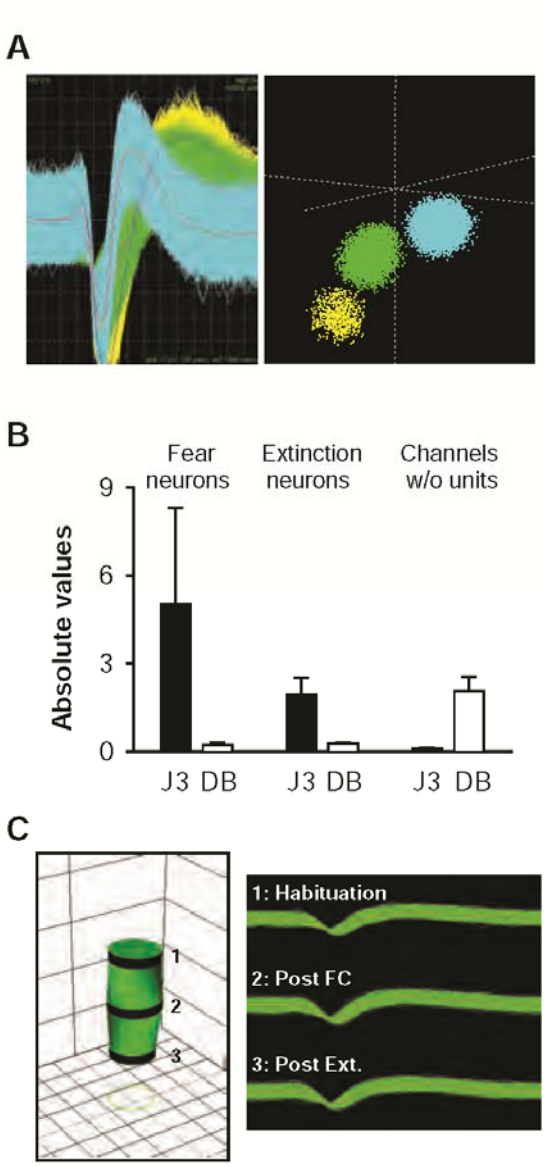
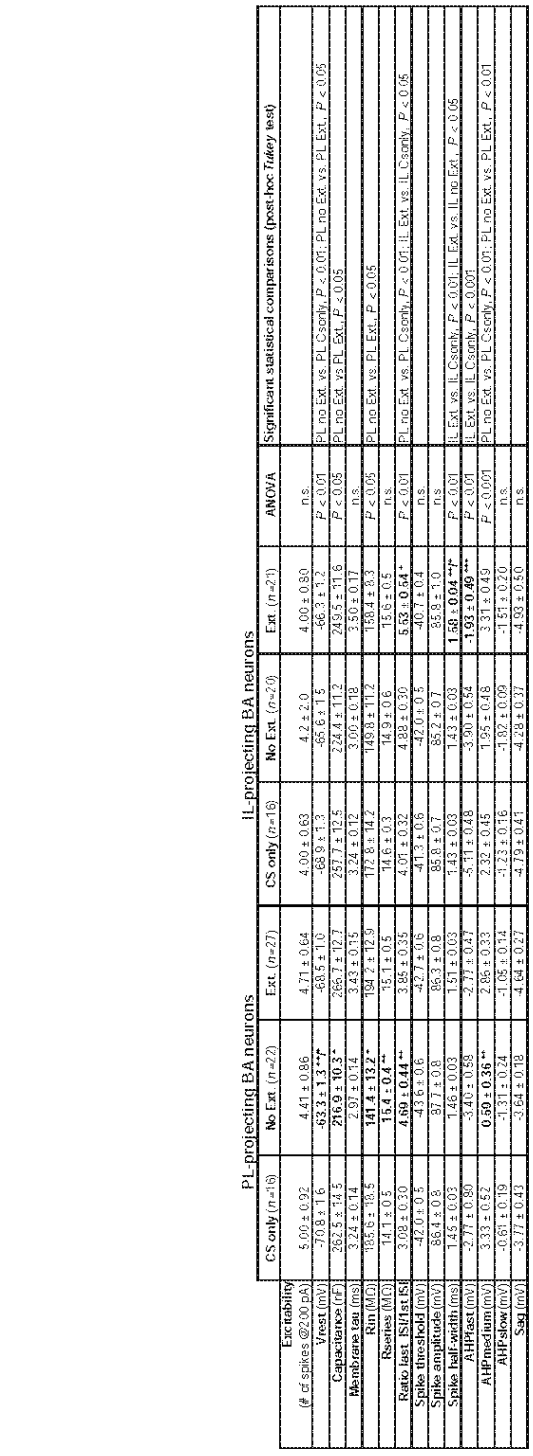


Figure S14. Stability of chronic single unit recordings.
(A)**,** Left: Superimposed waveforms recorded from three different units. Right: Spikes originating from individual units were sorted using 3D-principal component analysis. (B), Quantitative J3 and Davies Bouldin validity index (DB) statistics calculated for fear and extinction neurons. Controls values were obtained using two clusters defined from the centered cloud of points from channels in which no units could be detected. High values for the J3 and low values for the DB are indicative of good single unit isolation. (C), Left: Stability of clustered waveforms during long-term recordings was assessed by calculating principal component (PC) space cylinders. Straight cylinders suggest that the same set of single units was recorded during the entire training session. Right: Superimposed waveforms used to calculate the PC space cylinder recorded before habituation, after fear conditioning, and after extinction sessions**.**

Counts

Injection site	Group	n_{Fos}	n_{beads}	$n_{overlap}$	n_{mice}
mPFC	CS only	691	1589	83	10
	FC	937	932	112	8
	no Ext.	1321	1947	188	11
	Ext.	692	1166	96	8
HC	CS only	400	1042	49	7
	FC	389	373	33	3
	no Ext.	830	893	59	9
	Ext.	332	749	33	5
PL	CS only	212	449	23	4
	FC	463	434	62	3
	no Ext.	898	1414	157	8
	Ext.	406	888	63	5
IL	CS only	119	279	9	3
	FC	301	373	28	3
	no Ext.	326	412	22	3
	Ext.	286	278	33	3

Table S1. Summary table listing the non-normalized counts of retrogradely labeled cells, FOS-positive neurons and double-labeled cells for all experiments shown in Figures 1 and 2.



	PL-projecting BA neurons			IL-projecting BA neurons				
Excitability	**CS only** (n=16)	**No Ext.** (n=22)	**Ext.** (n=27)	**CS only** (n=16)	**No Ext.** (n=20)	**Ext.** (n=21)	**ANOVA**	**Significant statistical comparisons (post-hoc *Tukey* test)**
(# of spikes @200 pA)	5.00 ± 0.92	4.41 ± 0.86	4.71 ± 0.64	4.00 ± 0.63	4.2 ± 2.0	4.00 ± 0.90	n.s.	
Vrest (mV)	-70.8 ± 1.6	**-63.3 ± 1.3** $^{**,\#}$	-68.5 ± 1.0	-68.9 ± 1.3	-65.6 ± 1.5	-66.3 ± 1.2	$P < 0.01$	PL no Ext. vs. PL Csonly, $P < 0.01$; PL no Ext. vs. PL Ext., $P < 0.05$
Capacitance (nF)	262.5 ± 14.5	**216.9 ± 10.3** *	266.7 ± 12.7	257.7 ± 12.5	224.4 ± 11.2	249.5 ± 11.6	$P < 0.05$	PL no Ext. vs PL Ext., $P < 0.05$
Membrane tau (ms)	3.24 ± 0.14	2.97 ± 0.14	3.43 ± 0.15	3.24 ± 0.12	3.00 ± 0.18	3.50 ± 0.17	n.s.	
Rin (MΩ)	185.6 ± 16.5	**141.4 ± 13.2** *	194.2 ± 12.9	172.8 ± 14.2	149.8 ± 11.2	158.4 ± 9.3	$P < 0.05$	PL no Ext. vs. PL Ext., $P < 0.05$
Rseries (MΩ)	14.1 ± 0.5	**15.4 ± 0.4** **	15.1 ± 0.5	14.6 ± 0.3	14.9 ± 0.6	15.6 ± 0.5	n.s.	
Ratio last ISI/1st ISI	3.08 ± 0.20	**4.59 ± 0.44** **	3.86 ± 0.35	4.01 ± 0.32	4.88 ± 0.30	**5.63 ± 0.54** *	$P < 0.01$	PL no Ext. vs. PL Csonly, $P < 0.01$; IL Ext. vs. IL Csonly, $P < 0.05$
Spike threshold (mV)	-42.0 ± 0.5	-43.6 ± 0.6	-42.7 ± 0.6	-41.3 ± 0.6	-42.0 ± 0.5	-40.7 ± 0.4	n.s.	
Spike amplitude (mV)	86.4 ± 0.8	87.7 ± 0.8	86.3 ± 0.8	85.8 ± 0.7	85.2 ± 0.7	85.8 ± 1.0	n.s.	
Spike half-width (ms)	1.45 ± 0.03	1.46 ± 0.03	1.51 ± 0.03	1.43 ± 0.03	1.43 ± 0.03	**1.58 ± 0.04** $^{**,\#}$	$P < 0.01$	IL Ext. vs. IL Csonly, $P < 0.01$; IL Ext. vs IL no Ext., $P < 0.05$
AHPfast (mV)	-2.77 ± 0.90	-3.40 ± 0.58	-2.71 ± 0.47	-5.11 ± 0.48	-3.90 ± 0.54	**-1.93 ± 0.49** $^{**,\#}$	$P < 0.01$	IL Ext. vs. IL Csonly, $P < 0.001$
AHPmedium (mV)	3.33 ± 0.52	**0.59 ± 0.36** **	2.86 ± 0.33	2.32 ± 0.45	1.95 ± 0.48	3.31 ± 0.49	$P < 0.001$	PL no Ext. vs. PL Csonly, $P < 0.01$; PL no Ext. vs. PL Ext., $P < 0.01$
AHPslow (mV)	-0.61 ± 0.19	-1.31 ± 0.24	-1.05 ± 0.14	-1.23 ± 0.16	-1.02 ± 0.09	-1.51 ± 0.20	n.s.	
Sag (mV)	-3.77 ± 0.43	-3.64 ± 0.18	-4.64 ± 0.27	-4.79 ± 0.41	-4.28 ± 0.37	-4.93 ± 0.50	n.s.	

All values are calculated for the first spike (unless noted otherwise)

Table S2. Summary table listing all parameters recorded using *ex-vivo* whole-cell recordings from identified IL- or PL-projecting neurons in slices obtained from different behavioral groups.

	PL (CS; n=16)	IL (CS; n=26)	
Excitability (# of spikes; +200 pA step)	5.00 +/- 0.92	4.00 +/- 0.63	n.s.
V_{rest} (mV)	-70.8 +/- 1.6	-68.9 +/- 1.3	n.s.
Capacitance (nF)	262.5 +/- 14.5	257.7 +/- 12.5	n.s.
$\tau_{membrane}$ (ms)	3.24 +/- 0.14	3.24 +/- 0.12	n.s.
R_{input} (MOhm)	185.6 +/- 18.5	172.8 +/- 14.2	n.s.
R_{series} (MOhm)	14.1 +/- 0.5	14.6 +/- 0.3	n.s.
Ratio last ISI / 1st ISI	3.08 +/- 0.30	4.01 +/- 0.32	n.s.
Threshold (mV)	-42.0 +/-0.5	-41.3 +/- 0.6	n.s.
Spike amplitude (mV)	86.4 +/- 0.8	85.8 +/- 0.7	n.s.
Spike half-width (ms)	1.45 +/- 0.03	1.43 +/-0.03	n.s.
AHP_{fast} (mV)	-2.77 +/- 0.80	-5.11 +/- 0.48	n.s. (p=0.053)
AHP_{medium} (mV)	3.33 +/- 0.52	2.32 +/- 0.45	n.s.
AHP_{slow} (mV)	-0.61 +/- 0.19	-1.23 +/- 0.16	n.s.
Sag (mV)	-3.77 +/- 0.43	-4.79 +/- 0.41	n.s.

Table S3. Summary table listing basic electrophysiological parameters of IL- and PL-projecting BA neurons in control (CS-only) animals.

	Fear cell	Fear cell	Ext cell	Ext cell	Ext cell	Ext cell	Ext cell
HAB-X1	0.9891	0.9693	0.9545	0.9958	0.9531	0.9871	0.9878
HAB-X2	0.9843	0.9702	0.9904	0.9883	0.9511	0.9807	0.9907
X1-X2	0.9918	0.9726	0.9653	0.9901	0.9965	0.997	0.9927
	Random cell	Random cell	Random cell	Random cell	Random cell	Random cell	Random cell
Hab-X1	0.7345	0.7786	0.9433	0.3933	0.9715	0.7877	0.9644
Hab-X2	0.7196	0.752	0.7425	0.4227	0.9456	0.9573	0.8894
X1-X2	0.8549	0.8633	0.8785	0.8533	0.8785	0.7442	0.9381

Table S4. To quantitatively evaluate similarity of different spike shapes recorded on different days, linear correlation values between time-shifted average waveforms were calculated for fear and extinction neurons. As a control we computed the r values from average waveforms of randomly selected neurons.

EXPERIMENTAL PROCEDURES

Animals

Male C57BL/6J mice (2 – 4 month, RCC Ltd.) were individually housed in a 12 h light/dark cycle while the experiments were running. Food and water was available *ad libitum*. All animal procedures were performed in accordance with institutional guidelines and were approved by the Veterinary Department of the Canton of Basel-Stadt.

Behavior

Mice were submitted to an auditory fear conditioning paradigm in which the CS (total CS duration of 30 s, consisting of 50-ms pips repeated at 0.9 Hz, 2-ms rise and fall; pip frequency: 7.5 kHz, 80 dB sound pressure level) was paired to the US (1 s foot shock, 0.6 mA, 5 CS/US pairings; inter-trial interval: 20–180 s). The onset of the US coincided with the offset of the CS. Fear conditioning was always performed in a context (context A) different from that used for all other behavioral sessions (context B). Context A was cleaned with 70% ethanol and context B with 1% acetic acid. Extinction training was performed over two days in context B by presenting 16 CSs each day (Herry and Mons 2004). The "No Extinction" group was presented with 4 CSs in context B; for either two days (histology data) or one day (electrophysiology data). CS only control mice were not subjected to fear conditioning but exposed to 4 CSs on three consecutive days. Freezing behavior was quantified in each behavioural session using an automatic infrared beam detection system (Coulbourn Instruments) as described previously (Herry et al. 2007). The animals were considered freezing if no movement was detected for 2 s. Changes in freezing values (Δ freezing) were calculated as differences between CS-induced freezing after learning vs. CS-induced freezing before learning (for Ext. and FC groups) or CS-induced freezing vs. pre-CS baseline freezing (for CS-only and no-Ext. groups).

Retrograde labeling and surgeries

For retrograde labeling of BA neurons projecting to the mPFC or to the hippocampus, we used fluorophore-coated latex microspheres (red exc.= 530 nm, em.= 590 nm and green exc.= 460 nm, em.=505 nm) referred to as retrobeads (RBs) (Lumafluor Inc., Naples, FL). RBs were dialyzed against a sucrose solution (0.32 M) to reduce osmotic stress in injected tissues. Polycarbonate membrane filters (Sterlitech Corp., Washington, USA; pore-size 0.01 µm, diameter 25 mm) were placed in a cell culture dish containing ~ 10 ml sucrose solution and incubated overnight at room temperature. Subsequently, RBs were stored at 4 °C until use.

Mice were anesthetized with initially 3% Isoflourane (Minrad, Inc., Buffalo, NY, USA) and with 2% during stereotactical surgery. Mice were placed into a stereotactical frame (David Kopf Instruments; Bilaney GmbH, Düsseldorf, Germany) and injected with RBs to the mPFC or the HC using a syringe (600 Series, 5 µl, Hamilton Bonaduz AG, Bonaduz, GR, Switzerland) or a picospritzer (Föhr Medical Instruments GmbH, Seeheim-Ober Beerbach, Germany) in combination with glass pipettes (borosilicate glass capillaries, World Precision Instruments,

Inc., Sarasota, Florida, USA pulled on a Flaming/Brown micropipette puller P-97, Sutter Instruments, Novato, CA, USA). The injection volumes for the mPFC and the hippocampus were 0.2 and 0.3 µl, respectively. Coordinates to target bilaterally the mPFC were originating from bregma: rostral + 1.75, lateral ± 0.3 and ventral 2.3 (PL) and 2.5 (IL); the hippocampus coordinates were caudal – 3.6, lateral ± 3.6 and ventral 3.7 (all coordinates calculated from the Mouse Brain Atlas by Franklin and Paxinos). During surgery, mice were locally treated with lidocaine (Boehringer Ingelheim, Germany) and naropin (Astrazeneca, Germany) for analgesia. Post-surgery treatment of mice involved subcutaneous injection of 0.06 ml Metacam (Boehringer Ingelheim, Germany) to reduce pain and inflammation risk.

Immunohistochemistry and imaging

Mice were deeply anaesthetized using urethane 2 h after the completion of the behavioral testing and perfused transcardially with ice-cold solution of 4% paraformaldehyde in phosphate buffer (PFA; pH 7.4). After postfixation overnight in the same fixative at 4 °C, coronal sections (50 µm) were cut on a vibratome (Leica, Nussloch, Germany) and collected in phosphate-buffered saline (PBS). Free-floating sections were rinsed in PBS four times. Subsequently, sections were incubated in blocking solution (10% bovine serum albumin (BSA) and 0.2% Triton X-100 in PBS) containing the primary polyclonal rabbit anti-FOS antibody (Oncogene Research Products; 1: 20000 dilution) over night at 4 °C. Subsequently, sections were washed with PBS and incubated for 2 h at room temperature with fluorescent goat anti-rabbit IgG (Santa Cruz; 1:1000 in PBS). Finally, immuno-labeled sections were rinsed three times with PBS, mounted on gelatin-coated slides, dehydrated and coverslipped. 3D z-stacks and 2D overview pictures (tiles) were acquired using a LSM 510 or a LSM 510 meta confocal microscope (Carl Zeiss AG, Germany) and three different laser lines (488, 543 and 633 nm). Tiles and z-stacks were acquired using a 40x/1.3 oil immersion DIC lens. Settings for acquisition (photomultiplier assignment and contrast values) were adjusted for different staining batches using a pixel saturation tool on very bright and dark spots. Pinhole was always adjusted to 1. Tiled images from the entire amygdala were collected with an opened pinhole. Confocal z-stack image series were sectioned in 1 µm thin optical planes (325.8 µm x 325.8 µm x 10-35 µm), tiles were taken in 9 x 9 images (2592.8 µm x 2602.2 µm). For quantification of IEG induction, six tiled overview pictures were analysed for each animal. For the quantification of the colocalization of RBs and FOS, five stacks were analysed for each animal. Quantification of FOS-expressing neurons was determined using the Imaris software package (Bitplane, Zürich, Switzerland) and LSM Image Browser (Carl Zeiss AG, Germany). The different subnuclei of the BLA (BA, LA, BMA) were defined with the Image Browser and then imported to Imaris to automatically count positive nuclei with the spot detection software. Parameters were kept constant within comparable batches of raw files. The spot diameter was set to 10 µm for every analysis, whereas the detection threshold was changed depending on the experiment. For the RBs/IEG colocalisation, double labeled neurons were counted manually in the LSM Image Browser.

Viral expression of optogenetic constructs

For optogenetic manipulation of IL- or PL-projecting BA neurons an intersectional viral approach was used. A mixture of Herpes Simplex virus (HSV-1)(Lima et al. 2009)(BioVex, London, UK) and Canine Adeno Virus (CAV-2)(Hnasko et al. 2006) was bilaterally injected into the IL or PL. Blue fluorescent latex beads were co-injected to localize injection sites. Both HSV-1 and CAV-2 have the ability to enter axon terminals and thereby retrogradely infect projection neurons. The two viruses infected largely non-overlapping subpopulations of the IL- or PL-projecting BA neurons found with conventional tracing approaches (Figures S4 and S5). Therefore, a mixture of the two viruses was used to infect a larger population. Both HSV-1 and CAV-2 delivered constructs for expression of Cre recombinase. An adeno-associated virus (AAV, serotype 2/7, Vector Core, University of Pennsylvania) for Cre-dependent coexpression of channelrhodopson-2 (ChR2), Halorhodopsin (NpHR) and Venus (AAV2/7 EF1a::DIO-ChR2(H134R)-2A-eNpHR-2A-Venus)(Tang et al. 2009) was injected into the BA of the same animals. This approach allowed for specific opsin expression only in IL-projecting BA cells. All viruses (approximately 0.3 µl HSV/CAV mix (1:1) and 0.5 µl AAV per hemisphere) were injected from glass pipettes (tip diameter 10–20 µm) connected to a Picospritzer at the following coordinates: IL rostral + 1.75, lateral ± 0.3 and ventral 2.5; PL rostral + 1.75, lateral ± 0.3 and ventral 2.5; BA caudal – 1.7, lateral ± 3.4 and ventral 4.2.

Optrode recordings and optogenetic manipulation of behaviour

For optogenetic manipulation of behavior animals were bilaterally implanted with custom-built optic fibre connectors (0.48 numerical aperture, 200 µm diameter, Thorlabs). Fibre ends were lowered to 500 µm above the BA at the injection site. Implants were fixed to the skull with cyanoacrylate glue (Ultra Gel, Henkel) and dental cement (Paladur, Heraeus). After 1-2 months, the implanted fibers were connected to a custom-built laser bench (lasers: MBL473 and MGL593.5, CNI Lasers). To allow the animals to move freely, the fibres were suspended over the experimental context.

Animals were subjected to fear conditioning and extinction training as described above, with some modifications. Animals were conditioned to two different CSs (pip structure and properties as described) of distinct frequencies (3 kHz and 12 kHz, respectively). Both CSs were paired with foot-shocks (5 pairings per CS). During extinction training both CSs were presented 16 times. The first 4 presentations of each CS were used to test initial freezing levels. Subsequent CS presentations were paired during each pip with blue light illumination for one of the CSs and yellow light for the other CS. Light was delivered for 300 ms, starting 50 ms before pip onset. This 2x2 experimental design (2 opposing optogenetic manipulations of 2 distinct pathways in 2 groups of animals) maximized the effects in each individual animal, and allowed for statistical analysis using 2way ANOVA. During retrieval on the next day CSs were presented without light. The order of the frequencies and the pairing with blue or yellow light were counterbalanced over the experimental set.

For validation of the optogenetic approach animals underwent the same virus injection procedure as described above. Here, one optic fibre connector was replaced by a custom-built optrode. Optrodes consisted of an optic fibre connector with an attached electrode. Electrodes were made of 16 individually insulated, gold-plated nichrome wires (13 µm inner diameter, impedance 30 to 100 kΩ, Sandvik), attached to a connector (18 pin, Omnetics). The whole implant was fixed to the skull as described above. Single unit recordings were performed as described below. For the validation of the approach implanted animals were transferred to a behavioural context and pulses of blue and yellow light (300 ms, 60 presentations each, 2 s inter pulse interval) were presented during recording of single unit activity to test for light responsiveness of individual recorded units. Neurons were considered as light responsive if they showed bidirectional time-locked, short-latency activity changes upon illumination (Figure S13).

For the functional characterization of optogenetically identified IL- and PL-projecting BA cells, animals were subjected to a discriminative fear conditioning and extinction paradigm as described before (Herry et al. 2008). On day 1, mice were habituated by 4 presentations of the CS+ and the CS– (pip frequency: 7.5 kHz or white-noise) in context B. Discriminative fear conditioning was performed on day 2 by pairing the CS+ with a US (5 CS+/US pairings) in context A. The CS– was presented after each CS+/US association but was never reinforced. The frequencies used for CS+ and CS– were counterbalanced across animals. On day 3 and day 4, conditioned mice underwent extinction training in context B, consisting of 4 and 12 presentations of the CS– and the CS+, respectively.

Single unit activity was recorded during all behavioral sessions. Before the habituation session and after the second extinction session light responsiveness of individual recorded units was tested. Extinction-, fear- and persistent-cells were defined as previously described (Herry et al. 2008). Extinction cells showed a significant increase in the *z*-score transformed neuronal activity in the first 100 ms after CS+ onset only in the late phase of extinction (last 4 CSs; low fear) and not in the early phase (first 4 CSs; high fear). Conversely, fear cells showed a significant increase in the *z*-score transformed neuronal activity in the first 100 ms after CS+ onset only in the early phase of extinction (first 4 CSs; high fear) and not in the late phase (last 4 CSs; low fear). Persistent cells showed significant activity changes during both high and low fear states. After completion of the experiment, recording sites were marked with electrolytic lesions before perfusion. Viral injection sites and electrode locations were validated with immunofluorescence and standard histological techniques.

For the optogenetic behavior experiments (Figure 3F) and for the functional characterization of IL- or PL-projecting BA neurons (Figure S8), only mice with bilateral injections exclusively in the IL or PL, respectively, and the BA were included. For the validation of the optogenetic approach animals with bilateral mPFC injections, BA injection on the recorded hemisphere and electrode implantation in the BA were included (Figure 3C).

In vivo electrophysiology and analysis

Single-unit spike sorting was performed using Off-Line Spike Sorter (OFSS, Plexon) as described (Herry et al. 2008). Briefly, principal component scores were calculated for unsorted waveforms and plotted on three-dimensional principal component spaces, and clusters containing similar valid waveforms were manually defined (Figure S14). A group of waveforms was considered to be generated from a single neuron if it defined a discrete cluster in principal component space that was distinct from clusters for other units and if it displayed a clear refractory period (>1 ms) in the auto-correlogram histograms. In addition, two parameters were used to quantify the overall separation between identified clusters in a particular channel. These parameters include the J3 statistic, which corresponds to the ratio of between-cluster to within-cluster scatter, and the Davies–Bouldin validity index (DB), which reflects the ratio of the sum of within-cluster scatter to between-cluster separation. High values for the J3 and low values for the DB are indicative of good single-unit isolation. Control values for these statistics were obtained by artificially defining two clusters from the centered cloud of points in the principal component space from channels in which no units could be detected. Template waveforms were then calculated for well-separated clusters and stored for further analysis. Clusters of identified neurons were analyzed offline for each recording session using principal component analysis and a template-matching algorithm. Only stable clusters of single units recorded over the time course of the entire behavioral training were considered. Long-term single-unit stability isolation was first evaluated using Wavetracker (Plexon) in which principal component space-cylinders were calculated from a 5 min segment of data spontaneously recorded before any training session. Straight cylinders suggest that the same set of single units was recorded during the entire training session (Figure S14).

Second, we quantitatively evaluated the similarity of waveform shape by calculating linear correlation (r) values between average waveforms obtained over training days (Herry et al. 2008). As a control, we computed the *r* values from average waveforms of different neurons (Table S4).

Third, for each unit we used correlation analysis to quantitatively compare the similarity of waveform shape during CS^+-stimulation and during a 60 s period of spontaneous activity recorded before each behavioral session. To avoid analysis of the same neuron recorded on different channels, we computed cross-correlation histograms. If a target neuron presented a peak of activity at a time that the reference neuron fires, only one of the two neurons was considered for further analysis. CS-induced neural activity was calculated by comparing the firing rate after stimulus onset with the firing rate recorded during the 500 ms before stimulus onset (bin size, 20 ms; averaged over blocks of 4 CS presentations consisting of 108 individual sound pips in total) using a z-score transformation. z-score values were calculated by subtracting the average baseline firing rate established over the 500 ms preceding stimulus onset from individual raw values and by dividing the difference by the baseline standard deviation. Only CS-excited neurons were considered for analysis. Classification of

units was performed by comparing the largest significant z-score values within 100 ms after CS-onset during post-fear conditioning and extinction sessions according to the freezing levels. For high-fear states, the entire post-fear conditioning session was analyzed, whereas, for low-fear states, analysis was restricted to the block of 4 CS presentations during which the fear level was the lowest. A unit was classified as a fear neuron if it exhibited a significant *z*-score value after fear conditioning (when freezing levels were high), but no significant z-score value after extinction (when freezing levels were low), and vice versa for extinction neurons. Finally, units were classified as extinction-resistant neurons if they displayed a significant z-score value during both post-fear conditioning and extinction sessions, independently of freezing levels.

Analysis of spike width and bursting activity of fear and extinction neurons was performed on data from previously collected *in vivo* single unit data (Herry et al. 2008). Tone-evoked changes in spike width in fear and extinction neurons were analyzed during the habituation (4 CSs), early extinction (first 4 CSs) and late extinction (last 4 CSs) sessions. For each pip in a given CS presentation, only spikes occurring within the first 500 ms after CS onset were analyzed. Furthermore, to prevent potential confounds with burst-induced changes in firing frequency that may influence the spike width, the analysis was restricted to spikes exhibiting a minimum interspike interval of 50 ms. Spike width was measured from trough to peak, and average values were normalized to the habituation session. To analyze bursting activity we used the Poisson surprise method (Legendy and Salcman 1985) applied to the entire habituation, early extinction and late extinction sessions, but excluding periods of CS presentation. Briefly, this method detects bursts by finding consecutive interspike intervals (ISI) shorter than half the mean ISI, and calculates the "unlikeliness" or "surprise" that a given ISI would be expected if the spike train was a Poisson process with random temporal patterning. Furthermore, the Poisson surprise method is a rigorous detector of bursts because it is non-sensitive to fluctuations in average firing rate and treats each spike train as an independent entity. We used a surprise value of 3 in our analysis which corresponds to a probability of 0.05 of finding a burst of spikes within a random sequence. Surprise values were calculated for bursts with at least 3 spikes with consecutive ISIs less than half the mean ISI of the entire sequence.

***In vitro* electrophysiology**

Brain coronal slices were prepared from six to nine week old male mice injected with RBs. Two hours after behavioural training, brains were dissected in ice-cold artificial cerebrospinal fluid (ACSF), and sliced (300 µm thick) with a Microm slicer (HM 650 V; Walldorf, Germany) at 4 ºC. Sapphire blades (Delaware Diamond Knives, USA) were used to improve slice quality. Slices were recovered for 45 min at 37 ºC in an interface chamber containing ACSF equilibrated with 95% O2/5% CO2. The ACSF contained (in mM): 124 NaCl, 2.7 KCl, 2 $CaCl_2$, 1.3 $MgCl_2$, 26 $NaHCO_3$, 0.4 NaH_2PO_4, 18 glucose, 2.25 ascorbate. Whole-cell patch-clamp recordings were obtained from RB-labeled projection neurons in the BA at room temperature in a submerged chamber with constant ACSF perfusion. Neurons were

visually identified with infrared video microscopy using an upright microscope equipped with x5 and x40 objectives (Olympus, Hamburg, Germany). Fluorescent cells were identified using a Polychrome V Till imaging system (Till Photonics, Gräfelfing, Germany). Patch electrodes (3-5 MΩ) were pulled from borosilicate glass tubing and were filled with an intracellular solution consisting of (in mM): 130 potassium-gluconate, 10 HEPES, 10 phosphocreatineNa_2, 4 Mg-ATP, 0.4 Na-GTP, 5 KCl, 0.6 EGTA (pH adjusted to 7.25 with KOH, ~ 280 – 300 mOsm). In voltage- and current-clamp recordings, membrane potential was held at −70 mV. All recordings were performed at room temperature in the presence of 100 µM picrotoxin.

Data were acquired with pClamp9.2 (Axon Instruments, Union City, CA, USA) and recorded with a Multiclamp700B amplifier. Voltage-clamp experiments were filtered at 2 kHz and sampled at 5 kHz, whereas all current-clamp experiments were filtered at 10 kHz and sampled at 20 kHz. Series resistance was monitored throughout the experiments by applying hyperpolarizing pulses. When series resistance changed by more than 20% the experiment was terminated. Spiking patterns were assessed by applying two different protocols to the cells in current-clamp; one applying hyperpolarizing currents from -200 pA to depolarizing currents at +250 pA in 50 pA steps, the other ranging from -400 pA to +500 pA in 100 pA steps. Excitability was measured by comparing between current steps of the same amplitude. Firing accommodation was measured by calculating the ratio of the last over the first inter-spike interval using sweeps with 5 action potentials. AHP_{fast} was measured by determining the minimum V_m in a 7 ms time window after action potential onset. Action potential half-width was measured at 50% of the peak amplitude. All values are averages of three measurements per cell using the first action potential elicited by the minimal supra-threshold current injection. For recordings of spontaneous synaptic currents, two 5 min periods were recorded and before and after series resistance was measured. All values are expressed as means ± s.e.m. Statistical comparisons were performed with paired or unpaired Student's t-test or with a one- or two-way ANOVA followed by a post-hoc Tukey test as appropriate (two-tailed $p < 0.05$ was considered significant). One neuron per slice was filled with a 1:1 Neurobiotin/Biocytin mix (5 mg/ml; Vector Laboratories, Inc., Burlingame, CA, USA; Sigma-Aldrich Chemie GmbH, Steinheim, Germany). Subsequently, slices were transferred to 4% PFA, kept for 1 – 3 weeks at 4°C for fixation, and revealed using standard methods (ABC-elite kit; Vector Laboratories, Inc., Burlingame, CA, USA).

Acknowledgements

We thank all members of the Lüthi laboratory for discussions and comments. We thank P. Argast and J. Lüdke for excellent technical assistance, and K. Deisseroth for generously sharing materials. This work was supported by grants from the Swiss National Science Foundation (to A.L. and to J.J.L., Ambizione Fellowship), the National Competence Center in Research (NCCR) of the Swiss Confederation on the synaptic basis of mental disorders, the

Schering Foundation (S.B.E.W.), EMBO Long-Term Fellowships and Marie-Curie Actions (J.G. and J.F.), the Aquitaine Regional Council (C.H.), the European Commission (Eurospin Project, Contract HEALTH-F2-2009-241498), and the Novartis Research Foundation.

4. DISCUSSION:

4.1 The role of inhibition and disinhibition in associative learning

Associative learning depends on the temporally contingent co-activation of neurons by sensory stimuli. This co-activation causes synaptic plasticity on a cellular level and changes in connectivity and information processing on the level of neuronal circuits. To ensure appropriate adaptations to the environment, learning and plasticity have to be tightly controlled. A major mechanism for this is the regulation of the stimulus-evoked activity which triggers plasticity - putting inhibition into the focus of interest. In fact, inhibition has been implicated in certain forms of associative learning, like fear conditioning (Ehrlich et al. 2009). Inhibition can modulate the strength, timing and precision of neuronal responses. However, how inhibition modulates plasticity and controls long-term modifications in neuronal circuits beyond mere changes of general cellular activity remains elusive.

Recently, interneuron-interneuron interactions and disinhibition are emerging as crucial circuit mechanisms for the control of plasticity and learning (Ciocchi et al. 2010; Letzkus et al. 2011; Wolff et al. 2013). Disinhibition is an attractive concept for the control of learning, since it gates strong activation of PNs and is thereby permissive for concurrent plasticity induction, but not necessarily causative. What is the advantage of disinhibitory mechanisms over selective excitation? Disinhibitory mechanisms require a certain level of tonic inhibition and are therefore relatively costly for the neuronal circuit. However, disinhibition allows to quickly and flexibly change the impact the very same sensory stimuli can make, depending on the state of the animal. Depending on previous experience and network state, disinhibition can gate sensory inputs and thereby select which circuit elements can exhibit plasticity. This way, disinhibition can also control the specificity of learning, without the need for a change of the incoming inputs themselves. Disinhibitory mechanisms may have evolved as additional control elements for activity and plasticity, since they can be placed on top of existing circuits and do not require changes in the general wiring of incoming inputs.

In this thesis, I demonstrated that disinhibition is indeed a crucial factor in the control of associative fear learning, conserved across neuronal circuits in different brain areas (Letzkus et al. 2011; Wolff et al. 2013). I identified disinhibitory microcircuits both in the auditory cortex (chapter 3.1) – between INs in layer 1 and PV^+ INs in layer 2/3 - and in the BLA (chapter 3.2) – between PV^+ and SOM^+ INs. Disinhibition of PNs during salient stimuli is critical in both brain areas for the acquisition of fear memories. Despite differences in their precise architecture, the presence of disinhibitory microcircuits and their general function appears to be a recurrent theme in different brain areas.

In both the AC and the BLA, PV^+ INs are inhibited during the shock, which results in perisomatic disinhibition of PNs (Letzkus et al. 2011 - Fig. 3; Wolff et al. 2013 - Fig. 2). This enhances the activation of PNs by the shock and in turn boosts the plasticity of co-activated

auditory inputs. In the BLA the shock-induced disinhibition is not restricted to the perisomatic region. Concurrent, shock-induced inhibition of dendrite-targeting SOM^+ INs leads to disinhibition along the entire somatodendritic axis of PNs (Wolff et al. 2013 - Fig. 3, 5). This is of particular importance, since BLA PNs receive auditory inputs at their dendrites (Farb and LeDoux 1997; McDonald 1998).

It is appealing to assume that a concurrent disinhibition of the perisomatic and dendritic compartments is a general mechanism for the gating of strong salient inputs. Its presence in the AC may enhance shock-induced PN activation and plasticity. Furthermore, this mechanism could also be employed in other areas, like the visual cortex, where we already observed a part of the pathway for perisomatic disinhibition during a footshock (Letzkus et al. 2011 - Suppl. Fig. 5). In sum, concurrent perisomatic and dendritic disinhibition in response to salient stimuli represents an attractive general circuit mechanism to gate stimulus-evoked activity and to boost plasticity.

Dendritic disinhibition appears to be a critical concept also beyond shock-evoked responses. In the BLA, it as well enhances tone responses in PNs. Tone-excited PV^+ INs inhibit SOM^+ INs, causing dendritic disinhibition, which can regulate the strength of the acquired fear responses (Wolff et al. 2013 - Fig. 3, 4). Notably, dendritic disinhibition could not only regulate the strength of a fear memory, but would also be an appealing mechanism to control its precision. Dendritic disinhibition could be stimulus-specific on several levels. Depending on the auditory input – and the state of the animal – only certain cells, or even only locally restricted branches of a dendritic tree could be disinhibited (Chiu et al. 2013). This would allow for a further selection of the potentiated circuit elements. It remains to be tested whether the tone-evoked dendritic disinhibition in the BLA in fact is locally specific and whether this process is also employed in other brain areas and distinct forms of learning.

The interaction between PV^+ and SOM^+ INs during the tone does not only cause the disinhibition of PN dendrites, but at the same time it induces a shift of inhibition from the dendritic to the perisomatic compartment of PNs. While the activation of PV^+ INs leads indirectly to dendritic disinhibition, it increases perisomatic inhibition at the same time. A possible consequence of this heightened inhibition could be an improved signal-to-noise ratio. Ongoing background and network activity in the PN could be reduced, so that the auditory inputs at the (disinhibited) dendrites have an even stronger impact on the cell's activity. This boosted saliency of the inputs may in turn increase the chances for potentiation of the activated synapses. Preliminary results on the role of PV^+ IN activity for generalization may support these ideas (chapter 4.1.1).

Overall, these considerations suggest that interneuron-interneuron interactions provide ample opportunities for a highly complex regulation of activity and plasticity in neuronal circuits. Still, it remains to be tested whether disinhibition is a general mechanism and recurrent motive in multiple brain areas and learning processes.

4.1.1 Regulation of Generalization

Inhibition and disinhibition are crucial control elements for fear learning. Beyond the mere strength of the acquired associations, they may also regulate more complex aspects, like the specificity of fear memories. The balance between precision and generalization is crucial for an animal's survival. Both too strong and too little generalization of fearful events and memories are maladaptive. A disregulation of generalization can cause pathologies, like Post-Traumatic Stress Disorder (PTSD), in which patients over-generalize and even stimuli, only slightly related to traumatic experiences, can trigger fear responses (Kent and Rauch 2003; Adolphs 2013). Therefore, the precision of fear memories must be carefully tuned.

Generalization may be regulated on multiple levels in the fear circuit. Plasticity in the auditory cortex may already determine the strength and the precision of auditory inputs to the amygdala. A broad potentiation in the AC may in turn lead to unspecific activations of the amygdala and to generalized fear responses. The specificity of this tone response plasticity may be regulated by (dis-)inhibitory mechanisms.

Preliminary experiments suggest that disinhibition indeed underlies memory precision in the BLA, downstream of the AC. Activation of PV^+ INs during a CS+ during conditioning leads to a strong generalization to a CS- and to non-conditioned stimuli. This effect may be mediated by the same disinhibitory microcircuit which determines the strength of fear responses (Wolff et al. 2013). The optogenetically enhanced PV^+ IN activity will also strengthen the dendritic disinhibition of PNs by SOM^+ IN inhibition. This, and possibly also the improved signal-to-noise ratio by stronger perisomatic inhibition, will enhance the PN tone responses. Importantly, the boosting of these mechanisms could also lead to the potentiation of auditory inputs, which are physiologically only mildly activated and which would normally not be strengthened. Subsequent presentation of the preferred stimuli for these additionally potentiated afferents could then lead to an inappropriate, generalized fear response. This suggests that the described disinhibitory microcircuit also controls the precision of fear memories. The stronger and the more widespread the dendritic disinhibition, the less similar auditory inputs have to be to the CS+ to be potentiated and the more generalized the fear response will be.

4.1.2 Disinhibition in the extinction and expression of fear

Fear extinction is a distinct learning process, but its mechanisms partially overlap with fear conditioning. The acquisition of extinction memories has to be tightly regulated and it is conceivable that the same or similar mechanisms are involved as in fear conditioning. During extinction training the CS is presented multiple times without reinforcement. Dendritic disinhibition will probably regulate the intensity of the auditory responses in this paradigm as well. The amplitude of the tone responses should determine the strength of the

prediction errors if the pairing with the shock is omitted. This in turn should affect the rate of extinction. Based on these assumptions, optogenetic manipulations of PV^+ or SOM^+ INs during presentation of the CS during extinction training should produce effects on learning analog to manipulations during fear acquisition. Enhanced dendritic disinhibition, causing stronger tone responses and possibly prediction errors, should boost, while impairment should interfere with extinction. It remains to be determined whether this mechanism is indeed involved in fear extinction.

If dendritic disinhibition determines the amplitude of PN activation by the CS, does it also affect the expression of fear memories after learning? It might be possible that enhanced tone responses also lead to stronger fear responses. However, this effect was not observed in preliminary experiments, using manipulations of IN activity during CS presentation after conditioning. This indicates that the control of tone responses via disinhibition might play an important role solely for the plasticity of inputs, but not for the expression of fear responses. However, it is possible that IN activity manipulations during fear expression only have an effect on simultaneous ongoing extinction, but not on the immediate expression. Further experiments are necessary to confirm the lack of an expression effect and to reveal the underlying reasons.

4.1.3 Regulation of disinhibition

While inhibition and disinhibition regulate at least the acquisition of fear, they have to be regulated themselves, too. This becomes apparent by the stimulus-specific disinhibition in the BLA during fear conditioning. The same microcircuit can cause disinhibition along the somatodendritic axis of PNs or can induce a shift from dendritic to perisomatic inhibition, depending on the stimulus (Wolff et al. 2013).

What determines the differential effects of the identified disinhibitory microcircuit? A major factor are the inputs to the different interneuron types. Different stimulus-evoked responses can be due to differential tone and shock inputs to PV^+ and SOM^+ INs. The strong tone-evoked activation of PV^+ INs suggests that they receive direct auditory inputs (Woodson et al. 2000; Sah et al. 2003). In contrast, I observed tone-evoked inhibition of SOM^+ INs, which is most likely mediated by PV^+ INs. This does not exclude that SOM^+ INs also receive direct auditory inputs – in fact, a minority of SOM^+ INs is tone-excited (Wolff et al. 2013 - ED Fig. 4) - but if they do, these inputs are too weak to overcome the PV^+ IN mediated inhibition.

During the shock however, PV^+ as well as SOM^+ INs in the BLA are inhibited (Wolff et al. 2013 - ED Fig. 4). Both types of INs may be contacted by a third IN class. Likely candidates are calretinin- (CR^+) or vasointestinal peptide-expressing (VIP^+) INs, which are known to mainly contact other INs (Fishell and Rudy 2011; Jiang et al. 2013). They may receive direct

somatosensory inputs or may be controlled by neuromodulation. This scenario resembles the situation in the AC, in which PV^+ INs are inhibited by layer 1 INs during the shock (Letzkus et al. 2011 - Fig. 3, Suppl. Fig. 10, 11). Interestingly, subgroups of both PV^+ and SOM^+ INs exhibit excitatory shock responses, which could reflect differential shock-inputs onto these subgroups. These shock-excited INs could also provide a source of inhibition for the majority of PV^+ and SOM^+ INs (Wolff et al. 2013 - ED Fig. 4).

Interneurons are also targets for neuromodulation. Several neuromodulatory systems are known to affect different subtypes of INs. Acetylcholine affects both PV^+ and SOM^+ INs, but due to differences in receptor expression, the effects may differ. The inhibitory muscarinic M2 receptor is for example only expressed in BLA SOM^+ INs, but not in PV^+ INs (McDonald and Mascagni 2011). Serotonin affects inhibition differentially as well. While all IN subtypes in the amygdala are targeted by serotonergic axon terminals (Muller et al. 2007b), the 5HT3a receptor is neither expressed by PV^+, nor by SOM^+ INs, but e.g. by VIP^+ INs (Rudy et al. 2011). This differential expression could lead to a selective activation of VIP^+ INs, which in turn may inhibit PV^+ and SOM^+ INs during the shock. Differential effects of neuromodulators on PV^+ or SOM^+ INs or effects on other INs may explain differential responses of the two IN classes (Wolff et al. 2013 - ED Fig. 4). Notably, different tone- or shock-responses in subgroups of PV^+ or SOM^+ INs may also be due to differential expression of receptors for neuromodulators.

What is the need for different disinhibitory mechanisms during the tone and the shock in the BLA and possibly in cortical areas? During the shock, the stimulus which drives associative fear learning, PNs are disinhibited along their entire somatodendritic axis. This allows strong shock-induced excitation and will assure the unselective boosting of plasticity of all tone-excited auditory inputs. In contrast, during the tone, most likely, only the dendrites are disinhibited. The concurrent increase in perisomatic inhibition may enhance the signal-to-noise ratio (see chapter 4.1). It can be assumed that disinhibition is not homogenous across the dendritic tree. Local differences may favor certain auditory inputs for potentiation over others. This may be an additional mechanism to ensure specificity of the memory trace (see chapter 4.1.1). Overall, a more refined disinhibition only of the dendrites will allow a more complex regulation of plasticity. This may be advantageous to filter out less salient or less predictive stimuli and to suppress the potentiation of the respective inputs. Controlled dendritic disinhibition may therefore be an intricate regulator of plasticity by itself, while global disinhibition during the shock ensures maximally efficient plasticity of the auditory inputs, which have been previously selected by dendritic disinhibition.

4.2 Association of non-overlapping CS and US

A puzzling phenomenon in fear learning, which at first glance contradicts the notion of Hebbian mechanisms underlying associative plasticity, is that learning occurs even if the CS and the US do not temporally overlap. Presentations of the CS and the US back-to-back or even with short delays between the CS and the US also induce reliable learning – without recruiting circuits required for trace fear conditioning (Fanselow and Poulos 2005). This is surprising, since already after a few hundred milliseconds no tone-evoked activity can be detected anymore in PNs. How can a shock, delivered after subsiding of the tone response, still induce associative plasticity? According to Hebbian models, non-overlapping CS and US, not evoking simultaneous activation, should not induce plasticity and learning (Hebb 1949; Blair et al. 2001).

This suggests that other mechanisms, beyond mere pairing of pre- and postsynaptic activity and possibly on the neuronal circuit level, play a crucial role in this form of associative plasticity. Such processes must be able to preserve the tone-evoked activity to some extent in PNs to allow plasticity upon pairing with the shock. A potential mechanism could be persistent sub-threshold membrane depolarization, outlasting the tone-induced spiking of PNs. This prolonged depolarization might be sufficient to trigger plasticity upon occurrence of US-induced activation.

The short time scale of the delay between the CS and the US likely excludes most molecular mechanisms which would retain a trace of the CS-induced activity. Synaptic tagging (Redondo and Morris 2011), based on local dendritic translation and protein expression could not account for such rapid changes at the synapse. In contrast, fast posttranslational modifications of receptors or channels may occur, which prime the auditory input synapses for plasticity. This might represent a particular short-term form of synaptic tagging and may increase the excitability of cells or render NMDARs more likely to open upon presentation of the US. Remarkably, the mechanisms underlying this phenomenon are completely elusive. Their study is especially challenging due to the short time scales and the probably highly transient nature of the involved molecular changes.

Since such processes will have an important impact on plasticity, it is likely that they are tightly regulated. They may even be affected by the same circuit mechanisms which control the tone-induced spiking activity. For the regulation of these processes on the level of neuronal circuits, inhibition is a prominent candidate – as it is for neuronal activity. A disinhibitory microcircuit in the BLA controls tone-induced activity and the strength of acquired fear memories (Wolff et al. 2013). This circuit may, via dendritic disinhibition, also regulate a prolonged sub-threshold depolarization at the synapses or molecular mechanisms, active during the CS-US delay. The level of activation of these processes could determine their duration and the available time window for plasticity induction. A boosting of these processes might therefore prolong the window of opportunity for a delayed shock

to induce plasticity. If the described disinhibitory microcircuit is involved in the control of these processes, its activity should affect the duration of the time window for effective CS-US associations. Optogenetic manipulations of the dendritic disinhibition either during the tone or during the delay may therefore prolong or shorten the available time window for this kind of associative plasticity.

In summary, experiments addressing the role of inhibition during the CS-US delay may be a crucial first step in understanding the unknown underlying mechanisms, which allow associations between non-overlapping stimuli. The results may not only resolve a long-standing question in fear conditioning, but also significantly extend our general view about Hebbian plasticity.

4.3 Selection of cells for a memory trace and memory allocation

Fear learning creates a memory trace in the neuronal fear circuits, allowing for long-term storage of the salient information and for the recall of the memory. Neurons which are recruited to this memory trace undergo synaptic and cellular plasticity - the molecular and cellular underpinnings of learning (Blair et al. 2001). However, only a fraction of the cells e.g. in the BLA is recruited to a specific memory trace (Herry et al. 2008; Zhou et al. 2009; Senn et al. 2013) – how these cells are selected and how the memory is allocated still remains elusive. Different lines of research could show that the excitability and the specific activation of PNs by the conditioned stimulus are critical factors for their plasticity and for their recruitment to the memory trace (Han et al. 2007; Han et al. 2009; Silva et al. 2009; Zhou et al. 2009; Liu et al. 2012; Kim et al. 2013; Ramirez et al. 2013). Furthermore, different experiences are represented by individual memory traces (Han et al. 2009; Ramirez et al. 2013), which allows for the specificity of memories and also of behavioral responses.

How is the allocation of memory physiologically regulated? The potential of a neuron to enter a memory trace could be hard-wired or flexibly regulated depending on the stimulus. Could a certain pool of neurons intrinsically have a higher excitability, favoring them for the recruitment to a memory trace? Alternatively, several mechanisms could contribute to both a fixed and a flexible selection of candidate neurons. For instance, the impact of neuromodulators on the excitability of certain cells may be hard-wired by selective targeting of neuromodulator release or by differential expression of the appropriate receptors. The impact could also be flexible, though, since both release and receptor expression can be regulated according to the state of the animal.

Another critical factor for including neurons in the memory trace are the received inputs. To exhibit plasticity, PNs must receive and be activated by both CS and US inputs (Blair et al. 2001), which restricts the pool of potential candidate neurons. The relative strength of different inputs will vastly vary even among neurons in this pool, which may possibly

underlie the stimulus-dependent selection of neurons for the memory trace. However, the naïve tuning-curve of BLA PNs is not a predictor for the recruitment to the memory trace, since the tuning can be heavily shifted towards the CS after conditioning (Bordi and LeDoux 1992; Rogan and LeDoux 1996). Such plasticity of non-optimally tuned neurons argues against a selection based merely on the initial strength of their inputs.

In addition, also circuit mechanisms like inhibition may control the recruitment of PNs to a memory trace (Ehrlich et al. 2009). Inhibition regulates the impact of excitatory inputs onto PNs. Differential inhibition can therefore gate plasticity in one PN, while it prevents plasticity in another PN with the same excitatory inputs, thereby selecting which cell enters the memory trace. Also inhibition may be both, hard-wired or flexible. Differential connectivity could favor certain neurons for plasticity. Inhibition may itself be regulated, though. Not only the strength of sensory inputs onto the interneurons, but also differences in neuromodulation, internal state of the animal, experience history and previous activity may shape the discrete pattern of inhibition during the acquisition of each individual memory trace and thereby select the participating neurons.

Disinhibitory mechanisms, as described in the BLA and AC (Letzkus et al. 2011; Wolff et al. 2013) may be involved in the selective recruitment of neurons to a memory trace by gating of stimulus-evoked activity. In this case, optogenetic manipulations of these microcircuits should drive neurons, which are disinhibited by the manipulation, into the memory trace. Especially cells in which neuronal responses are only uncovered by the manipulation should strongly increase their chance to be recruited (Wolff et al. 2013 - ED Fig. 7).

The allocation of memory and the selection of cells which are incorporated into a memory trace are likely controlled by a combination of the described mechanisms. A limited pool of neurons, possibly defined by their intrinsic properties and their anatomical connectivity, has the potential to undergo plastic changes during learning. Neuromodulation of the PNs and inhibitory and disinhibitory mechanisms may then add a more intricate and flexible level of control and together determine the selection of neurons driven into the memory trace.

4.3.1 Testing factors for the incorporation of neurons into a memory trace

What makes a fear cell a fear cell (Herry et al. 2008)? How is a fear memory trace created and how is the memory allocated? Is there a group of BA neurons, which are the dedicated fear cells, independent of the stimulus? This group of cells could be unselectively activated in response to each conditioned stimulus to trigger a fear reaction. The activation level of these cells may in turn determine the strength and the nature of the fear response. For a mere signaling of the initiation of a fear response, a further differentiation of fearful stimuli might not be necessary on the level of the BA.

However, a more complex selection would allow to form individual memory traces for distinct experiences. Specific memory traces with distinct sets of recruited neurons would also allow to trigger appropriate behavioral responses, according to the fearful stimulus. While in mice and other prey species fear responses are rather stereotypical (LeDoux 2000; Fanselow and Poulos 2005), it is advantageous for higher mammals to flexibly select their responses from a broad behavioral repertoire to adapt to the level of threat. There might be a pool of potential fear cells, possibly the PL-projecting neurons, from which cells are recruited, depending on the stimulus (Senn et al. 2013 - Fig. 3, Suppl. Fig. 8). The partial overlap of PL-projecting neurons and fear cells could reflect the selective recruitment of certain cells to the memory trace for the given CS, while the others became non-responsive or persistent cells (Senn et al. 2013 - Suppl. Fig. 8). Different memories may then trigger different activation patterns of groups of fear cells.

Both scenarios are a priori possible. Studying fear cells may therefore provide the opportunity to identify mechanisms for the selection of cells which enter a memory trace. The first scenario would suggest that fear cells are a distinct population of neurons, which share the projection to the PL with other cells, but are otherwise separate. In contrast, the second scenario assumes that PL-projecting neurons (or at least a subgroup) have the potential to become fear cells and that there are mechanisms in place to select cells for the memory trace, depending on the stimulus.

An important step to distinguish these two scenarios is to test whether individual fear cells are fear cells for several CSs. Herry et al. previously described cells which exhibit activation to two CSs after conditioning (Herry et al. 2008). Furthermore, they could be identified as fear cells for one of the CSs, which was selectively extinguished. However, since no extinction to the other CS was performed, they cannot unambiguously be identified as fear cells for the second CS (Herry et al. 2008). Further experiments, using conditioning and extinction to several stimuli, could reveal whether the same set of cells forms the fear memory traces for multiple CSs. In case fear cells are a distinct population of neurons, multiple molecular, anatomical and electrophysiological approaches can be employed to determine what distinguishes them from other PL-projecting neurons. If instead all PL-projecting neurons are potential fear cells, the differences cannot be hard-wired, but a flexible, stimulus-dependent regulation must be in place to select the cells entering the memory trace.

The presence and strength of sensory inputs are critical factors for the selection of neurons for a memory trace. If fear neurons are a distinct population, they might exclusively receive both tone- and shock-inputs. In contrast, in a pool of potential fear cells, input strengths might differ, favoring certain cells for the incorporation into the memory trace.

Differential inputs to fear cells may be revealed using several methods. The starting point for most of these strategies is the combinatorial viral approach described in Senn et al. (Senn et al. 2013). The delivery of CRE recombinase to PL-projecting BA neurons by a

retrograde virus allows for specific conditional expression of opsins and optogenetic cell identification. Added specificity for PL-projecting BA cells which receive defined inputs can be achieved by the additional use of FLP recombinase (Branda and Dymecki 2004) (see chapter 4.4). The FLP/FRT system can be used analogously to the CRE/LoxP system. Since there are no cross-reactions, double-conditional vectors can be used which require presence of both CRE and FLP for expression. Delivery of FLP with an anterograde virus (e.g. certain HSVs (Frampton et al. 2005)) or tracer (e.g. WGA (Holstege and Vrensen 1988)) into a potential input region, like the hippocampus, will lead to a specific coexpression of both CRE and FLP only in PL-projecting BA cells receiving hippocampal input. Double-conditional expression of ChR2 and optogenetic identification allow then to determine the functional phenotype of these cells during fear learning. Multiple input regions can be tested, revealing whether fear cells receive differential input compared to other PL-projecting neurons. However, the present lack of reliability and precision of anterograde viruses may cause inefficient, leaky or polysynaptic delivery of the second recombinase, rendering this approach problematic.

Another strategy employs the rabies virus system. After expression of CRE in PL-projecting BA neurons, the conditional EnvA/TVA system (Wall et al. 2010) can be used to specifically infect these cells with rabies and to allow the virus to jump one synapse to their direct inputs. Delivery of ChR2 via the rabies will allow to optogenetically activate the inputs to the PL-projecting neurons. Illumination of the input area with a second optical fiber will induce short-latency firing in the BA cells receiving this input. This strategy allows to uncover whether functionally different PL-projecting BA cells indeed receive separate inputs. Importantly, the distinction between direct and polysynaptic inputs is crucial for the reliability of this approach.

The most precise and informative, but possibly also most challenging approach would be to perform juxta-cellular recordings from optogenetically identified PL-projecting BA cells (Pinault 1996). Such recordings allow an electroporation with a DNA construct coding for the rabies G protein, so that a rabies virus can jump synapses exclusively from the identified cell (Wickersham et al. 2007). Importantly, with this approach, the specifically labeled inputs to an individual, functionally described PL-projecting BA neuron can be precisely identified post-hoc. In contrast to the other approaches, no prior assumptions about differential inputs are necessary here. Moreover, a juxta-cellularly recorded, identified neuron can be further characterized in terms of its gene expression profile – using e.g. immunohistochemistry or even single-cell RT-PCR. This may separate fear cells from other PL-projecting neurons, for instance based on differences in receptor or channel expression, underlying increased excitability. However, juxta-cellular recordings in behaving animals are technically very challenging and laborious. Especially the reliable identification of fear cells is critical, since juxta-cellular recordings over sufficiently long periods of time may be not even feasible.

A further crucial factor which may distinguish a fear cell population, or which may allow the flexible selection from a pool of potential fear cells is differential inhibition. Predetermined fear cells may receive specific inhibitory inputs, favoring them for plasticity. It has to be tested whether the impact of inhibition onto fear cells differs from other PL-projecting neurons, both before and after conditioning. To this end, optogenetic identification of PL-projecting fear cells, using ARCH, can be performed in mouse lines expressing FLP recombinase in defined interneuron subpopulations. FLP-dependent expression of ChR2 will allow the specific activation of these interneurons to determine their inhibitory input onto identified PL-projecting cells. Using different FLP mouse lines will allow to compare the impact of different interneuron subclasses onto fear cells and other PL-projecting neurons, possibly revealing how differential inhibition may underlie memory allocation. Importantly, inhibition may also be stimulus-dependent to select cells from a pool of candidates for incorporation in the memory trace. This could be tested if directly connected pairs of PL-projecting cells and interneurons could be recorded *in vivo* (identified via inhibitory cross-correlations) during conditioning to multiple tones. Stimulus-dependent activation of the interneuron could determine whether the PL-projecting cell is selectively recruited as a fear cell.

The described approaches may uncover whether fear cells are a distinct population. If fear cells are selected from a pool of PL-projecting neurons instead, these experiments may indicate mechanisms which underlie this selection. Dissecting the identity of fear cells (and extinction cells) in detail, may therefore provide important insights into general mechanisms of memory allocation.

4.4 Diversity of neuronal subpopulations

To understand the neuronal circuits underlying fear learning, it is necessary to dissect the role of the individual basic elements of these circuits. Powerful tools are available today to target neuronal populations which account for these building blocks.

However, the main question remains – what defines a neuronal population? Is it the molecular profile, the connectivity, the function, the developmental origin? In the most extreme case, each individual neuron could be considered as its own cell type, since no two cells are exactly the same in all respects. To address the role and organization of neuronal circuits, though, classifications have to be meaningful and remain at a tractable level. Several attempts have been undertaken to establish classifications. An important example is the Petilla group, which established a nomenclature for the description and subsequent classification of the highly diverse cortical interneurons (Ascoli et al. 2008). Despite these approaches, accurate classification of neurons remains problematic - especially, since there is apparently tremendous variability in important cellular properties, like gene expression, current densities or synaptic strength not only in cells of the same type, but even amongst

identical neurons across animals (Prinz et al. 2004; Schulz et al. 2006; Goaillard et al. 2009). Amazingly, the same circuit functions can emerge even from considerably different combinations of properties (Prinz et al. 2004). This underlines that for the dissection of neuronal circuits meaningful classifications have to be found which are adapted to the addressed question and level of detail.

Importantly, classifying neuronal populations and targeting them to study their function are two very different problems. Neuronal populations as they can be experimentally targeted today – like in this thesis - still display considerable diversity and can be comprised of several subpopulations. Technical limitations so far prevent to break down the populations to the fundamental building blocks of neuronal circuits. As of now, neuronal populations are targeted based on only one property, like a molecular marker or a specific projection. However, it is evident that there hardly is a distinct, homogenous population of cells which can be defined by just one unique attribute. Furthermore, important other characteristics, like the developmental origin of a neuron, are hardly used at all for targeting. Interneurons for example fall into a vast diversity of different subtypes, with more than 20 described classes in the hippocampus (Freund and Buzsaki 1996; Somogyi and Klausberger 2005). These classes differ in molecular markers, morphology, connectivity, subcellular targeting, cellular properties, developmental origin and function. However, today's techniques only allow to separate them based on circa 5 molecular markers, e.g. parvalbumin or somatostatin. This obviously leads to the pooling of different cell types into the same group, like PV^+ basket and PV^+ chandelier cells (Freund and Buzsaki 1996; Somogyi and Klausberger 2005).

In contrast, principal neurons in the BLA at first glance appear much more uniform, showing similar morphologies and – at least as we know so far – lacking a further differentiation based on molecular markers. However, striking differences on a functional level, most prominently the examples of fear and extinction neurons (Herry et al. 2008), strongly suggest corresponding diversity also in other underlying properties. The best handle on PN diversity so far is to define them by their projection targets. As shown in this thesis (Senn et al. 2013 - chapter 3.3) this can reveal clear relationships between projection target and function. However, it is also evident that one projection target alone does not fully define a cell type or correspond to a functional class of neurons. For example, IL-projecting cells in the BLA comprise not only extinction, but also non-responsive and persistent cells (Herry et al. 2008; Senn et al. 2013 - Suppl. Fig. 8).

This diversity even in "defined" neuronal populations is usually reflected in recordings of their physiological activity during behavior, showing different response classes (Letzkus et al. 2011; Senn et al. 2013; Wolff et al. 2013). Manipulations of the neuronal activity of such populations therefore inevitably target non-uniform groups. While behavioral effects of such manipulations usually reflect the impact of the major subgroup in the targeted population, the diversity can also cause misleading results. If subgroups of a population

have opposite roles, effects of manipulations may cancel out each other. In addition, the role of small subgroups may be missed or misinterpreted, possibly leading to incorrect ideas about circuit mechanisms.

To address the problem of diverse neuronal populations and to achieve an even more thorough and detailed dissection of neuronal circuits, it will be necessary to advance the targeting strategies for monitoring and manipulating neuronal subpopulations even further. This will allow to increase the precision of the targeting in multiple steps, depending on the questions asked. Intersectional approaches promise enhanced specificity of targeting (Yizhar et al. 2011). Key in these approaches is a tightened stringency for the conditional expression of the effector proteins (like opsins). Since intersectional approaches depend on the conditional expression of not only one but at least two molecular markers or anatomical projection targets, specificity can be considerably improved. Moreover, these approaches also allow combinations of molecular and anatomical properties for more specific targeting of neuronal populations.

While not commonly used, yet, the technical foundations for these approaches have been developed. They are mainly based on the combination of the CRE/loxP system with the FLP-FRT recombination system (Branda and Dymecki 2004). The FLP-FRT system allows manipulations analogous to the CRE/loxP system, but uses a different enzyme and different DNA recognition sites. The systems can be combined without cross-reactions, and double-conditional expression vectors with recognition sites of the two systems allow expression of the transgene only upon presence of both CRE and FLP recombinases. Since FLP can be delivered analogous to CRE, they can be flexibly combined for enhanced targeting specificity. Mouse lines expressing CRE and FLP under the control of different promoters can be crossed, leading to coexpression of the recombinases only in neurons expressing both molecular markers. Using two retrograde viruses in two different brain areas will lead to CRE and FLP co-presence only in neurons which project to both target areas. Finally, mouse lines and viruses can be combined to achieve specific coexpression in neurons, which are defined both molecularly and anatomically. Double-conditional vectors will allow expression of the effector (e.g. an opsin) exclusively in neurons which coexpress both recombinases and which are therefore defined by at least two properties. Thereby, these intersectional approaches can significantly enhance the specificity of neuronal targeting. To extend these intersectional approaches even further and to achieve specific expression in neurons defined by three or more distinct criteria, further recombinases, like Dre or other enzymes from yeast (Nern et al. 2011), could be used in addition. Moreover, the integration of specific promoters in retrograde viruses could further restrict expression to subgroups of retrogradely infected neurons, fostering targeting specificity (Yizhar et al. 2011).

A variant of these intersectional approaches allows to use the targeting of a distinct neuronal population as a further defining characteristic. Identification and manipulation of, for example, only PV^+ INs which provide inhibitory input to SOM^+ INs would be possible. This

could be achieved by using a pseudotyped rabies virus (Wall et al. 2010) in mice expressing CRE in SOM^+ INs and FLP in PV^+ INs. Using AAVs, both the TVA receptor and the rabies G protein (Wickersham et al. 2007; Wall et al. 2010) can be conditionally expressed in SOM^+ INs. This enables the rabies virus to infect SOM^+ INs and to jump a synapse to the cells providing input to the SOM^+ INs. By using an engineered rabies virus with FLP-dependent expression of its transgene, it can be ensured that e.g. ChR2 will only be expressed in PV^+ INs (which express FLP) which contact SOM^+ INs. Such an approach (and its variants) would allow to more precisely target cells not only based on the brain area they project to, but on their specific target population.

While the described intersectional approaches have the potential to considerably advance the targeting of defined neuronal populations, important caveats have to be considered. Even the use of such sophisticated approaches are likely not sufficient to disentangle all different neuronal subgroups, so that still some diversity may remain in the defined populations. In some cases even the use of two or three characteristic properties may not be enough to target completely homogenous populations. Furthermore, even cells which, based on our knowledge, would be considered as homogenous may still differ in their physiological function. Targeting according to function, or at least to activity during a certain behavior, e.g. using systems based on immediate early gene expression (Han et al. 2009; Liu et al. 2012), may provide another level of separation for neuronal populations. Crucial for all intersectional approaches is the precision and reliability of the techniques, since leaky expression would drastically reduce their value and usefulness. Finally, especially for small neuronal populations, the level of targeting precision should be chosen carefully and increased in several steps to first gain general functional insights and to subsequently refine these and learn about the details.

Overall, intersectional approaches promise to advance the dissection of neuronal circuits significantly by allowing to target their fundamental building blocks and to investigate their precise role and function.

4.5 Future and caveats of optogenetics

Optogenetics has become one of the most important techniques in neuroscience research today (Deisseroth 2010). In less than a decade, the manipulation of neuronal activity by expression of light-sensitive channels or pumps and subsequent illumination has revolutionized the way experiments in neuroscience are conducted. Today, most studies on the level of neuronal circuits employ optogenetics in one way or another.
Although the methodology has rapidly and considerably been refined, the development is still ongoing. New opsins are continuously discovered and molecular engineering constantly improves opsin properties (Yizhar et al. 2011). An important goal is to extend the portfolio with multiple excitatory and inhibitory opsins with different activation spectra. This will

foster the parallel use of different opsins in the same animal to differentially activate or inhibit distinct neuronal populations and to study their interactions. Red-shifted activation spectra and lowered activation thresholds of new opsins will also allow extra-cranial illumination for optogenetic control, as recently demonstrated (Lin et al. 2013). Also the kinetics of opsins are constantly improved, allowing for temporally even more precise control of activity (Yizhar et al. 2011).
A further direction of development are switchable opsins, like the step-function opsins (SFOs, (Berndt et al. 2009)), which can be switched on or off by brief pulses of light and allow to continuously excite (or inhibit) neurons. First steps have also been taken to extend optical control of neurons beyond mere firing. Optically activatable GPCRs and light-sensitive components of cellular signaling pathways (Airan et al. 2009) promise a more refined control over neurons and their function.
In addition, the combination of optogenetics and pharmacogenetics promises new venues for research (Shapiro et al. 2012). Pharmacogenetic approaches allow manipulations of neuronal activity on extended time scales and are less invasive, since no optical fibers are implanted. With the combination of both methodologies, different neuronal populations in the same animal can be independently controlled on different time scales. Furthermore, effects of acute and chronic manipulations can be directly compared.

Although optogenetics is a powerful, robust and widely used technique, certain caveats have to be considered. Several technical aspects have to be controlled, like opsin expression and sufficient illumination. Furthermore, the efficiency and feasibility of optogenetic control may differ between cell types and has to be tested. Certain cell types may also not tolerate strong opsin expression. Importantly, the effects of the channel/pump activation itself may be problematic for the cell and for interpretation of experimental results. The massive influx of Cl^- ions upon NpHR activation may persistently affect neuronal firing (Raimondo et al. 2012). Moreover, the high Ca^{2+} permeability of certain ChR2 variants may affect cellular signaling beyond mere firing (Kleinlogel et al. 2011).

The combination of optogenetics with simultaneous electrophysiological recordings is extremely important in dissecting neuronal circuits, since it allows to determine network effects of the applied manipulations and, even more importantly, it can be used for the identification of the targeted neurons in the recordings (Lima et al. 2009; Senn et al. 2013; Wolff et al. 2013). However, also here certain caveats, beyond light artifacts, have to be kept in mind. The parameters used for optogenetic identification are critical for reliable results. The main criterion is a significant change in neuronal activity upon illumination. Since optogenetic manipulations also affect the network, it is essential to exclude indirectly activated or inhibited neurons from the analysis (Lima et al. 2009; Cohen et al. 2012; Kvitsiani et al. 2013; Wolff et al. 2013). Commonly used for this is the latency of the light-induced response, which should be shorter for direct light effects. In addition, the jitter of light-evoked spikes should vary less in directly activated neurons (Lima et al. 2009; Cohen et

al. 2012; Kvitsiani et al. 2013). Both criteria, though, can depend on the studied cell type, on the opsin expression level and the illumination. It is therefore of utmost importance to establish reliable thresholds for the system under study. Optogenetic inhibition appears advantageous for neuronal identification, since indirect effects are less likely. However, the concerted loss of inputs to a cell may also lead to an indirect decrease in activity. Furthermore, inhibition and its latency are hard to determine in neurons with low spontaneous firing rates.
Strong optogenetic activation can induce changes in the spike waveform of neurons. In extracellular recordings this can lead to a miss of the light-evoked spikes, if they cannot be assigned to the recorded cell. Furthermore, the concerted activation of large numbers of neurons may lead to a precise temporal overlap of multiple spikes and to a merge of the recorded waveforms. Such a population spike could also not be assigned to the individual neurons and the activation would not be detected.

Besides technical concerns, there are also critical conceptual caveats. Optogenetic manipulations are widely used to demonstrate causal relationships between the activity of a defined neuronal population and a behavior. But when can causality really be proven?
The most common experimental optogenetic intervention today is the activation of a certain neuronal population during some behavioral paradigm. If the manipulation induces or enhances the studied behavior, a causal relationship is assumed. However, such gain-of-function experiments do not prove that the targeted neuronal population is indeed causally involved in the physiological behavior. Under physiological conditions, the targeted neurons may play no role in the behavior and only network effects, induced by the massive activation of these neurons, drive the observed effects. Gain-of-function experiments alone can therefore only show that the activation of a neuronal population is sufficient to induce a certain behavior, but cannot show its necessity.
In contrast, loss-of-function experiments, based on optogenetic inhibition of a neuronal population, can show the necessity of certain neurons for a behavior, but not their sufficiency. Furthermore, also optogenetic inhibition may affect behavior via indirect network mechanisms. The decrease of activity in a physiologically not involved population may disturb network function sufficiently to disrupt the studied behavior.
Due to these potential problems, bidirectional manipulations should be performed, since bidirectional effects on behavior could not be easily explained by a mere disturbance of the network function. Even more important is to combine optogenetic manipulations with electrophysiological recordings of the targeted and identified cell population during the unperturbed behavior. If the targeted population shows physiological activity in line with the observed behavioral effects, a causal relationship is highly likely.

However, all experimental results and behavioral observations gained with optogenetic manipulations have to be considered carefully. Not considered network effects may easily lead to incorrect interpretations of the role of the targeted neuronal population. Incorrectly

assumed causal relationships may lead to an oversimplification of the suggested neuronal circuits.
Therefore, it will be important in the future to combine the experimental approaches with the modeling of neuronal circuits. This would allow to determine whether the experimental results indeed fit with the suggested circuit model and whether alternative models could also explain the results.

Overall, optogenetics is one of the most powerful and influential techniques in neuroscience today and has already provided crucial insights and advances in the understanding of neuronal circuits underlying behavior. However, it is essential to keep its limitations in mind and to apply this technique carefully, considering the potential caveats.

4.6 Conclusions

Research has been tremendously successful in revealing brain areas and molecular mechanisms underlying learning and memory. However, only a shift of the perspective to neuronal circuits will allow further and deeper insight into this most fundamental function of the brain. While the study of neuronal circuits was hampered in the past by the lack of appropriate tools, today, we can take first steps to dissect these circuits and to elucidate the role of the neuronal populations which account for their building blocks.

The studies presented here take advantage of state-of-the-art techniques for the dissection of neuronal circuits. Targeting, monitoring and manipulating of defined neuronal populations is the key to identify distinct circuits for learning and for its control and regulation. Establishing and refining these techniques allowed me to reveal important general circuit mechanisms of fear learning.

I could demonstrate marked relationships between the molecular or anatomical identity of defined neuronal populations and their function in fear learning. Moreover, my studies point out that the control of learning is a highly complex process, involving multiple types of neurons, and depends on an intricate interplay in both local and distributed neuronal circuits. How general these circuit mechanisms are and whether they are recurrently used in other brain areas and different forms of learning remains to be tested.

While today's techniques, as used in the presented studies, allow circuit dissection in unprecedented detail and have already provided remarkable insights into circuit function, their further improvement is an inevitable must for an advanced understanding of learning and memory. New intersectional approaches and techniques to target really homogeneous populations of neurons will allow to identify and study the actual building blocks of neuronal circuits.

In sum, this thesis represents a significant step towards an understanding of neuronal circuits, learning and memory and the general function of the brain.

5. ACKNOWLEDGEMENTS

First of all, I would like to thank Andreas! It was an amazing time! Much better, than I would have ever imagined! Thanks for giving me the amazing opportunities here, for giving me the best project ever, for being a great mentor, for science, for help with so many things, for showing me how to do and think science. And for just giving me the chance for all this – I learned so much! Thank you!

I would really like to thank Johannes – for helping me during all these years, for being so much more than just another mentor, for great discussions, for really doing science with me, for getting me started, for the great work together – especially on the papers - and for the green fairy!

Thanks also to Philip! For so much practical help and for so many discussions about science and our work in particular! And for taking care of me! In so many ways!

A big thanks also to Cyril – for invaluable help, for showing me how to record, for discussions, for telling me which way to go.

Thank you Ingrid – for being my mentor in the beginning, for being super helpful and super tough. For telling me what is right – and what is wrong! You put me on the right track in the beginning! Thank you!

Thanks to Jan and Jonathan! For amazing help in the last years – I know, it's already quite some time, but to me it still feels like you just arrived – and you already helped me so much!

And of course – thank you Francois!! For everything!! For being there, for science – and even more for life! Thank you!

Great thanks also to Verena and Stephane – you both spend a lot of time on teaching and explaining a lot of different things to me! And for being great examples as PhD students!

Thanks to the people who also helped me with experiments – Christian, Julia, Sabine, Kristine!

Thanks for helpful comments for writing my thesis – Philip, Jan and Jonathan!

And of course, I would like to thank the whole Lüthi Lab – not only for an amazing scientific environment, but also for all the fun! This is the lab!!!

I would also like to thank a lot of people outside the lab for great scientific help! Gilad for amazing analysis, Philip Schönenberger for unlimited knowledge about opsins and great discussions, Simon, Volker, Tobias, Botond and Flavio – and finally you leave first…

I would also like to thank my thesis committee for the great help over all the years and for the efforts with my thesis, my defense and my committee meetings – Carl Petersen and Rainer Friedrich.

A big thank you also to Paul Argast for incredible work in his machine shop!

Special Thanks also to Susan Thomas – for the great time, the great laughs, many cookies and gentle reminders!

I would also like to thank my friends at the FMI and in Basel during all these years for the amazing time, I will never forget! For always being there, for making Basel my home, for a life, and for being real friends! There are so many people, I am thankful to – and I am sure, I am forgetting many – but these people here I would really like to thank for sure!

Jakob (for spending all this time in the very same flat with me), Sylvia (for starting with me and not waiting for me in the end ;)), Claudi, Eszter, Frank, Xavier (for being there from the beginning and making my first years in Basel unforgettable!), Flo, Inga, Mike, Claudius (Jever!), Angeline, Paolo, Alessia, Thomas, Lynda (from the start!), Rabih, Enrico, Lisa, Dominik, Nik, Jun and Sibel (for being great friends), and finally Lars and Aaron (for the entrance to the Biozentrum and for being real friends)!

I would also like to thank my friends from University, who are still there! – unbelievable!!! Thanks, Achim, Johannes, Stefan, Helge, Paulina, Evy, Tina!!!

And thanks to the Rhein, my living room at Flora, for the best summers ever and for all the people there! SCHWIMMEN?!

Finally, I would like to thank my family for the great support, for waiting so long for the “last doctor” and for everything – Danke!

6. REFERENCES

Acquas, E., Wilson, C. and Fibiger, H. C. (1996). Conditioned and unconditioned stimuli increase frontal cortical and hippocampal acetylcholine release: effects of novelty, habituation, and fear. *J Neurosci* **16**(9): 3089-96.

Adolphs, R. (2013). The biology of fear. *Current biology : CB* **23**(2): R79-93.

Airan, R. D., Thompson, K. R., Fenno, L. E., Bernstein, H. and Deisseroth, K. (2009). Temporally precise in vivo control of intracellular signalling. *Nature* **458**(7241): 1025-9.

Amano, T., Duvarci, S., Popa, D. and Pare, D. (2011). The fear circuit revisited: contributions of the basal amygdala nuclei to conditioned fear. *J Neurosci* **31**(43): 15481-9.

Amaral, D. G. and Insausti, R. (1992). Retrograde transport of D-[3H]-aspartate injected into the monkey amygdaloid complex. *Exp Brain Res* **88**(2): 375-88.

Ascoli, G. A., Alonso-Nanclares, L., Anderson, S. A., Barrionuevo, G., Benavides-Piccione, R., Burkhalter, A., Buzsaki, G., Cauli, B., Defelipe, J., Fairen, A., Feldmeyer, D., Fishell, G., Fregnac, Y., Freund, T. F., Gardner, D., Gardner, E. P., Goldberg, J. H., Helmstaedter, M., Hestrin, S., Karube, F., Kisvarday, Z. F., Lambolez, B., Lewis, D. A., Marin, O., Markram, H., Munoz, A., Packer, A., Petersen, C. C., Rockland, K. S., Rossier, J., Rudy, B., Somogyi, P., Staiger, J. F., Tamas, G., Thomson, A. M., Toledo-Rodriguez, M., Wang, Y., West, D. C. and Yuste, R. (2008). Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature reviews. Neuroscience* **9**(7): 557-68.

Atallah, B. V., Bruns, W., Carandini, M. and Scanziani, M. (2012). Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73**(1): 159-70.

Bacci, A., Huguenard, J. R. and Prince, D. A. (2005). Modulation of neocortical interneurons: extrinsic influences and exercises in self-control. *Trends Neurosci* **28**(11): 602-10.

Bacon, S. J., Headlam, A. J., Gabbott, P. L. and Smith, A. D. (1996). Amygdala input to medial prefrontal cortex (mPFC) in the rat: a light and electron microscope study. *Brain research* **720**(1-2): 211-9.

Bauer, E. P., Schafe, G. E. and LeDoux, J. E. (2002). NMDA receptors and L-type voltage-gated calcium channels contribute to long-term potentiation and different components of fear memory formation in the lateral amygdala. *J Neurosci* **22**(12): 5239-49.

Berendse, H. W., Galis-de Graaf, Y. and Groenewegen, H. J. (1992). Topographical organization and relationship with ventral striatal compartments of prefrontal corticostriatal projections in the rat. *J Comp Neurol* **316**(3): 314-47.

Berndt, A., Yizhar, O., Gunaydin, L. A., Hegemann, P. and Deisseroth, K. (2009). Bi-stable neural state switches. *Nat Neurosci* **12**(2): 229-34.

Berretta, S., Pantazopoulos, H., Caldera, M., Pantazopoulos, P. and Pare, D. (2005). Infralimbic cortex activation increases c-Fos expression in intercalated neurons of the amygdala. *Neuroscience* **132**(4): 943-53.

Bi, G. Q. and Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J Neurosci* **18**(24): 10464-72.

Bienvenu, T. C., Busti, D., Magill, P. J., Ferraguti, F. and Capogna, M. (2012). Cell-type-specific recruitment of amygdala interneurons to hippocampal theta rhythm and noxious stimuli in vivo. *Neuron* **74**(6): 1059-74.

Bissiere, S., Humeau, Y. and Luthi, A. (2003). Dopamine gates LTP induction in lateral amygdala by suppressing feedforward inhibition. *Nat Neurosci* **6**(6): 587-92.

Blair, H. T., Schafe, G. E., Bauer, E. P., Rodrigues, S. M. and LeDoux, J. E. (2001). Synaptic plasticity in the lateral amygdala: a cellular hypothesis of fear conditioning. *Learning & memory* **8**(5): 229-42.

Bliss, T. V. and Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J Physiol* **232**(2): 331-56.

Boatman, J. A. and Kim, J. J. (2006). A thalamo-cortico-amygdala pathway mediates auditory fear conditioning in the intact brain. *Eur J Neurosci* **24**(3): 894-900.

Bordi, F. and LeDoux, J. (1992). Sensory tuning beyond the sensory system: an initial analysis of auditory response properties of neurons in the lateral amygdaloid nucleus and overlying areas of the striatum. *J Neurosci* **12**(7): 2493-503.

Bouton, M. E. and King, D. A. (1983). Contextual control of the extinction of conditioned fear: tests for the associative value of the context. *J Exp Psychol Anim Behav Process* **9**(3): 248-65.

Branda, C. S. and Dymecki, S. M. (2004). Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. *Dev Cell* **6**(1): 7-28.

Brown, S. P. and Hestrin, S. (2009). Intracortical circuits of pyramidal neurons reflect their long-range axonal targets. *Nature* **457**(7233): 1133-6.

Burgos-Robles, A., Vidal-Gonzalez, I. and Quirk, G. J. (2009). Sustained conditioned responses in prelimbic prefrontal neurons are correlated with fear expression and extinction failure. *J Neurosci* **29**(26): 8474-82.

Cajal, S. R. (1909 - 1911). Histologie du système nerveux de l'homme et des vertébrés. Paris, Maloine.

Campeau, S. and Davis, M. (1995a). Involvement of subcortical and cortical afferents to the lateral nucleus of the amygdala in fear conditioning measured with fear-potentiated startle in rats trained concurrently with auditory and visual conditioned stimuli. *J Neurosci* **15**(3 Pt 2): 2312-27.

Campeau, S. and Davis, M. (1995b). Involvement of the central nucleus and basolateral complex of the amygdala in fear conditioning measured with fear-potentiated startle in rats trained concurrently with auditory and visual conditioned stimuli. *J Neurosci* **15**(3 Pt 2): 2301-11.

Canteras, N. S. and Swanson, L. W. (1992). Projections of the ventral subiculum to the amygdala, septum, and hypothalamus: a PHAL anterograde tract-tracing study in the rat. *J Comp Neurol* **324**(2): 180-94.

Caporale, N. and Dan, Y. (2008). Spike timing-dependent plasticity: a Hebbian learning rule. *Annu Rev Neurosci* **31**: 25-46.

Cauller, L. J., Clancy, B. and Connors, B. W. (1998). Backward cortical projections to primary somatosensory cortex in rats extend long horizontal axons in layer I. *J Comp Neurol* **390**(2): 297-310.

Chhatwal, J. P., Myers, K. M., Ressler, K. J. and Davis, M. (2005). Regulation of gephyrin and GABAA receptor binding within the amygdala after fear acquisition and extinction. *J Neurosci* **25**(2): 502-6.

Chiu, C. Q., Lur, G., Morse, T. M., Carnevale, N. T., Ellis-Davies, G. C. and Higley, M. J. (2013). Compartmentalization of GABAergic inhibition by dendritic spines. *Science* **340**(6133): 759-62.

Chow, B. Y., Han, X., Dobry, A. S., Qian, X., Chuong, A. S., Li, M., Henninger, M. A., Belfort, G. M., Lin, Y., Monahan, P. E. and Boyden, E. S. (2010). High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* **463**(7277): 98-102.

Christophe, E., Roebuck, A., Staiger, J. F., Lavery, D. J., Charpak, S. and Audinat, E. (2002). Two types of nicotinic receptors mediate an excitation of neocortical layer I interneurons. *J Neurophysiol* **88**(3): 1318-27.

Chu, Z., Galarreta, M. and Hestrin, S. (2003). Synaptic interactions of late-spiking neocortical neurons in layer 1. *J Neurosci* **23**(1): 96-102.

Ciocchi, S., Herry, C., Grenier, F., Wolff, S. B., Letzkus, J. J., Vlachos, I., Ehrlich, I., Sprengel, R., Deisseroth, K., Stadler, M. B., Muller, C. and Luthi, A. (2010). Encoding of conditioned fear in central amygdala inhibitory circuits. *Nature* **468**(7321): 277-82.

Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. and Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**(7383): 85-8.

Collins, D. R. and Pare, D. (2000). Differential fear conditioning induces reciprocal changes in the sensory responses of lateral amygdala neurons to the CS(+) and CS(-). *Learning & memory* **7**(2): 97-103.

Corcoran, K. A., Desmond, T. J., Frey, K. A. and Maren, S. (2005). Hippocampal inactivation disrupts the acquisition and contextual encoding of fear extinction. *J Neurosci* **25**(39): 8978-87.

Corcoran, K. A. and Quirk, G. J. (2007). Activity in prelimbic cortex is necessary for the expression of learned, but not innate, fears. *J Neurosci* **27**(4): 840-4.

Couey, J. J., Meredith, R. M., Spijker, S., Poorthuis, R. B., Smit, A. B., Brussaard, A. B. and Mansvelder, H. D. (2007). Distributed network actions by nicotine increase the threshold for spike-timing-dependent plasticity in prefrontal cortex. *Neuron* **54**(1): 73-87.

Cousens, G. and Otto, T. (1998). Both pre- and posttraining excitotoxic lesions of the basolateral amygdala abolish the expression of olfactory and contextual fear conditioning. *Behav Neurosci* **112**(5): 1092-103.

Davis, M. (1979). Diazepam and flurazepam: effects on conditioned fear as measured with the potentiated startle paradigm. *Psychopharmacology* **62**(1): 1-7.

Davis, M. (2000). The role of the amygdala in conditioned and unconditioned fear and anxiety. The amygdala. J. P. Aggleton, Oxford University Press**:** 213-288.

Deisseroth, K. (2010). Controlling the brain with light. *Sci Am* **303**(5): 48-55.

Delgado, J. M., Rosvold, H. E. and Looney, E. (1956). Evoking conditioned fear by electrical stimulation of subcortical structures in the monkey brain. *Journal of comparative and physiological psychology* **49**(4): 373-80.

Desgranges, B., Levy, F. and Ferreira, G. (2008). Anisomycin infusion in amygdala impairs consolidation of odor aversion memory. *Brain research* **1236**: 166-75.

Disterhoft, J. F., Coulter, D. A. and Alkon, D. L. (1986). Conditioning-specific membrane changes of rabbit hippocampal neurons measured in vitro. *Proc Natl Acad Sci U S A* **83**(8): 2733-7.

Doyere, V., Schafe, G. E., Sigurdsson, T. and LeDoux, J. E. (2003). Long-term potentiation in freely moving rats reveals asymmetries in thalamic and cortical inputs to the lateral amygdala. *Eur J Neurosci* **17**(12): 2703-15.

Duvarci, S., Popa, D. and Pare, D. (2011). Central amygdala activity during fear conditioning. *J Neurosci* **31**(1): 289-94.

Eccles, J. (1965). The Synapse. *Sci Am* **212**: 56-66.

Ehrlich, I., Humeau, Y., Grenier, F., Ciocchi, S., Herry, C. and Luthi, A. (2009). Amygdala inhibitory circuits and the control of fear memory. *Neuron* **62**(6): 757-71.

Euston, D. R., Gruber, A. J. and McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron* **76**(6): 1057-70.

Faber, E. S. and Sah, P. (2002). Physiological role of calcium-activated potassium currents in the rat lateral amygdala. *J Neurosci* **22**(5): 1618-28.

Falls, W. A., Miserendino, M. J. and Davis, M. (1992). Extinction of fear-potentiated startle: blockade by infusion of an NMDA antagonist into the amygdala. *J Neurosci* **12**(3): 854-63.

Fanselow, M. S. and Poulos, A. M. (2005). The neuroscience of mammalian associative learning. *Annu Rev Psychol* **56**: 207-34.

Farb, C. R. and LeDoux, J. E. (1997). NMDA and AMPA receptors in the lateral nucleus of the amygdala are postsynaptic to auditory thalamic afferents. *Synapse* **27**(2): 106-21.

Fishell, G. and Rudy, B. (2011). Mechanisms of inhibition within the telencephalon: "where the wild things are". *Annu Rev Neurosci* **34**: 535-67.

Floresco, S. B. and Tse, M. T. (2007). Dopaminergic regulation of inhibitory and excitatory transmission in the basolateral amygdala-prefrontal cortical pathway. *J Neurosci* **27**(8): 2045-57.

Foehring, R. C., van Brederode, J. F., Kinney, G. A. and Spain, W. J. (2002). Serotonergic modulation of supragranular neurons in rat sensorimotor cortex. *J Neurosci* **22**(18): 8238-50.

Frampton, A. R., Jr., Goins, W. F., Nakano, K., Burton, E. A. and Glorioso, J. C. (2005). HSV trafficking and development of gene therapy vectors with applications in the nervous system. *Gene Ther* **12**(11): 891-901.

Freund, T. F. and Buzsaki, G. (1996). Interneurons of the hippocampus. *Hippocampus* **6**(4): 347-470.

Freund, T. F. and Katona, I. (2007). Perisomatic inhibition. *Neuron* **56**(1): 33-42.

Froemke, R. C., Merzenich, M. M. and Schreiner, C. E. (2007). A synaptic memory trace for cortical receptive field plasticity. *Nature* **450**(7168): 425-9.

Gabbott, P. L., Warner, T. A. and Busby, S. J. (2006). Amygdala input monosynaptically innervates parvalbumin immunoreactive local circuit neurons in rat medial prefrontal cortex. *Neuroscience* **139**(3): 1039-48.

Gabbott, P. L., Warner, T. A., Jays, P. R., Salway, P. and Busby, S. J. (2005). Prefrontal cortex in the rat: projections to subcortical autonomic, motor, and limbic centers. *J Comp Neurol* **492**(2): 145-77.

Gabernet, L., Jadhav, S. P., Feldman, D. E., Carandini, M. and Scanziani, M. (2005). Somatosensory integration controlled by dynamic thalamocortical feed-forward inhibition. *Neuron* **48**(2): 315-27.

Gallistel, C. R., Fairhurst, S. and Balsam, P. (2004). The learning curve: implications of a quantitative analysis. *Proc Natl Acad Sci U S A* **101**(36): 13124-31.

Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F. and Petersen, C. C. (2010). Membrane potential dynamics of GABAergic neurons in the barrel cortex of behaving mice. *Neuron* **65**(3): 422-35.

Gentet, L. J., Kremer, Y., Taniguchi, H., Huang, Z. J., Staiger, J. F. and Petersen, C. C. (2012). Unique functional properties of somatostatin-expressing GABAergic neurons in mouse barrel cortex. *Nat Neurosci* **15**(4): 607-12.

Gerren, R. A. and Weinberger, N. M. (1983). Long term potentiation in the magnocellular medial geniculate nucleus of the anesthetized cat. *Brain research* **265**(1): 138-42.

Gil, Z., Connors, B. W. and Amitai, Y. (1997). Differential regulation of neocortical synapses by neuromodulators and activity. *Neuron* **19**(3): 679-86.

Goaillard, J. M., Taylor, A. L., Schulz, D. J. and Marder, E. (2009). Functional consequences of animal-to-animal variation in circuit parameters. *Nat Neurosci* **12**(11): 1424-30.

Gonchar, Y. and Burkhalter, A. (2003). Distinct GABAergic targets of feedforward and feedback connections between lower and higher areas of rat visual cortex. *J Neurosci* **23**(34): 10904-12.

Goosens, K. A., Hobin, J. A. and Maren, S. (2003). Auditory-evoked spike firing in the lateral amygdala and Pavlovian fear conditioning: mnemonic code or fear bias? *Neuron* **40**(5): 1013-22.

Goosens, K. A. and Maren, S. (2004). NMDA receptors are essential for the acquisition, but not expression, of conditional fear and associative spike firing in the lateral amygdala. *Eur J Neurosci* **20**(2): 537-48.

Gozzi, A., Jain, A., Giovannelli, A., Bertollini, C., Crestan, V., Schwarz, A. J., Tsetsenis, T., Ragozzino, D., Gross, C. T. and Bifone, A. (2010). A neural switch for active and passive fear. *Neuron* **67**(4): 656-66.

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annu Rev Neurosci* **31**: 359-87.

Groenewegen, H. J., Berendse, H. W., Wolters, J. G. and Lohman, A. H. (1990). The anatomical relationship of the prefrontal cortex with the striatopallidal system, the thalamus and the amygdala: evidence for a parallel organization. *Prog Brain Res* **85**: 95-116; discussion 116-8.

Grubb, M. S. and Burrone, J. (2010). Activity-dependent relocation of the axon initial segment fine-tunes neuronal excitability. *Nature* **465**(7301): 1070-4.

Han, J. H., Kushner, S. A., Yiu, A. P., Cole, C. J., Matynia, A., Brown, R. A., Neve, R. L., Guzowski, J. F., Silva, A. J. and Josselyn, S. A. (2007). Neuronal competition and selection during memory formation. *Science* **316**(5823): 457-60.

Han, J. H., Kushner, S. A., Yiu, A. P., Hsiang, H. L., Buch, T., Waisman, A., Bontempi, B., Neve, R. L., Frankland, P. W. and Josselyn, S. A. (2009). Selective erasure of a fear memory. *Science* **323**(5920): 1492-6.

Harris, J. A. and Westbrook, R. F. (1998). Evidence that GABA transmission mediates context-specific extinction of learned fear. *Psychopharmacology* **140**(1): 105-15.

Hasselmo, M. E. and Sarter, M. (2011). Modes and models of forebrain cholinergic neuromodulation of cognition. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **36**(1): 52-73.

Hattox, A. M. and Nelson, S. B. (2007). Layer V neurons in mouse cortex projecting to different targets have distinct physiological properties. *J Neurophysiol* **98**(6): 3330-40.

Haubensak, W., Kunwar, P. S., Cai, H., Ciocchi, S., Wall, N. R., Ponnusamy, R., Biag, J., Dong, H. W., Deisseroth, K., Callaway, E. M., Fanselow, M. S., Luthi, A. and Anderson, D. J. (2010). Genetic dissection of an amygdala microcircuit that gates conditioned fear. *Nature* **468**(7321): 270-6.

Hebb, D. O. (1949). The Organization of Behavior; A Neuropsychological Theory. New York, Wiley.

Heldt, S. A. and Ressler, K. J. (2007). Training-induced changes in the expression of GABAA-associated genes in the amygdala after the acquisition and extinction of Pavlovian fear. *Eur J Neurosci* **26**(12): 3631-44.

Herry, C., Bach, D. R., Esposito, F., Di Salle, F., Perrig, W. J., Scheffler, K., Luthi, A. and Seifritz, E. (2007). Processing of temporal unpredictability in human and animal amygdala. *J Neurosci* **27**(22): 5958-66.

Herry, C., Ciocchi, S., Senn, V., Demmou, L., Muller, C. and Luthi, A. (2008). Switching on and off fear by distinct neuronal circuits. *Nature* **454**(7204): 600-6.

Herry, C., Ferraguti, F., Singewald, N., Letzkus, J. J., Ehrlich, I. and Luthi, A. (2010). Neuronal circuits of fear extinction. *Eur J Neurosci* **31**(4): 599-612.

Herry, C. and Mons, N. (2004). Resistance to extinction is associated with impaired immediate early gene induction in medial prefrontal cortex and amygdala. *Eur J Neurosci* **20**(3): 781-90.

Herry, C., Trifilieff, P., Micheau, J., Luthi, A. and Mons, N. (2006). Extinction of auditory fear conditioning requires MAPK/ERK activation in the basolateral amygdala. *Eur J Neurosci* **24**(1): 261-9.

Hestrin, S. and Armstrong, W. E. (1996). Morphology and physiology of cortical neurons in layer I. *J Neurosci* **16**(17): 5290-300.

Hestrin, S., Nicoll, R. A., Perkel, D. J. and Sah, P. (1990). Analysis of excitatory synaptic action in pyramidal cells using whole-cell recording from rat hippocampal slices. *J Physiol* **422**: 203-25.

Hippenmeyer, S., Vrieseling, E., Sigrist, M., Portmann, T., Laengle, C., Ladle, D. R. and Arber, S. (2005). A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol* **3**(5): e159.

Hnasko, T. S., Perez, F. A., Scouras, A. D., Stoll, E. A., Gale, S. D., Luquet, S., Phillips, P. E., Kremer, E. J. and Palmiter, R. D. (2006). Cre recombinase-mediated restoration of nigrostriatal dopamine in dopamine-deficient mice reverses hypophagia and bradykinesia. *Proc Natl Acad Sci U S A* **103**(23): 8858-63.

Hobin, J. A., Ji, J. and Maren, S. (2006). Ventral hippocampal muscimol disrupts context-specific fear memory retrieval after extinction in rats. *Hippocampus* **16**(2): 174-82.

Holstege, J. C. and Vrensen, G. F. (1988). Anterograde tracing in the brain using autoradiography and HRP-histochemistry. A comparison at the ultrastructural level. *J Microsc* **150**(Pt 3): 233-43.

Hoover, W. B. and Vertes, R. P. (2007). Anatomical analysis of afferent projections to the medial prefrontal cortex in the rat. *Brain Struct Funct* **212**(2): 149-79.

Huang, Y. Y. and Kandel, E. R. (1998). Postsynaptic induction and PKA-dependent expression of LTP in the lateral amygdala. *Neuron* **21**(1): 169-78.

Hubener, M. and Bonhoeffer, T. (2010). Searching for engrams. *Neuron* **67**(3): 363-71.

Humeau, Y., Reisel, D., Johnson, A. W., Borchardt, T., Jensen, V., Gebhardt, C., Bosch, V., Gass, P., Bannerman, D. M., Good, M. A., Hvalby, O., Sprengel, R. and Luthi, A. (2007). A pathway-

specific function for different AMPA receptor subunits in amygdala long-term potentiation and fear conditioning. *J Neurosci* **27**(41): 10947-56.
Ji, J. and Maren, S. (2007). Hippocampal involvement in contextual modulation of fear extinction. *Hippocampus* **17**(9): 749-58.
Ji, W., Suga, N. and Gao, E. (2005). Effects of agonists and antagonists of NMDA and ACh receptors on plasticity of bat auditory system elicited by fear conditioning. *J Neurophysiol* **94**(2): 1199-211.
Jiang, X., Wang, G., Lee, A. J., Stornetta, R. L. and Zhu, J. J. (2013). The organization of two new cortical interneuronal circuits. *Nat Neurosci* **16**(2): 210-8.
Johansen, J. P., Wolff, S. B., Luthi, A. and LeDoux, J. E. (2012). Controlling the elements: an optogenetic approach to understanding the neural circuits of fear. *Biol Psychiatry* **71**(12): 1053-60.
Kandel, E. R. (2001). The molecular biology of memory storage: a dialog between genes and synapses. *Biosci Rep* **21**(5): 565-611.
Kandel, E. R. and Spencer, W. A. (1968). Cellular neurophysiological approaches in the study of learning. *Physiol Rev* **48**(1): 65-134.
Kawaguchi, Y. and Kubota, Y. (1997). GABAergic cell subtypes and their synaptic connections in rat frontal cortex. *Cereb Cortex* **7**(6): 476-86.
Kent, J. M. and Rauch, S. L. (2003). Neurocircuitry of anxiety disorders. *Curr Psychiatry Rep* **5**(4): 266-73.
Kim, J., Kwon, J. T., Kim, H. S. and Han, J. H. (2013). CREB and neuronal selection for memory trace. *Front Neural Circuits* **7**: 44.
Klausberger, T. and Somogyi, P. (2008). Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. *Science* **321**(5885): 53-7.
Kleinlogel, S., Feldbauer, K., Dempski, R. E., Fotis, H., Wood, P. G., Bamann, C. and Bamberg, E. (2011). Ultra light-sensitive and fast neuronal activation with the Ca(2)+-permeable channelrhodopsin CatCh. *Nat Neurosci* **14**(4): 513-8.
Krettek, J. E. and Price, J. L. (1978). Amygdaloid projections to subcortical structures within the basal forebrain and brainstem in the rat and cat. *J Comp Neurol* **178**(2): 225-54.
Kruglikov, I. and Rudy, B. (2008). Perisomatic GABA release and thalamocortical integration onto neocortical excitatory cells are regulated by neuromodulators. *Neuron* **58**(6): 911-24.
Kuba, H., Oichi, Y. and Ohmori, H. (2010). Presynaptic activity regulates Na(+) channel distribution at the axon initial segment. *Nature* **465**(7301): 1075-8.
Kvitsiani, D., Ranade, S., Hangya, B., Taniguchi, H., Huang, J. Z. and Kepecs, A. (2013). Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* **498**(7454): 363-6.
Lammel, S., Ion, D. I., Roeper, J. and Malenka, R. C. (2011). Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. *Neuron* **70**(5): 855-62.
Lawrence, J. J. (2008). Cholinergic control of GABA release: emerging parallels between neocortex and hippocampus. *Trends Neurosci* **31**(7): 317-27.
Le Be, J. V., Silberberg, G., Wang, Y. and Markram, H. (2007). Morphological, electrophysiological, and synaptic properties of corticocallosal pyramidal cells in the neonatal rat neocortex. *Cereb Cortex* **17**(9): 2204-13.
LeDoux, J. (1996). The Emotional Brain: The Mysterious Underpinnings of Emotional Life. New York, Simon & Schuster.
LeDoux, J. (2007). The amygdala. *Curr Biol* **17**(20): R868-74.
LeDoux, J. E. (2000). Emotion circuits in the brain. *Annu Rev Neurosci* **23**: 155-84.
LeDoux, J. E., Farb, C. and Ruggiero, D. A. (1990). Topographic organization of neurons in the acoustic thalamus that project to the amygdala. *J Neurosci* **10**(4): 1043-54.
LeDoux, J. E., Farb, C. R. and Milner, T. A. (1991). Ultrastructure and synaptic associations of auditory thalamo-amygdala projections in the rat. *Exp Brain Res* **85**(3): 577-86.

LeDoux, J. E., Iwata, J., Cicchetti, P. and Reis, D. J. (1988). Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear. *J Neurosci* **8**(7): 2517-29.

Legendy, C. R. and Salcman, M. (1985). Bursts and recurrences of bursts in the spike trains of spontaneously active striate cortex neurons. *J Neurophysiol* **53**(4): 926-39.

Letzkus, J. J., Wolff, S. B., Meyer, E. M., Tovote, P., Courtin, J., Herry, C. and Luthi, A. (2011). A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature* **480**(7377): 331-5.

Likhtik, E., Pelletier, J. G., Paz, R. and Pare, D. (2005). Prefrontal control of the amygdala. *J Neurosci* **25**(32): 7429-37.

Likhtik, E., Pelletier, J. G., Popescu, A. T. and Pare, D. (2006). Identification of basolateral amygdala projection cells and interneurons using extracellular recordings. *J Neurophysiol* **96**(6): 3257-65.

Likhtik, E., Popa, D., Apergis-Schoute, J., Fidacaro, G. A. and Pare, D. (2008). Amygdala intercalated neurons are required for expression of fear extinction. *Nature* **454**(7204): 642-5.

Lima, S. Q., Hromadka, T., Znamenskiy, P. and Zador, A. M. (2009). PINP: a new method of tagging neuronal populations for identification during in vivo electrophysiological recording. *PLoS ONE* **4**(7): e6099.

Lin, J. Y., Knutsen, P. M., Muller, A., Kleinfeld, D. and Tsien, R. Y. (2013). ReaChR: a red-shifted variant of channelrhodopsin enables deep transcranial optogenetic excitation. *Nat Neurosci* **16**(10): 1499-508.

Lisman, J. E. (1997). Bursts as a unit of neural information: making unreliable synapses reliable. *Trends Neurosci* **20**(1): 38-43.

Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K. and Tonegawa, S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* **484**(7394): 381-5.

Lovett-Barron, M., Turi, G. F., Kaifosh, P., Lee, P. H., Bolze, F., Sun, X. H., Nicoud, J. F., Zemelman, B. V., Sternson, S. M. and Losonczy, A. (2012). Regulation of neuronal input transformations by tunable dendritic inhibition. *Nat Neurosci* **15**(3): 423-30, S1-3.

Lu, K. T., Walker, D. L. and Davis, M. (2001). Mitogen-activated protein kinase cascade in the basolateral nucleus of amygdala is involved in extinction of fear-potentiated startle. *J Neurosci* **21**(16): RC162.

Lynch, G. S., Dunwiddie, T. and Gribkoff, V. (1977). Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. *Nature* **266**(5604): 737-9.

Malenka, R. C. and Bear, M. F. (2004). LTP and LTD: an embarrassment of riches. *Neuron* **44**(1): 5-21.

Manookin, M. B., Beaudoin, D. L., Ernst, Z. R., Flagel, L. J. and Demb, J. B. (2008). Disinhibition combines with excitation to extend the operating range of the OFF visual pathway in daylight. *J Neurosci* **28**(16): 4136-50.

Maren, S. (1999). Neurotoxic basolateral amygdala lesions impair learning and memory but not the performance of conditional fear in rats. *J Neurosci* **19**(19): 8696-703.

Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. *Annual review of neuroscience* **24**: 897-931.

Maren, S. and Quirk, G. J. (2004). Neuronal signalling of fear memory. *Nat Rev Neurosci* **5**(11): 844-52.

Maren, S., Yap, S. A. and Goosens, K. A. (2001). The amygdala is essential for the development of neuronal plasticity in the medial geniculate nucleus during auditory fear conditioning in rats. *J Neurosci* **21**(6): RC135.

Margrie, T. W., Meyer, A. H., Caputi, A., Monyer, H., Hasan, M. T., Schaefer, A. T., Denk, W. and Brecht, M. (2003). Targeted whole-cell recordings in the mammalian brain in vivo. *Neuron* **39**(6): 911-8.

Markram, H., Lubke, J., Frotscher, M. and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **275**(5297): 213-5.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G. and Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nat Rev Neurosci* **5**(10): 793-807.

Martin, S. J. and Clark, R. E. (2007). The rodent hippocampus and spatial memory: from synapses to systems. *Cell Mol Life Sci* **64**(4): 401-31.

Martin, S. J., Grimwood, P. D. and Morris, R. G. (2000). Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu Rev Neurosci* **23**: 649-711.

Martin, S. J. and Morris, R. G. (2002). New life in an old idea: the synaptic plasticity and memory hypothesis revisited. *Hippocampus* **12**(5): 609-36.

McDonald, A. J. (1982). Neurons of the lateral and basolateral amygdaloid nuclei: a Golgi study in the rat. *The Journal of comparative neurology* **212**(3): 293-312.

McDonald, A. J. (1984). Neuronal organization of the lateral and basolateral amygdaloid nuclei in the rat. *J Comp Neurol* **222**(4): 589-606.

McDonald, A. J. (1998). Cortical pathways to the mammalian amygdala. *Prog Neurobiol* **55**(3): 257-332.

McDonald, A. J. and Betette, R. L. (2001). Parvalbumin-containing neurons in the rat basolateral amygdala: morphology and co-localization of Calbindin-D(28k). *Neuroscience* **102**(2): 413-25.

McDonald, A. J. and Mascagni, F. (2001). Colocalization of calcium-binding proteins and GABA in neurons of the rat basolateral amygdala. *Neuroscience* **105**(3): 681-93.

McDonald, A. J. and Mascagni, F. (2011). Neuronal localization of M2 muscarinic receptor immunoreactivity in the rat amygdala. *Neuroscience* **196**: 49-65.

McDonald, A. J., Mascagni, F. and Guo, L. (1996). Projections of the medial and lateral prefrontal cortices to the amygdala: a Phaseolus vulgaris leucoagglutinin study in the rat. *Neuroscience* **71**(1): 55-75.

McKernan, M. G. and Shinnick-Gallagher, P. (1997). Fear conditioning induces a lasting potentiation of synaptic currents in vitro. *Nature* **390**(6660): 607-11.

Mechawar, N., Cozzari, C. and Descarries, L. (2000). Cholinergic innervation in adult rat cerebral cortex: a quantitative immunocytochemical description. *J Comp Neurol* **428**(2): 305-18.

Mermelstein, P. G., Bito, H., Deisseroth, K. and Tsien, R. W. (2000). Critical dependence of cAMP response element-binding protein phosphorylation on L-type calcium channels supports a selective response to EPSPs in preference to action potentials. *J Neurosci* **20**(1): 266-73.

Metherate, R. (2004). Nicotinic acetylcholine receptors in sensory cortex. *Learning & memory* **11**(1): 50-9.

Milad, M. R. and Quirk, G. J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature* **420**(6911): 70-4.

Morgan, M. A., Romanski, L. M. and LeDoux, J. E. (1993). Extinction of emotional learning: contribution of medial prefrontal cortex. *Neurosci Lett* **163**(1): 109-13.

Morishima, M. and Kawaguchi, Y. (2006). Recurrent connection patterns of corticostriatal pyramidal cells in frontal cortex. *J Neurosci* **26**(16): 4394-405.

Muller, J. F., Mascagni, F. and McDonald, A. J. (2005). Coupled networks of parvalbumin-immunoreactive interneurons in the rat basolateral amygdala. *J Neurosci* **25**(32): 7366-76.

Muller, J. F., Mascagni, F. and McDonald, A. J. (2006). Pyramidal cells of the rat basolateral amygdala: synaptology and innervation by parvalbumin-immunoreactive interneurons. *J Comp Neurol* **494**(4): 635-50.

Muller, J. F., Mascagni, F. and McDonald, A. J. (2007a). Postsynaptic targets of somatostatin-containing interneurons in the rat basolateral amygdala. *J Comp Neurol* **500**(3): 513-29.

Muller, J. F., Mascagni, F. and McDonald, A. J. (2007b). Serotonin-immunoreactive axon terminals innervate pyramidal cells and interneurons in the rat basolateral amygdala. *J Comp Neurol* **505**(3): 314-35.

Myers, K. M. and Davis, M. (2004). AX+, BX- discrimination learning in the fear-potentiated startle paradigm: possible relevance to inhibitory fear learning in extinction. *Learning & memory (Cold Spring Harbor, N.Y* **11**(4): 464-75.

Myers, K. M. and Davis, M. (2007). Mechanisms of fear extinction. *Molecular psychiatry* **12**(2): 120-50.

Nern, A., Pfeiffer, B. D., Svoboda, K. and Rubin, G. M. (2011). Multiple new site-specific recombinases for use in manipulating animal genomes. *Proc Natl Acad Sci U S A* **108**(34): 14198-203.

Neves, G., Cooke, S. F. and Bliss, T. V. (2008). Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. *Nat Rev Neurosci* **9**(1): 65-75.

Nimmerjahn, A., Kirchhoff, F., Kerr, J. N. and Helmchen, F. (2004). Sulforhodamine 101 as a specific marker of astroglia in the neocortex in vivo. *Nature methods* **1**(1): 31-7.

Olveczky, B. P. (2011). Motoring ahead with rodents. *Curr Opin Neurobiol* **21**(4): 571-8.

Pape, H. C. and Pare, D. (2010). Plastic synaptic networks of the amygdala for the acquisition, expression, and extinction of conditioned fear. *Physiol Rev* **90**(2): 419-63.

Pare, D. and Gaudreau, H. (1996). Projection cells and interneurons of the lateral and basolateral amygdala: distinct firing patterns and differential relation to theta and delta rhythms in conscious cats. *J Neurosci* **16**(10): 3334-50.

Pare, D. and Smith, Y. (1993). The intercalated cell masses project to the central and medial nuclei of the amygdala in cats. *Neuroscience* **57**(4): 1077-90.

Peters, J., Kalivas, P. W. and Quirk, G. J. (2009). Extinction circuits for fear and addiction overlap in prefrontal cortex. *Learning & memory* **16**(5): 279-88.

Phelps, E. A., Delgado, M. R., Nearing, K. I. and LeDoux, J. E. (2004). Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* **43**(6): 897-905.

Pikkarainen, M., Ronkko, S., Savander, V., Insausti, R. and Pitkanen, A. (1999). Projections from the lateral, basal, and accessory basal nuclei of the amygdala to the hippocampal formation in rat. *J Comp Neurol* **403**(2): 229-60.

Pinard, C. R., Mascagni, F. and McDonald, A. J. (2012). Medial prefrontal cortical innervation of the intercalated nuclear region of the amygdala. *Neuroscience* **205**: 112-24.

Pinault, D. (1996). A novel single-cell staining procedure performed in vivo under electrophysiological control: morpho-functional features of juxtacellularly labeled thalamic cells and other central neurons with biocytin or Neurobiotin. *Journal of neuroscience methods* **65**(2): 113-36.

Pitkanen, A., Pikkarainen, M., Nurminen, N. and Ylinen, A. (2000). Reciprocal connections between the amygdala and the hippocampal formation, perirhinal cortex, and postrhinal cortex in rat. A review. *Ann N Y Acad Sci* **911**: 369-91.

Pitkanen, A., Savander, V. and LeDoux, J. E. (1997). Organization of intra-amygdaloid circuitries in the rat: an emerging framework for understanding functions of the amygdala. *Trends Neurosci* **20**(11): 517-23.

Prinz, A. A., Bucher, D. and Marder, E. (2004). Similar network activity from disparate circuit parameters. *Nat Neurosci* **7**(12): 1345-52.

Quirk, G. J. (2002). Memory for extinction of conditioned fear is long-lasting and persists following spontaneous recovery. *Learning & memory* **9**(6): 402-7.

Quirk, G. J., Armony, J. L. and LeDoux, J. E. (1997). Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. *Neuron* **19**(3): 613-24.

Quirk, G. J., Garcia, R. and Gonzalez-Lima, F. (2006). Prefrontal mechanisms in extinction of conditioned fear. *Biol Psychiatry* **60**(4): 337-43.

Quirk, G. J. and Mueller, D. (2008). Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology* **33**(1): 56-72.

Quirk, G. J., Repa, C. and LeDoux, J. E. (1995). Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: parallel recordings in the freely behaving rat. *Neuron* **15**(5): 1029-39.

Quirk, G. J., Russo, G. K., Barron, J. L. and Lebron, K. (2000). The role of ventromedial prefrontal cortex in the recovery of extinguished fear. *J Neurosci* **20**(16): 6225-31.

Raimondo, J. V., Kay, L., Ellender, T. J. and Akerman, C. J. (2012). Optogenetic silencing strategies differ in their effects on inhibitory synaptic transmission. *Nat Neurosci* **15**(8): 1102-4.

Rainnie, D. G., Mania, I., Mascagni, F. and McDonald, A. J. (2006). Physiological and morphological characterization of parvalbumin-containing interneurons of the rat basolateral amygdala. *J Comp Neurol* **498**(1): 142-61.

Ramirez, S., Liu, X., Lin, P. A., Suh, J., Pignatelli, M., Redondo, R. L., Ryan, T. J. and Tonegawa, S. (2013). Creating a false memory in the hippocampus. *Science* **341**(6144): 387-91.

Redondo, R. L. and Morris, R. G. (2011). Making memories last: the synaptic tagging and capture hypothesis. *Nature reviews. Neuroscience* **12**(1): 17-30.

Rescorla, R. A. (2001). Experimental extinction. Handbook of Contemporary Learning Theories. R. R. Mowrer and S. Klein. NJ, Erlbaum: Mahwah**:** 119-154.

Rescorla, R. A. and Heth, C. D. (1975). Reinstatement of fear to an extinguished conditioned stimulus. *J Exp Psychol Anim Behav Process* **1**(1): 88-96.

Rodrigues, S. M., Schafe, G. E. and LeDoux, J. E. (2001). Intra-amygdala blockade of the NR2B subunit of the NMDA receptor disrupts the acquisition but not the expression of fear conditioning. *J Neurosci* **21**(17): 6889-96.

Rogan, M. T. and LeDoux, J. E. (1995). LTP is accompanied by commensurate enhancement of auditory-evoked responses in a fear conditioning circuit. *Neuron* **15**(1): 127-36.

Rogan, M. T. and LeDoux, J. E. (1996). Emotion: systems, cells, synaptic plasticity. *Cell* **85**(4): 469-75.

Rogan, M. T., Staubli, U. V. and LeDoux, J. E. (1997). Fear conditioning induces associative long-term potentiation in the amygdala. *Nature* **390**(6660): 604-7.

Romanski, L. M., Clugnet, M. C., Bordi, F. and LeDoux, J. E. (1993). Somatosensory and auditory convergence in the lateral nucleus of the amygdala. *Behav Neurosci* **107**(3): 444-50.

Romanski, L. M. and LeDoux, J. E. (1992). Bilateral destruction of neocortical and perirhinal projection targets of the acoustic thalamus does not disrupt auditory fear conditioning. *Neurosci Lett* **142**(2): 228-32.

Rosen, J. B., Hitchcock, J. M., Miserendino, M. J., Falls, W. A., Campeau, S. and Davis, M. (1992). Lesions of the perirhinal cortex but not of the frontal, medial prefrontal, visual, or insular cortex block fear-potentiated startle using a visual conditioned stimulus. *J Neurosci* **12**(12): 4624-33.

Rosenkranz, J. A. and Grace, A. A. (2002). Dopamine-mediated modulation of odour-evoked amygdala potentials during pavlovian conditioning. *Nature* **417**(6886): 282-7.

Rosenkranz, J. A., Moore, H. and Grace, A. A. (2003). The prefrontal cortex regulates lateral amygdala neuronal plasticity and responses to previously conditioned stimuli. *J Neurosci* **23**(35): 11054-64.

Rubehn, B., Wolff, S. B., Tovote, P., Luthi, A. and Stieglitz, T. (2013). A polymer-based neural microimplant for optogenetic applications: design and first in vivo study. *Lab Chip* **13**(4): 579-88.

Rubio-Garrido, P., Perez-de-Manzo, F., Porrero, C., Galazo, M. J. and Clasca, F. (2009). Thalamic input to distal apical dendrites in neocortical layer 1 is massive and highly convergent. *Cereb Cortex* **19**(10): 2380-95.

Rudy, B., Fishell, G., Lee, S. and Hjerling-Leffler, J. (2011). Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev Neurobiol* **71**(1): 45-61.

Rumpel, S., LeDoux, J., Zador, A. and Malinow, R. (2005). Postsynaptic receptor trafficking underlying a form of associative learning. *Science* **308**(5718): 83-8.

Saar, D., Grossman, Y. and Barkai, E. (1998). Reduced after-hyperpolarization in rat piriform cortex pyramidal neurons is associated with increased learning capability during operant conditioning. *Eur J Neurosci* **10**(4): 1518-23.

Sah, P., Faber, E. S., Lopez De Armentia, M. and Power, J. (2003). The amygdaloid complex: anatomy and physiology. *Physiol Rev* **83**(3): 803-34.

Sah, P., Westbrook, R. F. and Luthi, A. (2008). Fear conditioning and long-term potentiation in the amygdala: what really is the connection? *Ann N Y Acad Sci* **1129**: 88-95.

Santini, E., Quirk, G. J. and Porter, J. T. (2008). Fear conditioning and extinction differentially modify the intrinsic excitability of infralimbic neurons. *J Neurosci* **28**(15): 4028-36.

Savander, V., Miettinen, R., Ledoux, J. E. and Pitkanen, A. (1997). Lateral nucleus of the rat amygdala is reciprocally connected with basal and accessory basal nuclei: a light and electron microscopic study. *Neuroscience* **77**(3): 767-81.

Schafe, G. E., Nader, K., Blair, H. T. and LeDoux, J. E. (2001). Memory consolidation of Pavlovian fear conditioning: a cellular and molecular perspective. *Trends Neurosci* **24**(9): 540-6.

Schulz, D. J., Goaillard, J. M. and Marder, E. (2006). Variable channel expression in identified single and electrically coupled neurons in different animals. *Nat Neurosci* **9**(3): 356-62.

Senn, V., Wolff, S. B., Herry, C., Grenier, F., Ehrlich, I., Gründemann, J., Fadok, J. P., Müller, C., Letzkus, J. J. and Luthi, A. (2013). Long-range connectivity defines behavioral specificity of amygdala neurons. *Neuron, accepted*.

Sesack, S. R., Deutch, A. Y., Roth, R. H. and Bunney, B. S. (1989). Topographical organization of the efferent projections of the medial prefrontal cortex in the rat: an anterograde tract-tracing study with Phaseolus vulgaris leucoagglutinin. *J Comp Neurol* **290**(2): 213-42.

Shapiro, M. G., Frazier, S. J. and Lester, H. A. (2012). Unparalleled control of neural activity using orthogonal pharmacogenetics. *ACS Chem Neurosci* **3**(8): 619-29.

Sierra-Mercado, D., Padilla-Coreano, N. and Quirk, G. J. (2011). Dissociable roles of prelimbic and infralimbic cortices, ventral hippocampus, and basolateral amygdala in the expression and extinction of conditioned fear. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **36**(2): 529-38.

Silva, A. J., Zhou, Y., Rogerson, T., Shobe, J. and Balaji, J. (2009). Molecular and cellular approaches to memory allocation in neural circuits. *Science* **326**(5951): 391-5.

Sjostrom, P. J., Rancz, E. A., Roth, A. and Hausser, M. (2008). Dendritic excitability and synaptic plasticity. *Physiol Rev* **88**(2): 769-840.

Somogyi, P. and Klausberger, T. (2005). Defined types of cortical interneurone structure space and spike timing in the hippocampus. *J Physiol* **562**(Pt 1): 9-26.

Sotres-Bayon, F., Bush, D. E. and LeDoux, J. E. (2007). Acquisition of fear extinction requires activation of NR2B-containing NMDA receptors in the lateral amygdala. *Neuropsychopharmacology* **32**(9): 1929-40.

Sotres-Bayon, F., Cain, C. K. and LeDoux, J. E. (2006). Brain mechanisms of fear extinction: historical perspectives on the contribution of prefrontal cortex. *Biol Psychiatry* **60**(4): 329-36.

Sotres-Bayon, F., Sierra-Mercado, D., Pardilla-Delgado, E. and Quirk, G. J. (2012). Gating of fear in prelimbic cortex by hippocampal and amygdala inputs. *Neuron* **76**(4): 804-12.

Spampanato, J., Polepalli, J. and Sah, P. (2011). Interneurons in the basolateral amygdala. *Neuropharmacology* **60**(5): 765-73.

Stosiek, C., Garaschuk, O., Holthoff, K. and Konnerth, A. (2003). In vivo two-photon calcium imaging of neuronal networks. *Proc Natl Acad Sci U S A* **100**(12): 7319-24.

Stuart, G. J. and Hausser, M. (2001). Dendritic coincidence detection of EPSPs and action potentials. *Nat Neurosci* **4**(1): 63-71.

Suga, N. and Ma, X. (2003). Multiparametric corticofugal modulation and plasticity in the auditory system. *Nature reviews. Neuroscience* **4**(10): 783-94.

Tang, W., Ehrlich, I., Wolff, S. B., Michalski, A. M., Wolfl, S., Hasan, M. T., Luthi, A. and Sprengel, R. (2009). Faithful expression of multiple proteins via 2A-peptide self-processing: a versatile and reliable method for manipulating brain circuits. *J Neurosci* **29**(27): 8621-9.

Taniguchi, H., He, M., Wu, P., Kim, S., Paik, R., Sugino, K., Kvitsiani, D., Fu, Y., Lu, J., Lin, Y., Miyoshi, G., Shima, Y., Fishell, G., Nelson, S. B. and Huang, Z. J. (2011). A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71**(6): 995-1013.

Tsvetkov, E., Carlezon, W. A., Benes, F. M., Kandel, E. R. and Bolshakov, V. Y. (2002). Fear conditioning occludes LTP-induced presynaptic enhancement of synaptic transmission in the cortical pathway to the lateral amygdala. *Neuron* **34**(2): 289-300.

Turner, B. H. and Herkenham, M. (1991). Thalamoamygdaloid projections in the rat: a test of the amygdala's role in sensory processing. *J Comp Neurol* **313**(2): 295-325.

Uylings, H. B. and van Eden, C. G. (1990). Qualitative and quantitative comparison of the prefrontal cortex in rat and in primates, including humans. *Prog Brain Res* **85**: 31-62.

Varga, C., Lee, S. Y. and Soltesz, I. (2010). Target-selective GABAergic control of entorhinal cortex output. *Nat Neurosci* **13**(7): 822-4.

Veening, J. G., Swanson, L. W. and Sawchenko, P. E. (1984). The organization of projections from the central nucleus of the amygdala to brainstem sites involved in central autonomic regulation: a combined retrograde transport-immunohistochemical study. *Brain research* **303**(2): 337-57.

Veinante, P. and Freund-Mercier, M. J. (1998). Intrinsic and extrinsic connections of the rat central extended amygdala: an in vivo electrophysiological study of the central amygdaloid nucleus. *Brain research* **794**(2): 188-98.

Vertes, R. P. (2004). Differential projections of the infralimbic and prelimbic cortex in the rat. *Synapse* **51**(1): 32-58.

Vidal-Gonzalez, I., Vidal-Gonzalez, B., Rauch, S. L. and Quirk, G. J. (2006). Microstimulation reveals opposing influences of prelimbic and infralimbic cortex on the expression of conditioned fear. *Learning & memory (Cold Spring Harbor, N.Y* **13**(6): 728-33.

Vogels, T. P. and Abbott, L. F. (2009). Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nat Neurosci* **12**(4): 483-91.

Waclaw, R. R., Ehrman, L. A., Pierani, A. and Campbell, K. (2010). Developmental origin of the neuronal subtypes that comprise the amygdalar fear circuit in the mouse. *J Neurosci* **30**(20): 6944-53.

Walker, D. L., Ressler, K. J., Lu, K. T. and Davis, M. (2002). Facilitation of conditioned fear extinction by systemic administration or intra-amygdala infusions of D-cycloserine as assessed with fear-potentiated startle in rats. *J Neurosci* **22**(6): 2343-51.

Wall, N. R., Wickersham, I. R., Cetin, A., De La Parra, M. and Callaway, E. M. (2010). Monosynaptic circuit tracing in vivo through Cre-dependent targeting and complementation of modified rabies virus. *Proc Natl Acad Sci U S A* **107**(50): 21848-53.

Weinberger, N. M. (2007a). Associative representational plasticity in the auditory cortex: a synthesis of two disciplines. *Learning & memory* **14**(1-2): 1-16.

Weinberger, N. M. (2007b). Auditory associative memory and representational plasticity in the primary auditory cortex. *Hear Res* **229**(1-2): 54-68.

Weiskrantz, L. (1956). Behavioral changes associated with ablation of the amygdaloid complex in monkeys. *Journal of comparative and physiological psychology* **49**(4): 381-91.

Wenk, G. L. (1997). The nucleus basalis magnocellularis cholinergic system: one hundred years of progress. *Neurobiol Learn Mem* **67**(2): 85-95.

Whitlock, J. R., Heynen, A. J., Shuler, M. G. and Bear, M. F. (2006). Learning induces long-term potentiation in the hippocampus. *Science* **313**(5790): 1093-7.

Wickersham, I. R., Lyon, D. C., Barnard, R. J., Mori, T., Finke, S., Conzelmann, K. K., Young, J. A. and Callaway, E. M. (2007). Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53**(5): 639-47.

Wilensky, A. E., Schafe, G. E. and LeDoux, J. E. (1999). Functional inactivation of the amygdala before but not after auditory fear conditioning prevents memory formation. *J Neurosci* **19**(24): RC48.

Wolff, S. B., Gründemann, J., Tovote, P., Jacobson, G. A., Müller, C., Herry, C., Ehrlich, I., Friedrich, R. W., Letzkus, J. J. and Luthi, A. (2013). Distinct subtypes of amygdala interneurons control fear learning through stimulus-specific disinhibition. *Nature, under review*.

Woodruff, A. R. and Sah, P. (2007a). Inhibition and synchronization of basal amygdala principal neuron spiking by parvalbumin-positive interneurons. *J Neurophysiol* **98**(5): 2956-61.

Woodruff, A. R. and Sah, P. (2007b). Networks of parvalbumin-positive interneurons in the basolateral amygdala. *J Neurosci* **27**(3): 553-63.

Woodson, W., Farb, C. R. and Ledoux, J. E. (2000). Afferents from the auditory thalamus synapse on inhibitory interneurons in the lateral nucleus of the amygdala. *Synapse* **38**(2): 124-37.

Wozny, C. and Williams, S. R. (2011). Specificity of synaptic connectivity between layer 1 inhibitory interneurons and layer 2/3 pyramidal neurons in the rat neocortex. *Cereb Cortex* **21**(8): 1818-26.

Wu, J. and Hablitz, J. J. (2005). Cooperative activation of D1 and D2 dopamine receptors enhances a hyperpolarization-activated inward current in layer I interneurons. *J Neurosci* **25**(27): 6322-8.

Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M. and Deisseroth, K. (2011). Optogenetics in neural systems. *Neuron* **71**(1): 9-34.

Zhang, F., Aravanis, A. M., Adamantidis, A., de Lecea, L. and Deisseroth, K. (2007). Circuit-breakers: optical technologies for probing neural signals and systems. *Nat Rev Neurosci* **8**(8): 577-81.

Zhang, W. and Linden, D. J. (2003). The other side of the engram: experience-driven changes in neuronal intrinsic excitability. *Nature reviews. Neuroscience* **4**(11): 885-900.

Zhou, F. M. and Hablitz, J. J. (1996). Morphological properties of intracellularly labeled layer I neurons in rat neocortex. *J Comp Neurol* **376**(2): 198-213.

Zhou, Y., Won, J., Karlsson, M. G., Zhou, M., Rogerson, T., Balaji, J., Neve, R., Poirazi, P. and Silva, A. J. (2009). CREB regulates excitability and the allocation of memory to subsets of neurons in the amygdala. *Nat Neurosci* **12**(11): 1438-43.