Cortical circuits for goal-directed cross-modal transfer learning

Maëlle Guyoton[1]†, Giulio Matteucci[1]†, Charlie G. Foucher [1] & Sami El-Boustani[1]*

[1] Department of Basic Neurosciences, Faculty of Medicine, University of Geneva, 1 Rue Michel-Servet, 1206 Geneva, Switzerland.

* Corresponding author e-mail: Sami El-Boustani (sami.el-boustani@unige.ch)

† These authors contributed equally to this work.

Abstract

In an environment full of complex multisensory stimuli, flexible and effective behaviors rely on our ability to transfer learned associations across sensory modalities. Here we explored the intertwined cortical representations of visual and whisker tactile sensations in mice and their role in cross-modal transfer learning. Mice trained to discriminate stimulations of two different whiskers seamlessly switched to the discrimination of two visual cues only when reward contingencies were spatially congruent across modalities. Using multi-scale calcium imaging over the dorsal cortex, we identified two distinct associative domains within the ventral and dorsal streams displaying visuo-tactile integration. We observed multimodal spatial congruency in visuo-tactile areas, both functionally and anatomically, for feedforward and feedback projections with primary sensory regions. Single-cell responses in these domains were tuned to congruent visuo-tactile stimuli. Suppressing synaptic transmission specifically in the dorsal stream impaired transfer learning. Our results delineate the pivotal cortical pathway necessary for visuo-tactile multisensory integration and goal-directed cross-modal transfer learning.

One Sentence Summary: Spatially organized representations of visual and tactile inputs in associative cortical areas facilitate effective cross-modal transfer following goal-directed sensorimotor learning.

Keywords: Multisensory integration, Associative cortices, Learning, Mice, Calcium imaging, Visual system, Somatosensory system, Cross-modal transfer, Goal-directed behavior

Introduction

Objects and events in the environment possess distinct physical properties, detectable through various sensory modalities. Specific brain circuits perform multisensory integration, resulting in a unified perception of these properties. Stimuli originating from the same spatial location and occurring simultaneously are likely to be attributed to a common cause (*1*). This is particularly true for the visual and haptic senses that share common inputs in the peri-personal space. Consequently, one can immediately recognize by visual inspection an object previously explored only by touch. This phenomenon is referred to as cross-modal object recognition and has been the focus of extensive studies in non-human primates and rodents (*2*).

Cross-modal transfer learning occurs when object or event recognition enables generalization of a learned behavior from one modality to another modality. Cross-modal transfer learning has been reported in rats for intensity or duration discrimination of auditory and visual stimuli (*3–6*). Additionally, lesions of the posterior parietal cortex in rats indicate that cross-modal transfer learning based on spatial information is impaired (*7*) and that cross-modal object recognition based on visual or tactile information during spontaneous exploration is abolished (*8*). Yet, the intricate functional organization underpinning these cognitive faculties still stands elusive.

Specialized brain regions generate multimodal representations, with aligned functional organizations for specific sensory features. Such architecture enables the nervous system to flexibly adapt to changing or noisy sensory scenes. Multisensory integration in the superior colliculus guides reflexive behaviors such as gaze and head orienting by augmenting the saliency of multisensory events via aligned topographic spatial representations of visual, auditory, and tactile stimuli (*9, 10*). In contrast, cortical

circuits play a central role in conscious perception and goal-directed behaviors. Neuroanatomical investigations in monkeys have indicated that multisensory integration might take place at various levels of the cortical processing hierarchy, including bidirectional interactions between early and late stages of processing (*11, 12*). Yet, the challenge lies in systematically characterizing these circuits at large scales while maintaining cellular resolution and linking their representations to behavior.

The mouse model offers powerful tools for dissecting circuits involved in multisensory integration and their contributions to behavioral outputs (*13–15*). Mice use tactile sensations transmitted by their large whiskers on the snout to explore their immediate surroundings. They rely on visuo-tactile information for detecting and crossing gaps (*16*), navigating in confined spaces (*17*), and recognizing object features (*18*). Given that whiskers occupy a significant part of the visual field, these two systems receive numerous concomitant inputs. The integration of whisker and visual information has been reported in the superior colliculus (*19–21*) and in RL, an associative cortical area rostro-lateral to V1 (*22, 23*). RL is part of a network of higher-visual areas situated between the primary visual and somatosensory cortex, some of which belong to the posterior parietal cortex known for its role in spatial reasoning in humans (*24, 25*). Nonetheless, the interaction between the anatomical and functional organization of cortical areas for visuo-tactile processing, and their potential role in cross-modal transfer learning, remains unresolved. In this study, we investigate these circuits in mice performing goal-directed behaviors.

Results

We designed a behavioral paradigm to test the ability of mice to transfer whisker sensorimotor associations they had previously learned to the visual modality, using a common spatial feature of stimuli (Fig. 1A). In the dark, head-fixed and water-restricted mice were first trained on a Go/No go tactile task, where they discriminated between two whiskers from the same column on the whisker pad, positioned along the ventro-dorsal axis of the snout. Mice could obtain a drop of water reward if they licked a spout upon stimulation of the top (B2) whisker whereas they were punished with a 10-second-long timeout if they licked for the bottom (C2) whisker. Once mice became expert at the task, performing stably with high percentage of correct trials over at least 3 consecutive sessions, we switched the task to a Go/No go visual discrimination task. In this condition we replaced top and bottom whisker stimulations along the rostro-caudal direction by black squares on a gray background drifting along the rostro-caudal direction. The screen was oriented to be centered and parallel to the right retina on the same side where the whisker stimulations were delivered. The locations of the moving square along the vertical axis were chosen to roughly match the locations of the whiskers on the visual field (see Methods).

To evaluate whether mice use spatial information from the tactile task to infer reward contingencies in the visual task, we examined two specific scenarios. In a cohort of mice, the visual stimulus that is spatially congruent with the whisker stimulus remained associated to a reward resulting in congruent reward contingencies. For other mice, we changed the rule after the switch and rewarded responses for the bottom stimulus resulting in incongruent reward contingences between the two modalities (Fig. 1B). After modality switch with congruent reward contingencies, we observed that mice immediately continued to perform the task with licking responses

and performance comparable to the previous session where whiskers were stimulated in the dark (Fig. 1C-D). When reward contingencies were spatially incongruent, mice initially attempted to lick for both stimuli. In some cases, this led to performance dropping below chance levels. Eventually, mice began to disengage from the task, ceasing any licking behavior in response to sensory stimuli (Fig. 1E-F). Mice strongly resisted engaging with this task, even if we manually delivered water drops, which typically induced prolonged licking bouts due to their heightened thirst-driven motivation. This suggests that mice manifest a conflicting prior regarding task rules rather than an insufficient motivational drive to perform the task. This was confirmed at the population level where we observed a characteristic behavior across all mice showing either a seamless transfer learning in the congruent case (Fig. 1G) or resistance to perform the task over several days in the incongruent case (Fig. 1H). We confirmed that this result was not caused by a biased preference for the top whisker or top visual stimulus by performing the same experiments with cohorts of mice trained to respond to the bottom whisker. Similar behavioral responses were observed, with mice performing normally after the switch with congruent reward contingencies but not when contingencies were spatially incongruent (Fig. 1I-J).

We further assessed if mice could equally transfer learning when they were trained to the visual task first. We found that mice maintained task performance in congruent scenarios, but failed in incongruent ones, irrespective of the rewarded stimulus during the first task (Supplementary Fig. S1). However, we observed a small but significant drop of performance from visual to tactile tasks (Supplementary Fig. S1F) that was absent in the tactile to visual switch (Fig. 1I). This asymmetry might arise from generally higher initial performances in the visual task, the distinct nature

of sensory representation (discrete for whiskers and continuous for visual inputs), or variations in circuit properties.

Despite the limited visibility of the capillary glass tubes used for whisker stimulation, mice could potentially rely on visual cues to perform the tactile task, thereby generalizing within the visual domain instead of across sensory modalities. We carried out control experiments where mice proficient in whisker discrimination underwent sessions with whiskers temporarily removed from the tubes, and subsequently reintroduced. We found that the mice's performance in both detection and discrimination tasks dropped to chance levels immediately after the whiskers were removed from the capillary tubes. However, their performance recovered as soon as the whiskers were reintroduced in the stimulated tubes (Supplementary Fig. S2). This demonstrates that mice are not using visual cues to perform the whisker discrimination task in dark conditions.

Besides the transfer of knowledge regarding the congruent reward contingencies between modalities, mice could also potentially use acquired knowledge about Go/No go task structure to continue performing after the switch. To evaluate learning trajectories post-switch with prior Go/No go task experience, similar motivation levels but no spatial information, we trained mice on an auditory Go/No go task using pure tones (Fig. 1K). These tones bear no clear spatial relationship to the visual stimuli introduced after the modality switch. In this condition, performance dropped to chance level after the switch but steadily recovered to expert level over the next few days (Fig. 1L-M). This indicates that mice can learn the new task significantly faster when they have no prior on spatial reward contingencies compared to when they do (Fig. 1N). In the latter scenario, mice, when trained for longer durations, would

158 eventually perform again but often with lower performance (Supplementary Fig.
159 S1E,G).

160 Finally, we compared our results with a situation where reward contingencies
161 are switched within the tactile modality, as previously done in other studies (*26, 27*).
162 After mice reached expert level, we reversed the reward contingencies between the
163 two whiskers. We observed a stark drop in discrimination performance, significantly
164 below chance levels, indicating that mice persist to perform the task following the
165 original rule (Supplementary Fig. S3). Following the switch, task performance
166 increased but remained below or at chance level for at least three consecutive days.
167 This demonstrates that mice behave differently when reward contingencies switch
168 occurs within the same sensory modality as opposed to across different modalities. In
169 the former case, they continue to inflexibly produce the same sensorimotor
170 transformation, likely reflecting ingrained habitual behaviors.

171 Together these results show that mice can swiftly and spontaneously transfer
172 previously learned associations across sensory modalities leveraging on the spatial
173 organization common to these senses. Conversely, mice display a pronounced
174 resistance to re-learning when required to perform against the previously valid spatial
175 rule. Cortical circuits are necessary for conscious perception and are believed to
176 mediate goal-directed cross-modal transfer learning (*2, 7*). To pinpoint the cortical
177 regions responsible for transferring spatial information between the visual and whisker
178 somatosensory systems, we mapped the topographic representation of vertical space
179 for both modalities in the dorsal cortex of transgenic mice expressing the calcium
180 indicator GCaMP6f in cortical layer 2/3 (see Methods). We first used standard
181 retinotopic and somatotopic mapping protocols (*28, 29*) to identify whisker responding
182 and retinotopically organized cortical areas through a 5 mm diameter cranial window

over the posterior part of the dorsal cortex (Fig. 2A-B). Whisker response pattern and retinotopic sign maps could be used to precisely fit a projection of the Allen Mouse Brain Atlas (see Methods). We used this atlas to collectively register all the brains, enabling us to generate average whisker response maps (Fig. 2C) and retinotopic maps (Fig. 2D and Supplementary Fig. S4A-C), which corroborated the previously reported functional organization of higher visual areas (*28*). We observed that whisker stimulations evoked activity in several visually responsive associative areas including the anterior (A), rostro-lateral (RL), antero-lateral (AL) and latero-intermediate (LI) areas as well as other somatosensory areas indicating that whisker representations might be present in a more extended cortical network than previously reported (*30*).

To explore representations of the vertical spatial dimension during unisensory and multisensory stimulations, we designed a visuo-tactile sparse noise protocol (Fig. 2E, see Methods). This protocol was first used to obtain retinotopic and somatotopic maps for elevation (vertical space) by computing preferred spatial position for each pixel. When we applied this approach to whisker stimuli, we identified the established somatotopic arrangement of the primary and secondary whisker somatosensory cortices, S1 and S2, which exhibit a topographic inversion at their boundary (Fig. 2F). Additionally, we observed organized somatotopic maps in the same associative visual areas (A, RL, AL and LI) as observed in Figure 2C. Strikingly, the map obtained with visual stimuli displayed a very similar organization in these associative areas as well as in the primary and secondary somatosensory cortices (Fig. 2G). The extended spatial representations evoked by visual or tactile stimuli were found consistently across mice with good cranial windows (Supplementary Fig. S4D). This suggests that spatially localized stimuli, whether evoked by visual or whisker tactile stimuli, might share a common representation, facilitating the mapping between sensory modalities,

as previously observed in the superior colliculus (*20, 21*). In particular, the spatial representations evoked by these two modalities displayed an angular offset that we estimated at 30 degrees by comparing the angle difference between elevation gradient vectors obtained from the maps (see Methods). This might represent the mouse's internal model of how whisker sensations align with their visual field.

We further investigated the properties of these representations by first computing a modality preference index to assess what sensory modality dominates each area (Fig. 2H). As expected, S1 and S2 were dominated by tactile inputs whereas primary visual cortex V1 was dominated by visual inputs. Interestingly, A, RL and the region at the border between AL and LI displayed a more balanced preference for both modalities though the lateral side was biased toward visual inputs and the medial side toward tactile inputs. In addition, we measured the spatial coherence between retinotopic and somatotopic elevation maps indicating how they locally correlate (Fig. 2I, see Methods). This confirmed a widespread co-alignment across most associative areas in the belt separating V1 and S1. To compare multisensory responses triggered by visuo-tactile stimuli with unisensory responses, we computed a multisensory modulation index (see Methods) which compares multisensory responses against the maximal unisensory response on a pixel-by-pixel basis. These maps were obtained under visuo-tactile conditions, delaying the whisker stimuli by 0.15 seconds to ensure that the evoked responses in the cortex synchronized (Supplementary Fig. S4E-H), as also documented in previous studies (*22, 23*). The resulting map revealed a strong multisensory enhancement in visuo-tactile associative areas and in S2 (Fig. 2J) confirming that these areas display neural computation classically attributed to multisensory brain regions. More generally, we found that multisensory enhancement

232 was more pronounced in regions with higher coherence between spatial maps and
233 with strong bimodal representation (Fig. 2K-L).

234 Functional maps measured with wide-field calcium imaging could result from
235 direct inputs from visual and tactile cortical areas, could by evoked by top-down inputs
236 (*31*) or even be the result of highly stereotypical uninstructed movements evoked by
237 sensory stimuli (*32*). To investigate the synaptic origin of these responses, we
238 performed additional experiments to anatomically map axonal projections from
239 primary sensory areas to associative areas displaying visuo-tactile representations.
240 We obtained visual and tactile functional maps for vertical spatial representation in
241 wild-type mice using intrinsic optical signal imaging under low isoflurane anesthesia
242 (see Methods). These maps were then used to identify two cortical locations
243 representing distinct iso-horizontal vertical positions in V1 or to target B2 and C2
244 barrels in S1. We re-opened the cranial window and injected two adeno-associated
245 viral vectors to induce expression of GFP and tdTomato in the respective locations
246 (Supplementary Fig. S5). After 10-15 days, transcardial perfusion was performed and
247 brains were extracted, flattened and sliced (see Methods). We used
248 immunohistochemical localization of M2 muscarinic acetylcholine receptors
249 (M2AChR) to identify the barrels in S1 and V1, given the enriched presence of these
250 receptors in these regions. We could then use these landmarks to fit the Allen Mouse
251 Brain Atlas on the reconstructed stack confirming the locations of injection sites along
252 the vertical representation of V1 and S1 (Fig. 2M-N). Axonal projections from the
253 primary sensory areas were found in associative cortical regions where visuo-tactile
254 responses were measured with the same spatial organization. This confirms that the
255 functional maps depicted in Figure 2F-G are, in part, inherited from direct feed-forward
256 projections from primary cortical areas.

Since cross-modal transfer learning could stem from sensory information transitioning between systems, we hypothesized that spatially organized feedback projections would be essential to generate responses like those observed in Figure 2F-G. In particular, no direct projections were found from V1 to S1 while visual stimuli could evoke responses in S1. Previous work has shown that feedback from high visual areas (including A, RL, AL and LI) to V1 are spatially organized along the vertical dimension (*33*) but it is not clear if this holds true for feedback projections to S1. Feedback projections from associative areas to S1 were assessed with the same strategy for anatomical mapping, using injections of Cholera Toxin Subunit B (CTB) conjugated either with Alexa555 or Alexa 647 (Fig. 2O). We observed retinotopically organized feedback projections from associative areas to S1 (see Methods).

These findings suggest that sensory representations elicited by either visual or whisker tactile stimuli have a consistent organization throughout much of the dorsal cortex. These representations are directly inherited from axonal projections originating from primary sensory areas. Furthermore, spatially organized feedforward and feedback projections allow spatial information to be transferred from one sensory modality to another. However, functional maps obtained with wide-field imaging do not reveal precise computation performed at single-cell level and could still be prone to artefacts produced by neuronal processes originating from other brain structures. We performed two-photon calcium imaging in a subset of mice implanted with a cranial window. Single neurons GCaMP6f signal was extracted during the visuo-tactile sparse noise protocol in fields of view covering different cortical areas identified with the atlas (Supplementary Fig. S6A-F). Properties for single neurons could then be realigned on the reference atlas across mice to compare local cellular responses extracted at the soma with the corresponding wide-field region. Many recordings were performed

across a large portion of the dorsal cortex to fully cover the responsive visuo-tactile areas (Supplementary Fig. S6G-H). In doing this, we reconstructed somatotopic maps for the vertical axis across the cranial window that aligned closely with the wide-field maps (Fig. 3A). Neurons with whisker tactile responses were found across the belt of associative areas following the somatotopic organization. This further confirms a more extended network of whisker responsive and visuo-tactile cortical regions (*30*). The same was true for neurons responding to visual stimuli which were found across most visual and tactile areas even extending to S1 and S2 (Fig. 3B). Importantly, we localized neurons that were bimodal in that they responded to unisensory visual or whisker stimuli revealing two distinct clusters of neurons (Fig. 3C) corresponding to the domains identified with wide-field imaging (Fig. 2H). As the centers of these clusters were located in areas associated with the dorsal (A, RL) and ventral (AL, LI, LM) streams, we will use these designations moving forward.

We then characterized the response properties of single neurons for visuo-tactile stimuli in comparison to their responses to unisensory stimulations (Fig. 3D) using delayed tactile stimuli as previously described (Supplementary Fig. S4E-H). We found that multimodal neurons are generally tuned to both a vertical visual position and to one whisker. For example, the neuron in Figure 3D responded to the bottom part of the visual field and to the bottom whisker therefore showing selectivity for spatially congruent visual and tactile inputs. Using these unisensory responses, we predicted the response pattern to visuo-tactile stimuli as the maximum response between the two modalities for each combination. When comparing the predicted with the measured responses, we observed suppression in incongruent combinations whereas congruent combinations were either unaffected or slightly enhanced. Hence, neurons conserved a specific selectivity for bottom stimuli regardless of the modalities

stimulated (unisensory or multisensory) reminiscent of supramodal encoding of object orientation in visuo-tactile neurons of the rat posterior parietal cortex (*18*). Responses selective for spatially congruent whisker and visual stimuli in single neurons were confirmed at the population level across all multimodal neurons (Fig. 3E). Additionally, we observed that neurons maintained a larger tuning for whisker and visual position in measured responses compared to predicted ones (Fig. 3F). Neurons in the ventral areas were more tuned to specific visual locations while neurons in the dorsal areas were more tuned to specific whiskers (Fig. 3G-H), in line with the modality preference observed with wide-field imaging (Fig. 2H). These results showed that neurons in visuo-tactile associative areas are indeed specifically responding to spatially congruent visuo-tactile stimuli and maintain a spatial preference independent from the sensory input despite specialized modality preferences in the ventral and dorsal areas.

Functional properties observed both at large scale and at single-cell level as well as the anatomical connectivity between visual and whisker somatosensory cortices indicate that they could mediate cross-modal transfer learning during goal-directed behaviors. To test this hypothesis, we performed loss-of-function manipulations. As these representations are present in naive mice (not yet exposed to any behavioral task), the propagation of evoked responses from the primary sensory cortex to other cortices during training could begin to recruit neurons supporting task execution early in the training process. Therefore, we reasoned that visuo-tactile associative cortical areas should be silenced prior to any learning as cross-modal transfer could happen already during learning. To do so, we expressed the tetanus toxin light chain (TeNT-P2A-GFP) through viral vectors delivery to prevent the release of synaptic neurotransmitters in transfected neurons (see Methods). Neurons expressing TeNT also co-expressed GFP, allowing comparison of the expression

pattern with the previously obtained atlas (Fig. 4A). After removing blood vessels patterns and comparing with expression before injection (see Methods), we characterized the extent of GFP expression and overlap with different visuo-tactile areas (Fig. 4B). To ensure that vesicle release in the transfected neurons was significantly suppressed before beginning the behavioral training, we waited at least four weeks after the viral injection (*34*). Mice were able to normally learn the whisker discrimination task as shown by the expert performance of an example mouse in Figure 4C. However, upon switching to the visual task with congruent reward contingencies, performance dropped to chance level. The mouse responded to visual stimuli at the same frequency as its spontaneous reactions, suggesting an inability to detect or discriminate (Fig. 4D).

To assess which visuo-tactile area was necessary for the cross-modal transfer learning, we expressed TeNT-P2A-GFP in each mouse using a different location (see Methods). Across all mice, we managed to cover all visuo-tactile areas with different degrees of overlap for each mouse (Supplementary Fig. S7). For every mouse, we then performed the modality switch and measured changes in performance following the switch. Collecting data from all the mice, we could then compute, for each cortical area, how much overlap of GFP expression with this area correlates with the performance drop (Supplementary Fig. S7F). This analysis revealed that only RL silencing results in strong significant correlation with behavior impairment (Fig. 4E). To confirm this result, we also performed a reverse-correlation analysis mapping the average GFP coverage that evoked a complete impairment of transfer learning (average performance below 55% post-switch) also highlighting RL as being necessary for cross-modal transfer learning (Fig. 4F). The boundary of S1 was also

discernible, potentially due to a consistent spread of fluorescence resulting from the dense axonal projections from RL to S1 (Fig. 2O).

When targeting injections specifically to areas within the dorsal or ventral streams, we found that silencing the dorsal stream notably hindered transfer learning, leading to slower re-learning post-congruent switch (Fig. 4G-H). In contrast, silencing the ventral stream did not prevent mice from performing above chance level, despite an initial performance drop that was rapidly recovered (Fig. 4I-J). Mice with silenced RL could learn the whisker task normally but were not able to transfer learned associations to the visual task using spatial correspondence. However, they were able to learn the visual task in the following days indicating that their ability to learn both tasks was not impaired, only the transfer was. We therefore observed a dramatic difference in performance change after the switch with control mice described in Figure 1I compared to all mice expressing TeNT for the same protocol (Fig. 4K). To characterize the learning trajectory of mice expressing TeNT in visuo-tactile areas, we isolated all mice with complete transfer impairment and measured the learning rate following the switch. Interestingly, mice with impaired transfer learning could still learn the visual task with learning rate comparable to the one observed for the auditory to visual transfer (Fig. 4L) indicating that TeNT mice learn the visual task with the same rate as a task without spatial information.

Discussion

In this study, we investigated the behavioral and neurophysiological underpinning of cross-modal transfer learning in mice, focusing on the representation of space across modalities. We first described a novel behavioral phenomenon where

mice display an ability for cross-modal transfer learning across whisker and visual modalities using spatial information. We then examined spatial representations in the dorsal cortex of mice during unisensory and multisensory stimulations. Using a visuo-tactile sparse noise protocol, we obtained retinotopic and somatotopic maps for elevation, discovering organized somatotopic maps in various associative areas traditionally known as visual areas only. Notably, similar spatial organizations along the vertical axis were observed for both visual and whisker tactile stimuli, suggesting shared representations across modalities. We mapped anatomical axonal projections between primary sensory areas and associative areas, unveiling a network of both feedforward and feedback projections that allows permanent spatial information transfer between sensory modalities. Single-cell recordings corroborated these findings, with neurons showing a preference for spatially congruent visuo-tactile stimuli, together with a spatial profile of multisensory modulation enabling supramodal encoding of space. Lastly, we demonstrated that silencing the visuo-tactile cortical area RL prior to training disrupted mice's ability to generalize from a whisker-based task to a congruent visual-based task, highlighting the central role of this area in cross-modal transfer learning. These results thus reveal the existence of interconnected spatial representations in the dorsal cortex and their role in multisensory processing and learning.

Contrary to the longstanding belief that cross-modal transfer learning is exclusive to apes and humans (*35*), studies have unveiled this cognitive ability in various species, including rodents (*2*) and, more recently, bumblebees (*36*). Early research into the brain circuits involved in cross-modal processing underscores the necessity of different brain structures to transfer diverse types of information such as spatial, temporal, and object-related features (*4, 7, 8*). These findings advocate for the

existence of dedicated cortical circuits developed for specific multimodal representations and functions. The posterior parietal cortex (PPC) plays an important role in processing spatial information (*24, 25*). Our results point out that area RL, a subregion of PPC, is crucial for cross-modal transfer learning based on spatial properties. Nonetheless, other forms of cross-modal transfer learning, like object recognition or duration discrimination, may engage different cortical regions.

Wide-field imaging and anatomical mapping experiments revealed a more extensive network of visuo-tactile associative cortical areas than previously identified (*22, 23, 37*). Many associative areas exhibited a shared representation of vertical space between visual and whisker tactile inputs. Previous research reported aligned representations of whisker and visual sensory inputs in the superior colliculus, where each whisker's representation was proportional to the space it occupies on the visual field, favoring whiskers of the top rows (*20, 21*). Past work had also hinted at the existence of similar co-aligned maps in cortical area RL, albeit with different coordinates and lacking spatial resolution (*22*). The spatial maps we observed extended even to primary sensory cortices through associative areas, with strong responses observed in S1 following pure visual stimuli, despite no direct projections between V1 and S1 (*38*). This second-order response map could potentially correspond to area lateral to RL (RLL), previously reported as overlapping with S1 (*28*). Although recent findings advise caution in interpreting cross-modal signals between primary cortical areas due to potential confusion with signals evoked by uninstructed movements (*32*), our results suggest that certain fundamental features, such as spatial location, indeed have a shared multimodal representation across various cortical areas, possibly oriented toward object or event-oriented encoding (*39*). More complex features like orientation have been identified as being represented in

parts of the posterior parietal cortex (PPC) with supramodal encoding for visuo-tactile stimuli (*40*). Future work could explore whether these representations in PPC are essential for the cross-modal transfer learning of these features.

Indeed, the dorsal stream areas are traditionally associated with spatial, attentional, or movement processing, while ventral stream areas are linked to object recognition. Previous studies have reported visuo-tactile multimodal responses in both ventral and dorsal pathways in primates (*41–43*), including areas initially thought to be exclusive to visual processing like V4 (*44*) and MT/V5 (*45*). Our study uncovers an extended network of visuo-tactile areas, some belonging to the ventral or dorsal streams in mice (*37, 46*). Specifically, RL and A are typically allocated to the dorsal stream, while LI, with its significant projection to the postrhinal (POR) cortex, is associated with the ventral stream (*37*). Lesions in the perirhinal cortex in rats, downstream to POR, yield a severe impairment in learning to discriminate complex tactile features (*47*) and obstruct cross-modal object recognition during spontaneous exploration (*8*). This finding implies that the ventral pathway might contribute more to cross-modal object recognition rather than the cross-modal spatial mapping observed in RL (*48, 49*). Future research focusing on object discrimination could provide further insight into this hypothesis.

Area RL is strategically positioned between the primary visual cortex and the primary whisker somatosensory cortex, facilitating bidirectional information transfer between these sensory systems. It possesses a retinotopic map biased towards the lower part of the visual field (*28*) where most whiskers are visible, and its neurons are tuned to high binocular disparity, aligning with objects in close proximity, potentially within whiskers' reach (*50*). Our anatomical findings highlight that RL possesses robust feedback projections to S1. Feedback projections from higher-order visual

areas to the primary visual cortex in mice have been shown to influence V1 responses (*51*) and to shape the properties of the non-classical receptive field with contextual information (*52*). These feedback projections are retinotopically organized (*33*) and modulate responses through dendritic integration in a location-specific way (*53*). Consequently, feedback projections from RL could potentially perform two distinct functions: contextual integration within the visual modality and cross-modal information transfer.

At the single-neuron level, we observed multisensory modulations that favor spatially congruent visuo-tactile stimuli and suppresses responses to incongruent stimuli. This response pattern could potentially arise from surround suppression for incongruent inputs, mediated by local parvalbumin-positive interneurons as observed for conflicting visuo-auditory stimuli (*13*). Our protocol enables the comparison of congruent versus incongruent combinations, uncovering multisensory modulation rules that shape a supramodal representation of spatial inputs. Similar coding schemes have been reported for the orientation of visual or tactile gratings (*40*). This raises the question of spatio-temporal properties of visuo-tactile stimuli that are perceived as congruent. The evoked responses in the somatosensory cortex precede visual responses by around 150 ms, and multisensory modulation is enhanced when the onset of evoked activity aligns (*22, 23*). Given that objects are typically seen before being touched, we hypothesize that this could signify a delay in sensory information under natural conditions. A systematic characterization of the spatio-temporal properties of visuo-tactile stimuli perceived as congruent could help elucidate the cortical circuits that bind multisensory properties in experienced objects and events.

References and Notes:

1. U. Noppeney, Perceptual Inference, Learning, and Attention in a Multisensory World. *Annu Rev Neurosci*. **44**, 449–473 (2021).

2. J. M. Cloke, D. L. Jacklin, B. D. Winters, The neural bases of crossmodal object recognition in non-human primates and rodents: A review. *Behavioural Brain Research*. **285**, 118–130 (2015).

3. W. H. Meck, R. M. Church, Discrimination of intertrial intervals in cross-modal transfer of duration. *Bull Psychon Soc*. **19**, 234–236 (1982).

4. E. H. Yeterian, The effects of visual or auditory cortical lesions on specific cross-modal transfer in the rat. *Neuropsychologia*. **15**, 517–527 (1977).

5. R. Over, N. J. Mackintosh, Cross-modal Transfer of Intensity Discrimination by Rats. *Nature*. **224**, 918–919 (1969).

6. E. R. Delay, Effects of cross-modal transfer on direct and reversal learning in the rat. *Anim Learn Behav*. **14**, 427–434 (1986).

7. R. Tees, The effects of posterior parietal and posterior temporal cortical lesions on multimodal spatial and nonspatial competencies in rats. *Behavioural Brain Research*. **106**, 55–73 (1999).

8. B. D. Winters, J. M. Reid, A Distributed Cortical Representation Underlies Crossmodal Object Recognition in Rats. *The Journal of Neuroscience*. **30**, 6253–6261 (2010).

9. M. A. Meredith, B. E. Stein, Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol*. **56**, 640–662 (1986).

10. M. T. Wallace, M. A. Meredith, B. E. Stein, Multisensory Integration in the Superior Colliculus of the Alert Cat. *J Neurophysiol*. **80**, 1006–1010 (1998).

11. E. Macaluso, Multisensory Processing in Sensory-Specific Cortical Areas. *The Neuroscientist*. **12**, 327–338 (2006).

12. G. A. Calvert, T. Thesen, Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris*. **98**, 191–205 (2004).

13. Y. H. Song, J. H. Kim, H. W. Jeong, I. Choi, D. Jeong, K. Kim, S. H. Lee, A Neural Circuit for Auditory Dominance over Visual Perception. *Neuron*. **93**, 940-954.e6 (2017).

14. A. R. Garner, G. B. Keller, A cortical circuit for audio-visual predictions. *Nat Neurosci*. **25**, 98–105 (2022).

15. L. Godenzini, D. Alwis, R. Guzulaitis, S. Honnuraiah, G. J. Stuart, L. M. Palmer, Auditory input enhances somatosensory encoding and tactile goal-directed behavior. *Nat Commun*. **12** (2021), doi:10.1038/s41467-021-24754-w.

16. E. W. Jenkinson, M. Glickstein, Whiskers, barrels, and cortical efferent pathways in gap crossing by rats. *J Neurophysiol*. **84**, 1781–1789 (2000).

17. N. J. Sofroniew, J. D. Cohen, A. K. Lee, K. Svoboda, Natural whisker-guided behavior by head-fixed mice in tactile virtual reality. *Journal of Neuroscience*. **34**, 9537–9550 (2014).

18. N. Nikbakht, A. Tafreshiha, D. Zoccolan, M. E. Diamond, Supralinear and Supramodal Integration of Visual and Tactile Signals in Rats: Psychophysics and Neuronal Mechanisms. *Neuron*. **97**, 626-639.e8 (2018).

19. S. Gharaei, E. Arabzadeh, S. G. Solomon, Integration of visual and whisker signals in rat superior colliculus. *Sci Rep*. **8** (2018), doi:10.1038/s41598-018-34661-8.

20. U. C. Drager, D. H. Hubel, Physiology of visual cells in mouse superior colliculus and correlation with somatosensory and auditory input. *Nature*. **253**, 203–204 (1975).

21. U. C. Drager, D. H. Hubel, Topography of visual and somatosensory projections to mouse superior colliculus. *J Neurophysiol*. **39**, 91–101 (1976).

22. U. Olcese, G. Iurilli, P. Medini, Cellular and synaptic architecture of multisensory integration in the mouse neocortex. *Neuron*. **79**, 579–593 (2013).

23. J. Caron-Guyon, J. Corbo, Y. Zennou-Azogui, C. Xerri, A. Kavounoudias, N. Catz, Neuronal encoding of multisensory motion features in the rat associative parietal cortex. *Cerebral Cortex*. **30**, 5372–5386 (2020).

24. J. R. Whitlock, Posterior parietal cortex. *Current Biology*. **27**, R691–R695 (2017).

25. S. Pasalar, T. Ro, M. S. Beauchamp, TMS of posterior parietal cortex disrupts visual tactile multisensory integration. *European Journal of Neuroscience*. **31**, 1783–1790 (2010).

26. R. Chéreau, T. Bawa, L. Fodoulian, A. Carleton, S. Pagès, A. Holtmaat, Dynamic perceptual feature selectivity in primary somatosensory cortex upon reversal learning. *Nat Commun*. **11**, 3245 (2020).

543 27. A. Banerjee, G. Parente, J. Teutsch, C. Lewis, F. F. Voigt, F. Helmchen, Value-guided
544 remapping of sensory cortex by lateral orbitofrontal cortex. *Nature*. **585**, 245–
545 250 (2020).

546 28. J. Zhuang, L. Ng, D. Williams, M. Valley, Y. Li, M. Garrett, J. Waters, An extended
547 retinotopic map of mouse cortex. *Elife*. **6**, 1–29 (2017).

548 29. G. Matteucci, M. Guyoton, J. M. Mayrhofer, M. Auffret, G. Foustoukos, C. C. H.
549 Petersen, S. El-Boustani, Cortical sensory processing across motivational states
550 during goal-directed behavior. *Neuron*. **110**, 4176-4193.e10 (2022).

551 30. T. Yamashita, A. Vavladeli, A. Pala, K. Galan, S. Crochet, S. S. A. Petersen, C. C. H.
552 Petersen, Diverse Long-Range Axonal Projections of Excitatory Layer 2/3
553 Neurons in Mouse Barrel Cortex. *Front Neuroanat*. **12** (2018),
554 doi:10.3389/fnana.2018.00033.

555 31. I. Choi, J.-Y. Lee, S.-H. Lee, Bottom-up and top-down modulation of multisensory
556 integration. *Curr Opin Neurobiol*. **52**, 115–122 (2018).

557 32. C. Bimbard, T. P. H. Sit, A. Lebedeva, C. B. Reddy, K. D. Harris, M. Carandini,
558 Behavioral origin of sound-evoked activity in mouse visual cortex. *Nat Neurosci*.
559 **26**, 251–258 (2023).

560 33. M. M. Morimoto, E. Uchishiba, A. B. Saleem, Organization of feedback projections
561 to mouse primary visual cortex. *iScience*. **24** (2021),
562 doi:10.1016/j.isci.2021.102450.

563 34. B. Zingg, B. Peng, J. Huang, H. W. Tao, L. I. Zhang, Synaptic Specificity and
564 Application of Anterograde Transsynaptic AAV for Probing Neural Circuitry. *The*
565 *Journal of Neuroscience*. **40**, 3250–3267 (2020).

566 35. R. K. Davenport, C. M. Rogers, Intermodal Equivalence of Stimuli in Apes. *Science*
567 *(1979)*. **168**, 279–280 (1970).

568 36. C. Solvi, S. Gutierrez Al-Khudhairy, L. Chittka, Bumble bees display cross-modal
569 object recognition between visual and tactile senses. *Science (1979)*. **367**, 910–
570 912 (2020).

571 37. Q. Wang, O. Sporns, A. Burkhalter, Network analysis of corticocortical connections
572 reveals ventral and dorsal processing streams in mouse visual cortex. *Journal of*
573 *Neuroscience*. **32**, 4386–4399 (2012).

574 38. Q. Wang, A. Burkhalter, Area map of mouse visual cortex. *Journal of Comparative*
575 *Neurology*. **502**, 339–357 (2007).

576 39. A. A. Ghazanfar, C. E. Schroeder, Is neocortex essentially multisensory? *Trends*
577 *Cogn Sci*. **10** (2006), pp. 278–285.

578 40. N. Nikbakht, A. Tafreshiha, D. Zoccolan, M. E. Diamond, Supralinear and
579 Supramodal Integration of Visual and Tactile Signals in Rats: Psychophysics and
580 Neuronal Mechanisms. *Neuron*. **97**, 626-639.e8 (2018).

581 41. A. Amedi, R. Malach, T. Hendler, S. Peled, E. Zohary, Visuo-haptic object-related
582 activation in the ventral visual pathway. *Nat Neurosci*. **4**, 324–330 (2001).

583 42. T. W. James, S. Kim, "Dorsal and Ventral Cortical Pathways for Visuo-haptic Shape
584 Integration Revealed Using fMRI" in *Multisensory Object Perception in the Primate*
585 *Brain* (Springer New York, New York, NY, 2010), pp. 231–250.

586 43. S. Lacey, K. Sathian, Visuo-haptic multisensory object recognition, categorization,
587 and representation. *Front Psychol*. **5** (2014), doi:10.3389/fpsyg.2014.00730.

588 44. J. H. R. Maunsell, G. Sclar, T. A. Nealey, D. D. DePriest, Extraretinal representations
589 in area V4 in the macaque monkey. *Vis Neurosci*. **7**, 561–573 (1991).

590 45. M. C. Hagen, O. Franzén, F. McGlone, G. Essick, C. Dancer, J. V. Pardo, Tactile
591 motion activates the human middle temporal/V5 (MT/V5) complex. *European*
592 *Journal of Neuroscience*. **16**, 957–964 (2002).

593 46. J. H. Marshel, M. E. Garrett, I. Nauhaus, E. M. Callaway, Functional specialization of
594 seven mouse visual cortical areas. *Neuron*. **72**, 1040–1054 (2011).

595 47. J. M. J. Ramos, Essential Role of the Perirhinal Cortex in Complex Tactual
596 Discrimination Tasks in Rats. *Cerebral Cortex*. **24**, 2068–2080 (2014).

597 48. G. Matteucci, R. Bellacosa Marotti, M. Riggi, F. B. Rosselli, D. Zoccolan, Nonlinear
598 processing of shape information in rat lateral extrastriate cortex. *The Journal of*
599 *Neuroscience*, 1938–18 (2019).

600 49. S. Tafazoli, H. Safaai, G. De Franceschi, F. B. Rosselli, W. Vanzella, M. Riggi, F.
601 Buffolo, S. Panzeri, D. Zoccolan, Emergence of transformation-tolerant
602 representations of visual objects in rat lateral extrastriate cortex. *Elife*. **6** (2017),
603 doi:10.7554/eLife.22794.

604 50. A. La Chioma, T. Bonhoeffer, M. Hübener, Area-Specific Mapping of Binocular
605 Disparity across Mouse Visual Cortex. *Current Biology*. **29**, 2954-2960.e5 (2019).

606 51. M. N. Oude Lohuis, A. C. Canton, C. M. A. Pennartz, U. Olcese, Higher order visual
607 areas enhance stimulus responsiveness in mouse primary visual cortex. *Cerebral*
608 *Cortex*. **32**, 3269–3288 (2022).

52. A. J. Keller, M. M. Roth, M. Scanziani, Feedback generates a second receptive field in neurons of the visual cortex. *Nature*. **582**, 545–549 (2020).

53. M. Fişek, D. Herrmann, A. Egea-Weiss, M. Cloves, L. Bauer, T. Y. Lee, L. E. Russell, M. Häusser, Cortico-cortical feedback engages active dendrites in visual cortex. *Nature*. **617**, 769–776 (2023).

54. K. K. Sit, M. J. Goard, Distributed and retinotopically asymmetric processing of coherent motion in mouse visual cortex. *Nat Commun*. **11**, 3565 (2020).

55. E. S. Lein, M. J. Hawrylycz, N. Ao, M. Ayres, A. Bensinger, A. Bernard, A. F. Boe, M. S. Boguski, K. S. Brockway, E. J. Byrnes, L. Chen, L. Chen, T.-M. Chen, M. Chi Chin, J. Chong, B. E. Crook, A. Czaplinska, C. N. Dang, S. Datta, N. R. Dee, A. L. Desaki, T. Desta, E. Diep, T. A. Dolbeare, M. J. Donelan, H.-W. Dong, J. G. Dougherty, B. J. Duncan, A. J. Ebbert, G. Eichele, L. K. Estin, C. Faber, B. A. Facer, R. Fields, S. R. Fischer, T. P. Fliss, C. Frensley, S. N. Gates, K. J. Glattfelder, K. R. Halverson, M. R. Hart, J. G. Hohmann, M. P. Howell, D. P. Jeung, R. A. Johnson, P. T. Karr, R. Kawal, J. M. Kidney, R. H. Knapik, C. L. Kuan, J. H. Lake, A. R. Laramee, K. D. Larsen, C. Lau, T. A. Lemon, A. J. Liang, Y. Liu, L. T. Luong, J. Michaels, J. J. Morgan, R. J. Morgan, M. T. Mortrud, N. F. Mosqueda, L. L. Ng, R. Ng, G. J. Orta, C. C. Overly, T. H. Pak, S. E. Parry, S. D. Pathak, O. C. Pearson, R. B. Puchalski, Z. L. Riley, H. R. Rockett, S. A. Rowland, J. J. Royall, M. J. Ruiz, N. R. Sarno, K. Schaffnit, N. V. Shapovalova, T. Sivisay, C. R. Slaughterbeck, S. C. Smith, K. A. Smith, B. I. Smith, A. J. Sodt, N. N. Stewart, K.-R. Stumpf, S. M. Sunkin, M. Sutram, A. Tam, C. D. Teemer, C. Thaller, C. L. Thompson, L. R. Varnam, A. Visel, R. M. Whitlock, P. E. Wohnoutka, C. K. Wolkey, V. Y. Wong, M. Wood, M. B. Yaylaoglu, R. C. Young, B. L. Youngstrom, X. Feng Yuan,

632 B. Zhang, T. A. Zwingman, A. R. Jones, Genome-wide atlas of gene expression in
633 the adult mouse brain. *Nature*. **445**, 168–176 (2007).

634 56. M. Pachitariu, C. Stringer, S. Schröder, M. Dipoppa, L. F. Rossi, M. Carandini, K. D.
635 Harris, Suite2p: beyond 10,000 neurons with standard two-photon microscopy.
636 *bioRxiv*, 061507 (2016).

637

638

Acknowledgements:

We thank members of the El-Boustani laboratory for discussions.

Funding:

This work was supported by the Swiss National Science Foundation: PCEFP3_181070 (S.E.-B.).

Author Contribution:

M.G., G.M. and S.E.-B. conceived experiments. M.G. performed surgeries, viral infections, wide-field and two-photon imaging, behavioral experiments, and post-mortem analysis. G.M. and C.G.F. performed behavioral experiments. G.M. coded acquisition software and performed data analysis. M.G., G.M. and S.E.-B. wrote the manuscript.

Competing interest:

Authors declare no competing interests.

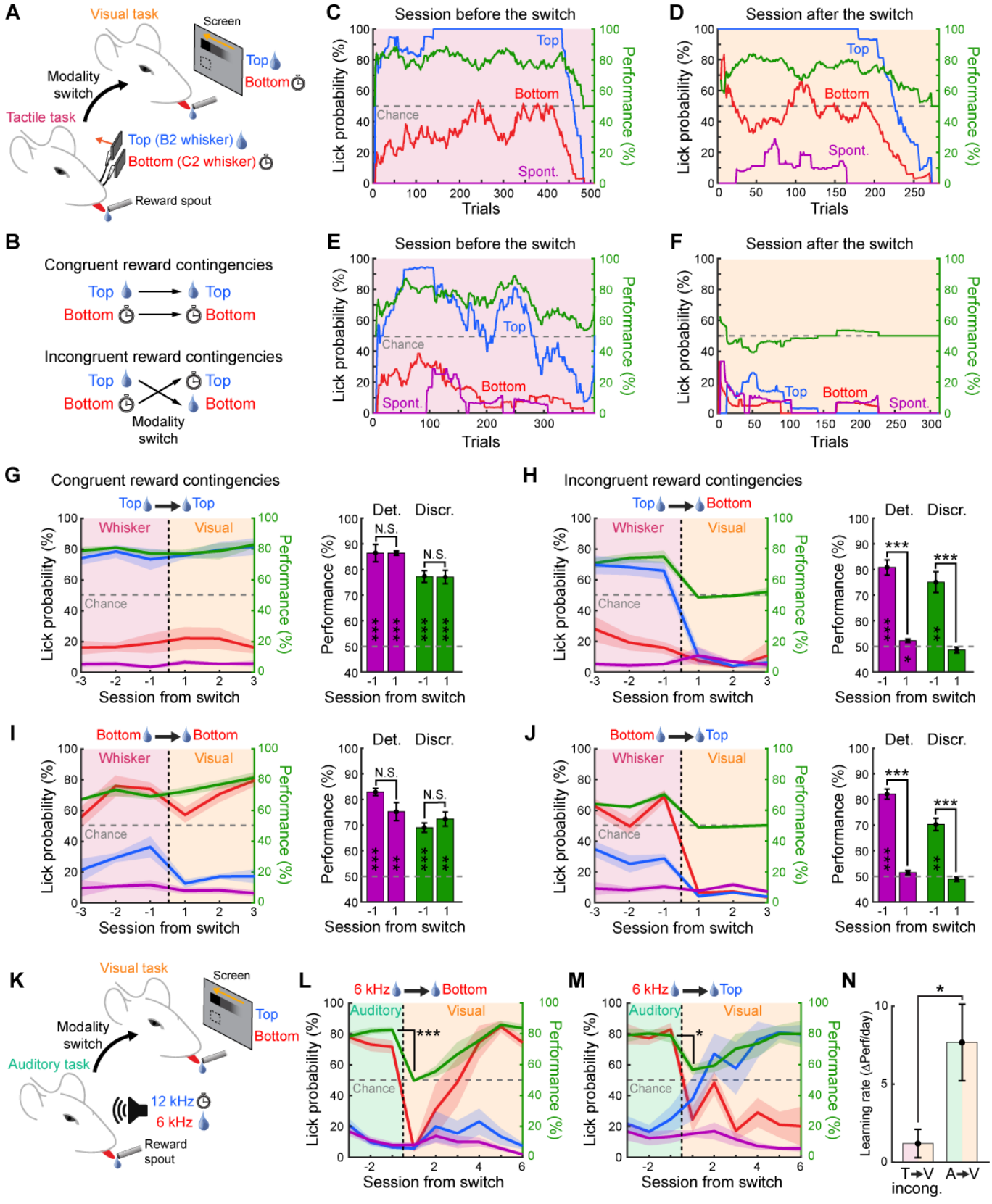

Figure 1

656

Fig. 1. Cross-modal transfer learning in mice performing Go/No go discrimination tasks.

(**A**) Schematic of behavioral paradigm for cross-modal transfer learning from a whisker-based tactile task to a visual task. Mice are trained to a Go/No go whisker discrimination task with two whiskers along a column (B2 and C2 whiskers). Once expert at the task, they are switched to a Go/No go visual discrimination task. (**B**) Two possible scenarios for modality switch: reward contingencies are either congruent between the two modalities (top stimuli remain associated to a reward after the switch and bottom stimuli remain associated to timeout punishment) or incongruent (bottom stimuli are associated to reward after the switch and top become associated to timeout punishment). (**C**) Example session taking place the last day before modality switch for a mouse expert at the whisker discrimination task where the top whisker is rewarded upon licking. Lick probabilities over trials are shown for the top whisker (blue), the bottom whisker (red) and in absence of stimuli (pink). Performance computed as the percentage of correct discrimination trials (see Methods) is shown in green. Performance chance level is indicated with a grey dashed line. Traces shown were computed using a sliding window of 60 trials. (**D**) Same as **C** for the first session after modality switch to a visual discrimination task with congruent reward contingencies. (**E-F**) Same as **C-D** but for a modality switch with incongruent reward contingencies between the tactile and visual task. **(G)** Average daily task performance and lick rates across days for mice population (N=5 mice) switching from the whisker task to the visual task with congruent reward contingencies where the top stimulus is rewarded. Shaded area: S.E.M. Black dashed line indicates the switch between modalities. Color code as in **C**. Histograms in the right indicate the detection (purple) and discrimination (green) performance distribution the day before and after the switch (two-sided paired

t-test comparing days, *p<0.05, **p<0.01, ***p<0.001, N.S. Not significant). Performances are also tested against chance level (two-sided t-test, *p<0.05, **p<0.01, ***p<0.001, Blank: Not significant). **(H)** Same as **G** but for a modality switch with incongruent reward contingencies between the tactile and visual task where the top whisker was rewarded. **(I-J)** Same as **G-H** but for modality switches where bottom whisker was rewarded. **(K)** Schematic of behavioral paradigm where switch occurs between a Go/No go auditory discrimination task with two pure tones (6 kHz and 12 kHz) and the visual task. The 6 kHz tone is always associated to a water reward. **(L)** Same as **G** with the bottom visual stimulus rewarded after switch from auditory task (N=5 mice, paired two-sided t-test comparing the day before and after switch, ***p<0.001; discrimination performance after switch is compared to chance level with a two-sided t-test, p=0.593). **(M)** Same as **L** with the top visual stimulus rewarded after switch from auditory task (N=5 mice, paired two-sided t-test comparing the day before and after switch, *p<0.05; discrimination performance after switch is compared to chance level with two-sided t-test, p=0.104). **(N)** Comparison of learning rate between mice that underwent switch from a tactile to a visual task with incongruent reward contingencies and mice that underwent switch from an auditory task to a visual task (N=10 mice for tactile group and N=10 mice for auditory group, unpaired two-sided t-test, *p<0.05).

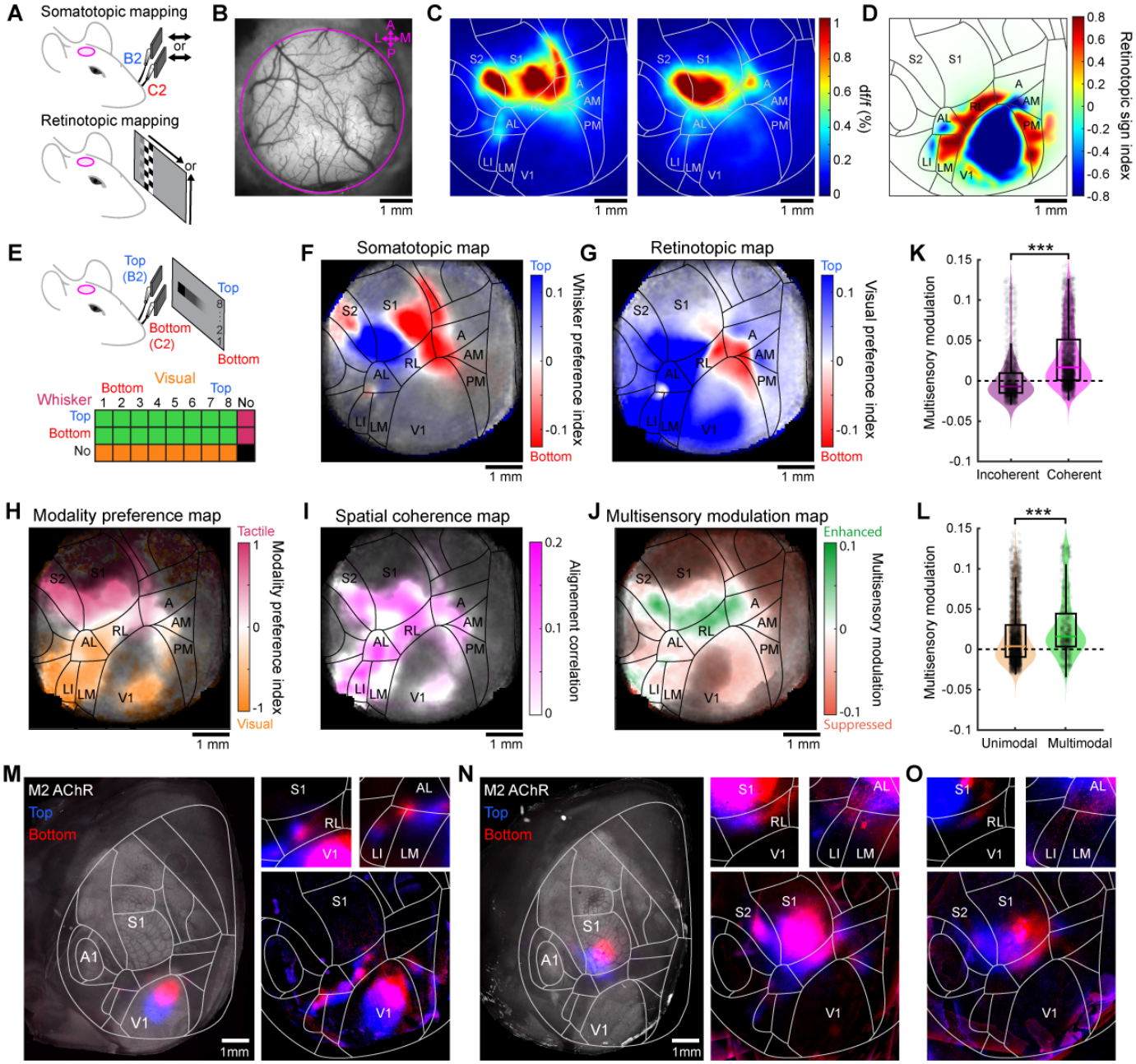

Figure 2

Fig. 2. Functional and anatomical mapping of vertical space in retinotopically and somatotopically organized areas of the dorsal cortex.

(**A**) Schematic of the protocols used to measure single whisker response maps (top) and retinotopic maps (bottom) in transgenic mice expressing GCaMP6f in layer 2/3 imaged through a cranial window. (**B**) Example of a 5 mm diameter cranial window with blood vessels pattern. The perimeter of the window is highlighted in pink and orientation is indicated in top right. (**C**) Response maps averaged across mice (N=41 mice) for C2 whisker (left) or B2 whisker (right) stimulations. Projection of the Allen Mouse Brain Atlas is overlaid on top with areas names. (**D**) Retinotopic sign maps averaged across N=50 mice and realigned to the atlas shown in **C**. (**E**) Schematic of the visuo-tactile sparse noise protocol. All possible combinations of visual and whisker stimuli are depicted in the matrix below. (**F**) Somatotopic maps for vertical space computed from whisker stimuli average across mice (N=29 mice), with transparency defined by response significance in each pixel (see Methods). (**G**) Retinotopic maps for vertical space computed from visual stimuli average across the same mice as in **F**, with transparency defined by response significance in each pixel (see Methods). (**H**) Modality preference maps between visual and tactile responses averaged across the same mice as in **F**. (**I**) Spatial coherence maps between visual and tactile representations (see Methods) averaged across the same mice as in **F**. (**J**) Multisensory maps comparing visuo-tactile responses and combination of unisensory responses (see Methods) averaged across the same mice as in **F**. (**K**) Multisensory modulation index for pixels belonging in regions of high spatial coherence compared to regions with lack of spatial coherence (n=2482 pixels versus n=1444 pixels, unpaired two-sided Wilcoxon test, ***$p<10^{-15}$). (**L**) Multisensory modulation index for pixels belonging to unimodal or multimodal regions (n=3471 pixels versus n=600

pixels, unpaired two-sided t-test, *** $p<10^{-15}$). (**M**) Anterograde labeling of V1 projections using pAAV-CAG-GFP and pAAV-CAG-tdTomato viral constructs injected in the top and bottom retinotopic part of V1, respectively. Left: injection sites in V1 along the iso-horizontal axis. Right: conserved retinotopic organization of projections in associative areas for the mouse shown in the left (top) and averaged across mice (N=3 mice, bottom). (**N**) Same as **M** but for anterograde labeling of S1 projections with two injection sites in S1 in B2 and C2 barrels (N=6 mice). (**O**) Retrograde labeling of S1-projecting neurons using CTB-555 and CTB-647. Top: Example of CTB-labelled neurons spatially organized in associative areas. Bottom: Average CTB distribution across mice of associative neurons projecting to S1 (N=3 mice).

740

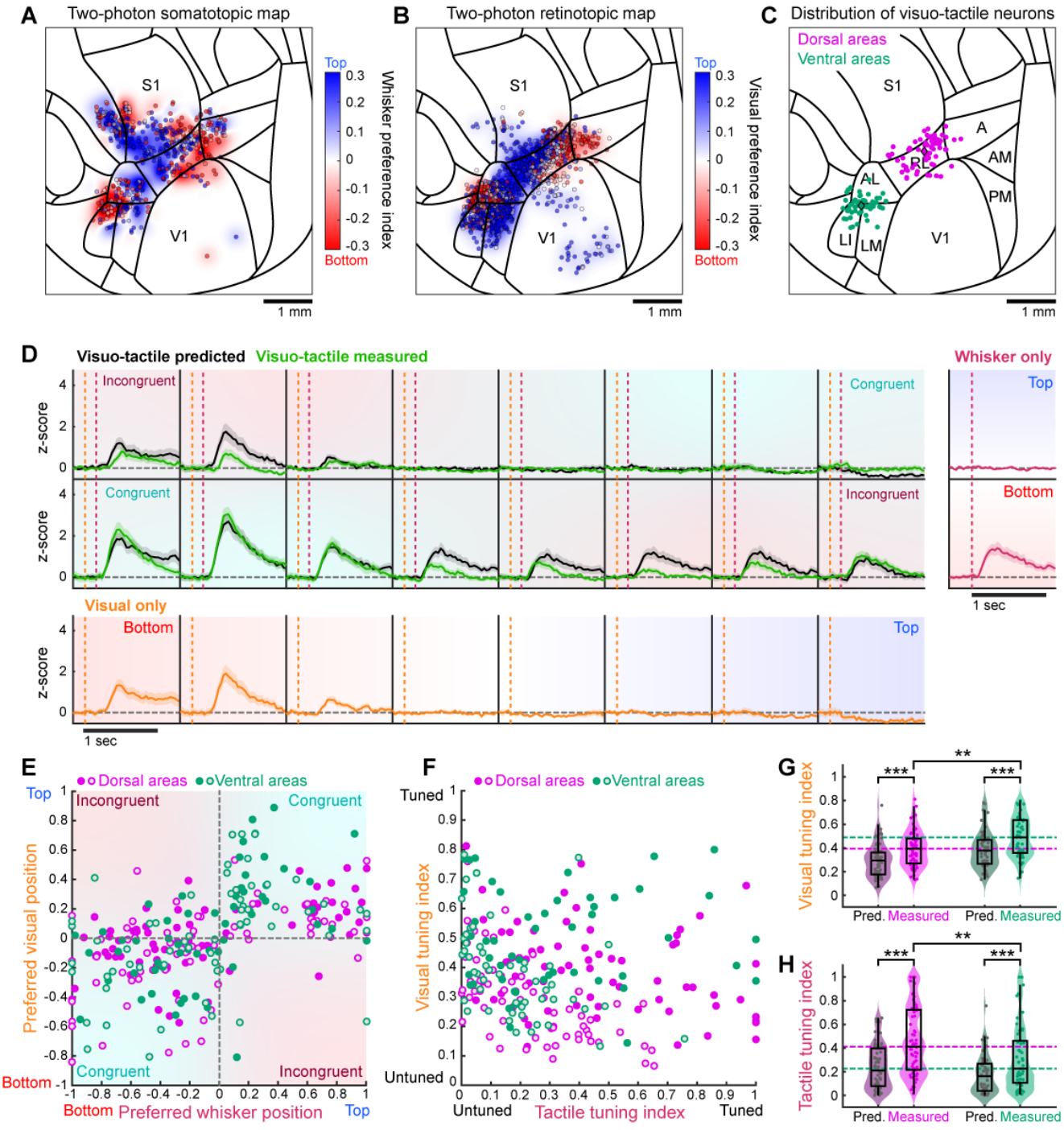

Figure 3

Fig. 3. Visuo-tactile representations of vertical space in single neurons of the dorsal cortex.

(**A**) Population of single neurons imaged with two-photon microscopy with GCaMP6f responses to whisker (n=623 neurons significantly responsive to tactile stimulation) stimuli. Each spot represents a single neuron and the color code indicate preference for the top (blue) or bottom (red) stimuli. Neurons are all realigned to a reference atlas. The reconstructed wide-field map is displayed in the background (see Methods, Pearson coefficient of correlation: 0.373, $p<10^{-16}$). (**B**) Same as **A** for neurons responding to visual stimuli (n=1595 neurons significantly responsive to visual stimulation, Pearson coefficient of correlation: 0.811, $p<10^{-16}$). (**C**) Distribution of all visuo-tactile multimodal neurons identified. Neurons are classified as part of the ventral (cyan) or dorsal (pink) pathway depending on their location (see Methods). (**D**) Response tuning properties of an example neuron from RL. Unisensory responses in z-score are shown on the side for visual (orange) or whisker (violet) stimuli. Based on these responses, predicted responses are shown for visuo-tactile stimuli (gray) together with measured responses (green). Error bars: S.E.M. (**E**) Comparison of preferred visual position and preferred whisker along the vertical space in visuo-tactile stimulation condition for predicted (open circles) and measured (full circles) responses of ventral neurons (n=59 significantly responsive multimodal neurons, Pearson coefficient: 0.39 for predicted with $p< 2.589*10^{-3}$ and 0.53 for measured with $p<1.648*10^{-5}$; 81% of neurons in congruent quadrants) and dorsal neurons (n=70 significantly responsive multimodal neurons, Pearson coefficient: 0.44 for predicted with $p<1.395*10^{-4}$ and 0.57 for measured with $p<2.509*10^{-7}$; 74% of neurons in congruent quadrants). (**F**) Comparison between tactile and visual tuning indices computed from the predicted (open circles) or measured (full circles) visuo-tactile

766 responses of neurons in **E**. (**G**) Comparison of the visual tuning indices in visuo-tactile
767 stimulation condition between predicted and measured responses for ventral and
768 dorsal stream neurons (two-sided paired Wilcoxon test between measured and
769 predicted for ventral: n=51 neurons, ***p=4.096*10^{-4}; for dorsal: n=68 neurons,
770 ***p=2.777*10^{-6}; two-sided unpaired t-test comparing dorsal and ventral measured
771 responses: ***p=1.121*10^{-3}). (**G**) Same as **F** for the tactile tuning indices (two-sided
772 paired Wilcoxon test between measured and predicted for ventral: n=51 neurons,
773 ***p=1.92*10^{-6}; for dorsal: n=68 neurons, ***p=8.53*10^{-10}; two-sided unpaired t-test
774 comparing dorsal and ventral measured responses: ***p=4.932*10^{-3}).

775

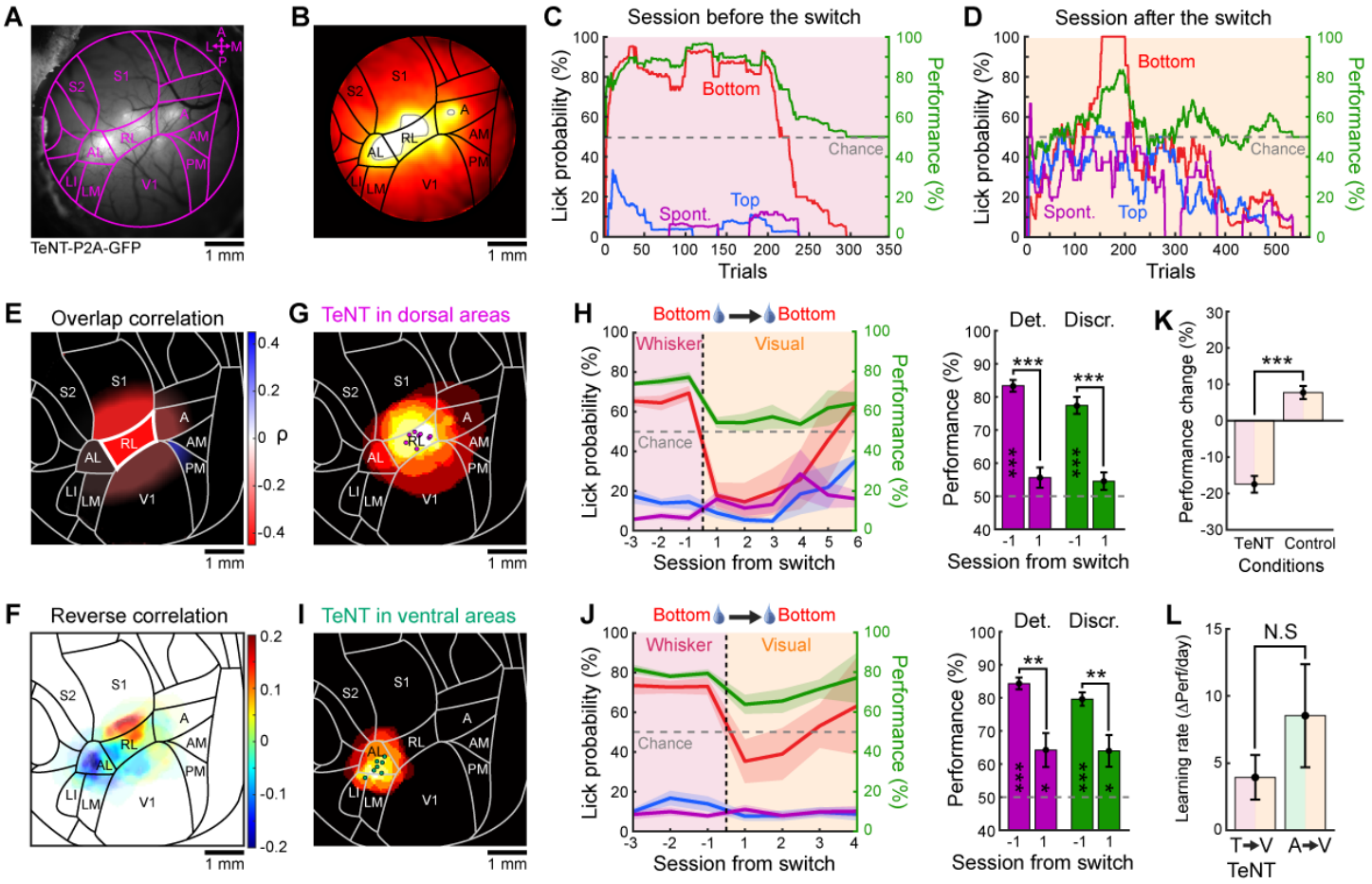

Figure 4

Fig. 4. Loss-of-function of visuo-tactile cortical areas during transfer learning. (**A**) Image of a 5 mm diameter cranial window over the dorsal cortex with expression of TeNT-P2A-GFP in regions where viral vectors were injected. Atlas obtained through functional mapping is overlaid for reference. (**B**) Vessels-free normalized fluorescent expression pattern over the cranial window (see Methods). Gray contour indicates areas considered for correlation analysis. (**C**) Example session taking place the last day before modality switch for a mouse expert at the whisker discrimination task where the bottom whisker is rewarded upon licking. Lick probabilities over trials are shown for the top whisker (blue), the bottom whisker (red) and in absence of stimuli (pink). Performance computed as the percentage of correct discrimination trials (see Methods) is shown in green. Performance at chance level is indicated with a grey dashed line. Traces shown were computed using a sliding window of 60 trials. (**D**) Session following modality switch from a tactile task in **C** to a visual task. (**E**) Area-based correlation between GFP expression overlap and performance drop following modality switch. Color map indicates the Pearson coefficient ρ and areas with $p<0.05$ are indicated with a thick border (RL with p=-0.43 and S1 with p=-0.39). (**F**) Average GFP coverage for mice with impaired transfer learning (average discrimination performance lower than 55%). The map is displayed after subtraction of the average coverage. (**G**) Average GFP coverage of all mice where only dorsal neurons were silenced (N=8 mice). (**H**) Average daily task performance and lick rates across days for mice population in **G** (N=8 mice) switching from the whisker task to the visual task with congruent reward contingencies where the bottom stimulus is rewarded. These mice are selected based on GFP expression overlap with dorsal visuo-tactile areas. Shaded area: S.E.M. Black dashed line indicates the switch between modalities. Color code as in **C**. Histograms in the right indicate the detection (purple) and discrimination (green) performance distribution the day before and after the switch (paired two-sided

t-test comparing days, *p<0.05, **p<0.01, ***p<0.001, N.S. Not significant). Performances are also tested against chance level (two-sided t-test, *p<0.05, **p<0.01, ***p<0.001, Blank: Not significant). (**I**) Same as **G** but for GFP expression in the ventral stream (N=7 mice). (**J**) Same as **H** for ventral area (paired two-sided t-test comparing days, *p<0.05, **p<0.01, ***p<0.001, N.S. Not significant). Performances are also tested against chance level (two-sided t-test, *p<0.05, **p<0.01, ***p<0.001, Blank: Not significant). (**K**) Comparison of performance change following switch between all mice expressing TeNT-P2A-GFP and control mice described in Figure 1I (N=22 mice for TeNT and N=5 mice for control, two-sided unpaired t-test, ***p=0.0008). (**L**) Learning rate averaged over the first three days following switch for mice expressing TeNT-P2A-GFP with impaired transfer learning and control mice described in Figure 1L-M (N=22 mice for TeNT and N=5 mice for control, two-sided unpaired t-test, N.S. p=0.127).

Supplementary Materials:

Materials and Methods:

Animals

All experiments were carried out in accordance with the Institutional Animal Care and Use Committee of the University of Geneva and with permission of the Geneva cantonal authorities (GE/95/19). C57BL/6J and transgenic mouse lines with C57BL/6J background were housed 2–6 mice per cage under a 12/12-h non-inverted light/dark cycle with ad libitum access to food and water. Transgenic mice were obtained as a crossing between Ai148D mice (Jackson Laboratories, stock number 030328) and Rasgrf2-2A-dCre mice (Jackson Laboratories, stock number 022864). Ai148D mouse line is a Cre-dependent reporter line containing a gene encoding the calcium indicator GCaMP6f at the Igs7 locus. Exposure to Cre recombinase through viral vector injections or crossing with Cre-expressing mice resulted in expression of GCaMP6f. Rasgrf2-2A-dCre mouse line expressed a trimethoprim-inducible Cre recombinase directed by endogenous Rasgrf2 promoter/enhancer elements. When induced, Cre recombinase activity is observed in cortical layers 2/3 and other scattered cells of the cortex, hypothalamus, thalamus, and midbrain. Trimethoprim (TMP) i.p. injections were performed for 3 consecutive days at least two weeks before any surgical intervention (0.25 mg/g of body weight diluted in DMSO and 0.9% NaCl). As a result, TMP injected crossed Ai148DxRasgrf2-dCre mice expressed GCaMP6f in cortical layer 2/3. Both males and females of 2-6 months and a weight of approximately 20-25 g were used for the experiments.

Viral vectors and markers

For anterograde labeling, pAAV-CAG-GFP (Addgene, #37825-AAV2, titer: 7×10^{12} vg/mL) and pAAV-CAG-tdTomato (Addgene, #59462-AAV2, titer: 4×10^{12} vg/mL) were used. Volumes of virus injected was usually 50-75nL in each site, without dilution. Brains were collected 3 weeks after the injections. For retrograde labeling, Cholera Toxin subunit B conjugated to Alexa Fluor 555 (Invitrogen, reference number C22843) and Alexa Fluor 647 (Invitrogen, reference number C34778) diluted in PBS were injected (~50nL in each site). Brains were collected 10 days after the injections. For both anterograde and retrograde labeling, the unique depth of injection was 400 µm. For inactivation experiments, an AAV-DJ-CMV-eGFP-2A-TeNT virus was injected in the region of interest (Wu Tsai Neurosciences Institute, reference number GVVC-AAV-70) at 3 different depths along the cortical column (800 µm, 500 µm and 200 µm respectively). Behavioral protocols started at least one month after the injections were done. For both anterograde and inactivation experiments, pictures of the injection sites were taken with the wide-field microscope before brain collection.

Stereotaxic surgeries

Pain management was performed by first administrating the opioid Buprenorphine subcutaneously (0.1 mg/kg) before starting the surgery. A local anesthetic was injected under the skin of the head before the surgical incision (mix Lidocaine/Bupivacaine, 6mg/kg and 2.5 mg/kg respectively). Mice were anesthetized inside an induction chamber with 3% isoflurane mixed with oxygen and then fixed on the stereotaxic apparatus (Model 940, Kopf). A custom-made nose-clamp has been adapted to the apparatus to maintain the position of the animal, allowing head rotation. Body temperature was constantly monitored through a thermic probe and adjusted to ~37°C via a heating pad placed below the mouse (DC Temperature Controller, FHC).

867 Breathing rate was regularly monitored by visual inspection. Ophthalmic gel (Vitamin
868 A, Bausch Lomb) was applied on both eyes to ensure protection from light and prevent
869 them from drying out. During the surgery, mice were anesthetized with a constant
870 isoflurane level lowered around 2%. At the end of the surgery, an anti-inflammatory
871 was also administered subcutaneously (Carprofen, 7.5mg/kg) and animals were
872 warmed with a heating lamp for at least 15 minutes until recovery from anesthesia. A
873 second anti-inflammatory (Ibuprofen, Algifor) was added to drinking water for 3-days
874 post-op and the weight was checked daily to ensure that weight was kept above 15%
875 of the original weight prior to the surgery. All animals were implanted with a cranial
876 implant. After removing the skin and tissues on top of the head, the skull was cleaned,
877 dried and thinned. The mouse head was tilted approximately at a 30° angle, ensuring
878 a better access to the left hemisphere. A custom-made metallic implant was placed
879 on the top of the skull using a custom-made holder. It was then fixed with a layer of
880 glue (Loctite 401, Henkel) and additional layers of dental acrylic (Pala, Kulzer) to
881 solidify the implant. Dental acrylic was covered with black nail polish to prevent light
882 contamination from visual stimuli during imaging experiments. After at least 3 days of
883 recovery, mice could undergo additional procedures. Animals that underwent imaging
884 sessions were implanted with a custom-made glass window composed of a bigger top
885 round cover slip of 7 mm diameter respectively and two superimposed 5 mm diameter
886 cover slips, respectively. The three concentric cover slips were glued together with UV
887 glue (Optical Adhesive n°68, Norland). The craniotomy was the size of the smaller
888 window and was drilled around the region of interest. Before removing skullcap, we
889 used a custom-made perfusion chamber with saline solution (NaCl 0.9%) to rinse the
890 craniotomy continuously and reduce bleeding stains. A custom-made holder with air-
891 suction was used to hold and position the window. The cranial window was gently
892 brought down until it was in contact with the brain. The window was then fixed with UV

glue, super glue and dental acrylic. For stereotaxic injections, glass pipettes (5-000-2005, Drummond) were pulled (Model P-97, Sutter Instrument Co) and further broken to obtain a tip of ~10-15µm inner diameter and further beveled to create a sharp tip and avoid cortical damage during insertion. Injection sites were determined using functional mapping. Injections were done using a single-axis oil hydraulic micromanipulator (R.MO-10, Narishige). The pipette was slowly inserted inside the cortex until reaching the desired depth. Viruses or toxins were injected at a speed of ~2nL/s. When the whole volume was injected, we waited 5 minutes with the pipette in the same position before gently removing it.

Behavioral training

A Matlab custom-made graphical user interface (GUI) was developed from the Matlab App Designer in order to control the behavioral tasks and monitor performance. The GUI allowed real time visualization of the animal's performance and online modification of the parameters (e.g. stimuli parameters, punishment duration, stimuli proportion). Animals underwent water restriction 2 to 4 days before the training started and were handled every day by the experimenter for at least 10 minutes. During the pre-training phase, mice were habituated to head-fixation and placed on the setup. In the first session, only go trials were presented with 100% reward probability. Go trials proportion was progressively reduced and nonrewarded trials were added the following sessions. Mice were trained once per day every day at the same hour.

Six different behavioral tasks (two tactile tasks, two visual tasks and two auditory tasks) were used for our experiments (Fig. 1), all following a Go/No go discrimination paradigm. In the tactile tasks, two whiskers ("top" whisker B2 and "bottom" whisker

C2) were inserted inside glass capillaries each attached to a piezo actuator that could create a small deflection of about 1 mm along the rostro-caudal axis. In one version of the task, B2 whisker stimulations were associated to a reward (go trials). In the other version, C2 whisker stimulations were the go trials. In the visual tasks, two drifting squares (a “top” square and a “bottom” square, relative to a midline in the mouse visual field) were presented on the screen. In one version of the task, top square stimulations were the go trials while in the other version of the task, bottom squares stimulations were the go trials. The total trial duration of a single trial was 4 s: after a 2 s quiet window (during which licking resulted in trial abortion) the stimulus was presented, and the mouse was allowed to lick during a 2 s response window. During go trials, the mouse could obtain a water reward upon licking the spout following the stimulus (Hit trials). Failure to lick would result in Miss trials. During No go trials, the mouse had to refrain from licking (Correct rejection trials, CR) or it was punished with a time-out of 10 s (False Alarm trials, FA). Some trials were presented without stimulus presentation (catch trials). If the mouse licked during catch trials, no time-out was applied. If mice were too compulsive (i.e. licking during the quiet window too frequently), a 10sec time-out early lick punishment could also be applied. The proportion of each trial type was the following: Go trials = 30%, No go trials = 50%, Catch trials = 20%. All tactile stimuli were generated through Matlab data acquisition toolbox controlling a piezo actuator (Bimorph bendor piezo actuator PB4NB2S, Thorlabs) through a National Instrument card. All visual stimuli were generated using Matlab and PsychToolBox. Stimuli were presented on a gray background through a LCD monitor (20x15cm, pixels, 60Hz refresh rate, Pi-shop) positioned 10 cm from the eye, with a 30° angle to the right of the midline. The screen was also tilted with a 30° angle along the horizontal plane to match the mouse head angle with the intent of roughly aligning the bottom and top parts of the screen to the resting position of the C2 and B2 whiskers in the mouse

visual field. Stimulations during behavior consisted in black squares moving through the screen in the rostro-caudal direction (bar width was 15°, stimulus duration was 175ms, speed was 500°/sec), on a gray background. Finally, an auditory task was considered following the same structure as the other tasks but using two short tons of 6 kHz and 12 kHz as Go and No go stimuli, respectively. These tones were delivered from a speaker located next to the mouse on the same side as the visual and tactile stimuli.

Wide-field microscopy

We used a custom-made wide-field epifluorescence microscope setup (*54*) including a sCMOS camera (ORCA-Flash4.0 V3, Hammamatsu). Magnification was determined through a 0.63X C-mount camera adapter for Olympus Microscopes. The field of view size was 5.6 mm x 5.6 mm. The camera and adapters were mounted on a base allowing vertical movement with manual focus. LED white illumination (740mW, 1225mA, Thorlabs) could be controlled via a T-Cube LED driver (LEDD1B, Thorlabs). Filter cubes could be changed for different type of imaging. For imaging GCaMP6f, GFP excitation, emission and dichroic filters were used. For imaging intrinsic signals, Cy3/5 excitation, emission and dichroic filters were used. Objective (MVX Plan Apochromat with 2x, Olympus) was attached to the microscope base. The somatotopic mapping consisted in repetitive rostro-caudal pulsatile deflections (~1 mm amplitude) of either B2 or C2 whiskers for 20-80 trials followed by quiet window. The retinotopic mapping protocol consisted of drifting bars. A contrast reversing checkerboard was presented within the bar to better drive neural activity (0.04 cyc/ ° of spatial frequency and 2 Hz of temporal frequency). In each trial the bar was swept in the four cardinal directions: left to right, right to left, bottom to top, and top to bottom.

Single trials were repeated in average 20-40 times. For anatomical experiments, only C57BL/6J mice were used. To record the intrinsic optical signal, we used longer tactile stimulations and slower visual stimulations. Mice were fixed on the platform and anesthetized during the procedure with isoflurane level lowered to 1%. Body temperature was monitored with a probe and adjusted to 37°C using a heating pad (DC Temperature Controller, FHC). For all the other experiments, we used Ai148D x Rasgrf2-2A-dCre mice and *in vivo* calcium imaging. The visuo-tactile sparse noise protocol consisted in combinations of visuo-tactile stimuli: three whisker conditions (C2 whisker, B2 whisker and no whisker), eight vertical positions of the moving square. The ninth position corresponded to a no visual stimulus condition (i.e. blank screen). Visual and tactile stimulus onsets were either synchronous or delayed (150ms delay visual leading tactile stimulus). In total, 54 different combinations were presented in pseudo-random order with a 1 sec interval, and the full sequence was repeated 60 times. Total sparse noise protocol duration was approximately 1 hour. At the beginning of each recording, a picture of the window's surface with blood vessels pattern on focus was taken as reference image and the focus was then set ~300µm under the surface to maximize signal collection. Light was adjusted to prevent saturation. Before each imaging session, the window was cleaned with 70% ethanol and eyes hydrated with mineral oil.

Two-photon microscopy

The two-photon microscope was custom-made (INSS Company). It consisted in a femtosecond laser with wavelength range 690-1040nm (Tunable Ti:Sapphire with dispersion compensation MaiTai DeepSee, Spectra Physics) which beam was displaced with Resonant/Galvo scan mirrors (INSS) and the emitted signals were

detected with 2 GaAsP amplified PMTs (PMT2101/M, Thorlabs). Imaging was performed through ScanImage (Vidrio Technologies). Images were acquired at approximately 30 frames per second. Two-photon calcium recordings during the visuo-tactile sparse noise protocol described in the wide-field section were performed on N=25 Ai148D x Rasgrf2-2A-dCre mice. After cleaning the glass window, a hydrophobic chamber was made between the head plate and the imaging platform using liquid plastic (Smooth-Cast 325, Smooth-On) and the objective was immerged in distilled water. We ensured no polluting light could reach the objective by covering it with a dark protection and by turning off the light room. At the beginning of each recording, an anatomical picture of the field of view was taken with a LED. When switching the microscope in two-photon mode, we took a picture of the surface blood vessels at magnification x1 and x1.5 for further realignment. For each field of view, we usually recorded at three different depths ranging between 300 µm and 50 µm below the surface. No more than two recordings were done during a day on the same animal.

Histology

Mice were anesthetized with 3% isoflurane and euthanized with i.p. pentobarbital injection (Eskornarkon, 150mg/kg). They were then transcardiacally perfused using a peristaltic pump (ISM829, Cole-Parmer) with 0.01 M PBS, pH 7.4 for 2 min, and then 4% paraformaldehyde (PFA) in phosphate buffer (PB; pH 7.4) for 3 min. The brains were post-fixed at 4°C in PFA for 48h and then transferred into PBS. For anatomical experiments of Figure 2, the brains were post-fixed at 4°C in PFA for 2h and washed 3x15min in PBS. The left cerebral hemisphere was separated from the right hemisphere and subcortical parts were removed with a spatula. The left hemisphere was flattened between glass slides and kept flattened at 4°C in PFA for 12h. Flattened

hemispheres were embedded in agar gel 4% before cutting slices of 60µm thickness with a vibratome (VT 1000 S, Leica). The sections were mounted on glass slides using mounting medium with DAPI (ab104139, Abcam), coverslipped and kept in dark at 4°C until further imaging.

Immunohistochemistry

Brain sections were incubated with slight agitation (40 rotations/min) for 2h at room temperature in a saturation/permeabilization solution containing a mix of 5% BSA and NGS, 0.3% triton X-100 and PBS. Brain sections were then incubated with slight agitation (40 rotations/min) overnight at 4°C with a rat anti-muscarinic acetylcholine receptor m2 primary antibody (1:500, Sigma-Aldrich, reference number MAB367) in the same blocking solution. Sections were washed 3x15min in PBS before a 2h incubation with either a donkey anti-rat secondary antibody conjugated with Alexa Fluor 488 (dilution 1:500, Invitrogen, #A21208) or with a goat anti-rat secondary antibody conjugated with Cy5 (dilution 1:500, Invitrogen, #A10525) in blocking solution at room temperature without agitation. Hoechst solution (dilution 1:1000, Invitrogen, # 33342) was used for fluorescent nuclear counterstaining (Fig. 2M-O).

Histological imaging

Slices were mounted on Superfrost microscope slides (Epredia) with mounting medium (Fluoromount) and covered with 24x50 mm coverslips (Menzel-Gläser). All photomicrographs were taken using a Zeiss Axio Scan.Z1 or a Zeiss Confocal LSM800 Airyscan at the bioimaging platform of the University of Geneva.

Wide-field calcium imaging analysis - retinotopic and whisker mapping protocols

Responses to drifting checkerboard stimuli were averaged across trials for each condition and converted in df/f using a pre-stimulus time window as baseline. Azimuth and elevation preference maps were then computed as in (*28*).To segment visual areas, azimuth and elevation maps were combined to generate a visual field sign map. The sign map was computed as the sine of the difference between the vertical and horizontal retinotopic gradients for each pixel. Somatotopic maps were obtained by averaging the response df/f across B2 or C2 stimulation trials over a short time window of 200ms.

Wide-field calcium imaging analysis – sparse noise protocol

After an initial down sampling to a resolution of 100 x 100 pixels by bicubic interpolation, 50 Hz framerate fluorescence videos acquired during the sparse noise protocol underwent pixelwise notch filtering (f_0=12 Hz, f_w=6 Hz) and discrete wavelet transform (DWT) detrending (setting lowest frequency approximation coefficients to zero in a 6-level decomposition) to remove artifacts and low-frequency drifts. After these initial preprocessing steps, videos where z-scored using as reference distribution the ensemble of all pixel values corresponding to blank stimuli (i.e. trials with no visual and no tactile stimulation). Next, trials with a high baseline activity (>=0.5 z-score on average, corresponding to a bump of spontaneous activity preceding the stimulus) were discarded and z-scoring and baseline subtraction was performed again including only remaining trials. Starting from the z-scored data obtained in this way, response surprises (i.e. -log10($p_{t\text{-}test}$)) across trials were computed for each time

samples and each stimulus in order to define spatial responsivity masks. Average visual or tactile top or bottom response videos were calculated by averaging the median responses across trials to the relevant stimuli. For “visual top” all stimulation conditions (either unimodal of multimodal) in which the visual stimulus was present in positions 7 or 8 of our grid were used whereas for “visual bottom” all visual conditions including positions 3 or 4 were used. Retinotopic and somatotopic difference maps like the ones shown in Figure 2F-G and Supplementary Fig. S4D were obtained by taking the difference between average visual top and bottom response videos. Modality preference maps (Fig. 2H and Supplementary Fig. S4D) were obtained by taking an element-by-element max between the top and bottom videos of a given modality and subtracting to the result the same operation in the other modality. This was normalized by the sum of the same elements. Multisensory enhancement maps (Fig. 2J and Supplementary Fig. S4D) were obtained by taking an element-by-element max between the average unimodal response video to V and T (all positions) as “predicted response” and subtract it to the “observed response” obtained as the average of all VT conditions. Retinotopy-somatotopy spatial coherence maps (Fig. 2I) were computed by convolutionally (stride = 1 patch size = 8 pixels) calculating the correlation-based similarity of matching patches of the two difference maps taken as input and then scaling the result by their average peak-to-peak range (to make it more noise robust). Pixel level comparison of grand average maps (Fig. 2K-L) where performed on responsive pixels (average surprise threshold defined such as response $p_{t\text{-}test} <= 0.05$) comparing pixels with preference or coherence value significantly different or not from zero (threshold $p_{t\text{-}test}= 0.05$ across mice). Grand average maps displayed in Figure 2F,G,J are obtained by averaging frames of the corresponding grand-average video (obtained by realigning single mouse ones to the common atlas and averaging) in different windows: frames 16 to 20 for the somatotopy (V-T delay =

0 s), 26 to 30 for retinotopy (0 ms V-T delay condition) and 19 to 23 for modulation (150 ms V-T delay condition). These windows were chosen to roughly match the peak of the signal. Signal time courses displayed in Supplementary Fig. S4E-H were also obtained by integrating grand-average videos over regions of interest defined by the intersection of an area-based mask (obtained from the common atlas) and spatiotemporal responsivity mask selecting pixels with average (across mice) max (across relevant conditions) response surprise above a given threshold (corresponding to $p_{t\text{-}test} <= 0.01$). This was done to include only the responsive subregions of each area at every given timepoint. To measure angular mismatch between retinotopic and somatotopic maps we computed elevation gradient vectors from the grand-average version of these maps (using Matlab "gradient" function with a scale of 4 pixels followed by a gaussian smoothing of gradient components). Then a region of interest overlapping with RL was defined and we computed the average angular difference between the two gradient fields limited to pixels within it.

Atlas fitting and registration

The reference Allen Institute mouse brain atlas (*55*) projected to match the skull tilt in our experiments was manually fitted to the wide-field imaging field-of-view for each mouse by visually aligning the atlas boundary lines to reproducible landmarks from functional maps. Both the maps obtained from the sparse noise protocol (i.e. vertical retinotopy and somatotopy difference maps) and from the whisker and retinotopic mapping protocols were used to register the atlas. Landmarks used include the outlines of the sign map regions, the position of C2 and B2 whisker activity bumps in S1 and S2, the reversal of vertical retinotopy at the boundary of each visual area (see Supplementary Figure S4 for examples). To be able to reconnect microscale and

macroscale level information conveyed by two-photon and wide-field imaging experiments, we also manually reconstructed the position of each two-photon field-of-view (FOV) in the frame of reference of the atlas fitted to the wide-field microscope FOV. To do so, we aligned the blood vessel pattern visible in each two-photon FOV with the one visible in the wide-field blood vessels image (as shown in Supplementary Fig. S6F). To pull information across mice ad compute grand averages, we developed a Matlab pipeline to robustly realign atlases to one another. This pipeline is based on iterative application of the image registration algorithm implemented by Matlab function “imregtform” (considering “rigid” transformations in “monomodal” mode) to stacks of atlas boundary images. This pipeline enabled us to obtain the rotation and shift required to register all maps and neuron positions from one mouse to the frame of reference of a common atlas. For the tracing experiments performed in Figure 2M-O, a similar processing was performed with ImageJ using registration of atlases fitted using the M2AChR staining. Brain from all mice with similar injections could then be averaged together after highlighting projections structures with the function “unsharp mask” to highlight small structures away from the injection site.

Two-photon calcium imaging analysis

To extract time-varying somatic calcium signals, we used the Suite2p toolbox (*56*). Neuropil contamination was corrected by subtracting the fluorescent signal from a surrounding ring $F_{Surround}(t)$ from somatic fluorescence: $F(t) = F_{Soma}(t) - \alpha * F_{Surround}(t)$ with $\alpha=0.7$. Neuropil-corrected fluorescence signals F(t) where then converted in z-score by subtracting from each trace the mean value and dividing by its standard deviation of F(t) over the samples contained in the last 0.2 s of the baseline window preceding the stimulus (pooling across all trials). Starting from these z-scored

fluorescence traces we computed response surprises (i.e. -log10($p_{t\text{-}test}$)) across trials for each timesamples of each stimulus. Similarly, we computed the coefficient of variation (CV) across trials quantifying response variability. Neuron was considered reliably responsive to a given stimulus if the surprise exceeded the threshold $Surprise_{th}$ = 8 (i.e. $p_{t\text{-}test} <= 10^{-8}$) while remaining below a coefficient of variation threshold value CV_{th} = 4 for at least 4 consecutive time bins. Neurons responsive to at least one stimulus condition were considered "responsive" and included in subsequent analyses. Neurons responsive in at least one tactile-only condition (i.e. in absence of visual stimulation) are considered "tactilely responsive". Neurons responsive in at least one visual-only condition (i.e. in absence of visual stimulation) are considered "visually responsive". Neurons satisfying both conditions were considered "multimodal" neurons (and used for the analyses displayed in Figure 3). Observed multisensory responses were compared to a "max model" of multisensory interaction (*10*) (as in Fig. 3D). Predicted visuo-tactile z-scored traces for each multisensory stimulus condition were computed trial-wise by taking the max (for each timebin) of the observed response traces observed in response to the same visual and tactile stimulus presented alone (i.e. the corresponding unimodal conditions) and subsequently averaging across trials. Average responses (predicted or observed) were computed over a response window spanning from 0.2 s to ~0.7 s following stimulus onset for each stimulus and each responsive neuron. Preferred vertical position in the tactile or visual space of each neuron was quantified as the center of mass of these average responses along the corresponding dimension of the stimulus grid, either restricting to unimodal or multimodal (observed or predicted) conditions. For further analysis and visualization (as in Fig. 4A,B,E) this center of mass was converted in a "position preference index" ranging from +1 for neurons tuned to the top visual/B2 whisker stimulus to -1 for neurons tuned to the bottom visual/C2 whisker stimulus. The

sharpness of positional tuning of the neurons was assessed computing visual (as the inverse of the best fit sigma parameter of a gaussian fit of the average response to each visual position) and tactile (as the absolute value of the tactile position selectivity index described above) selectivity indices. These indices range from 0 for a neuron responding equally to all positions (i.e. completely positionally untuned neurons) to 1 for a neuron responding to only one position (i.e. maximally positionally tuned neurons). Figure 3E shows the value of the above-mentioned positional preference indices computed on the predicted and observed multisensory conditions whereas Figure 3F shows the value of the above-mentioned selectivity indices for the preferred (i.e. max response) multisensory condition. Neurons were labeled as ventral or dorsal by running a k-means clustering algorithm (with $n_{centroids}$=2) on the spatial distribution of multimodal neuron over the surface of dorsal cortex (displayed in Figure 3C).

Behavioral analysis

To quantify mouse behavior over time within single sessions we isolated different stimulation conditions (i.e. Go trials, No go trials and Catch trials) to compute a smoothed conditional lick probability over a sliding window of 60 trials. The identity of Go and No go stimuli depends on the task condition and vary across mice groups and switch days. Conditional lick probabilities or “lick rates” for short, where calculated as the number of times a mouse licked in a stimulus condition divided by the number of occurrences of that given condition over the sliding window. These curves are the ones displayed in Figures 1 and 4 and Supplementary Figure S1-3 and are labelled “Top”, “Bottom” and “Spont”. From Go and No go lick rates, we also computed a “Performance” variable as the percentage of correct trials for discrimination. Specifically, we defined behavioral performance as the average between the rate of

correct choices upon delivery of the Go stimulus (i.e. Hit rate = lick rate for Go trials) and the rate of correct choices upon delivery of the No go stimulus (i.e. Correct Reject rate = 1 – False Alarm rate where False Alarm rate = lick rate for No go trials). This way of computing performance makes it insensitive to the proportion of Go and versus No go trials. Following a similar logic, we defined a "Detection" performance as the max between the average of Hit rate and Catch Correct Reject rate (i.e. the Go detection rate) and the one of False Alarm rate and Catch Correct Reject rate (i.e. the No go detection rate). Behavioral states were classified as previously described in (*29*). Average rates and performance per session were computing only during engaged trials of each session (i.e. cutting the final part of the session in which the lick rate of animals drops due to task disengagement). Switch-aligned performance trajectories shown in Figure 1 and 4 and Supplementary Fig. S1, S3 and S7 were obtained realigning the curve of average performance and rates around the day of the task switch. Accompanying barplots display the difference of discrimination and detection performance between the last day of pre-switch behavior and the first day after switch. In Supplementary Figure S2 the average rates, performance and detection trajectories across mice that underwent whisker removal control sessions are shown. These average curves were obtained by dividing each session in three chunks: "before", "during" and "after" whisker removal and averaging them separately after a stretching interpolation to match their length. Accompanying barplots display the average discrimination and detection performance in these three chunks (excluding boundary timepoints to avoid contamination due to the temporal smoothing induced by the sliding window we used to compute the rates). Learning rates following switch (Fig. 1N and Fig. 4L) were obtained averaging the slope of the best fitted line for the 3 days after switch for all mice in each group, when necessary, excluding animals not showing a drop to chance level the day after the switch. The relationship between the TeNT-

P2A-GFP expression overlap with each brain areas and the congruent cross-modal transfer learning impairment was assessed using Pearson correlation. Specifically, we calculated the Pearson correlation coefficient between two vectors: 1) the vector representing the fractional overlap of TeNT expression binary masks (displayed in Supplementary Fig. S7E) and each atlas-defined brain area for each mouse, and 2) the vector representing the change in performance following a task switch for each mouse (performance averaged over 3 days after switch). This analysis was performed for each brain area (examples for areas RL and AL are shown in Supplementary Fig. S7F) and displayed as a correlation map in Fig. 4E. With a similar aim, inspired by reverse correlation techniques, we computed the difference between the average of all TeNT-injected mice expression binary masks (see next section) and the one of mice displaying impairment of cross-modal transfer ability (defined as mice performing below 55% of correct trials on average in the first day after congruent switch). This difference will take null or negative values in areas where the manipulation does not affect the behavior and positive values in regions important for producing the behavioral effect (i.e. regions where TeNT expression is enriched in the impairment-triggered ensemble compared to the all TeNT mice one).

Quantification of TeNT expression and ventral/dorsal grouping

Wide-field fluorescence images of dorsal cortex displaying the pattern of expression of TeNT-P2A-GFP in each injected mouse were first preprocessed in order to remove the blood vessels from the surface. This was necessary to obtain a smooth, non-occluded estimate of the spatial profile of construct expression across the surface of cortex below the cranial window. To do so we first segmented the blood vessels with adaptive thresholding (using Matlab function “adaptthresh”) and then inpaint them by

local interpolation (using Matlab function “regionfill”) (Fig. 4A-B and Supplementary Fig. S7A-D). Next, we computed a median-subtracted map of df/f calculated with respect to a manually annotated region of interest on the border of each window (selected to correspond to a fluorescence profile stable around baseline level, usually at the farthest end of the window with respect to the injection site). To make the images acquired in different mice comparable despite potential differences in imaging conditions we artificially re-saturated all images at the same level - i.e. clipping the values in the map of each mouse to the minimum (across mice) max value (within each map). After normalizing these resaturated df/f maps to 1 we computed their 0.90 contour to define the binary TeNT expression mask to be used in subsequent analyses to evaluate the spatial extent of toxin expression. The overlay of these masks for all mice is displayed in Supplementary Fig. S7E. Next, TeNT injected mice were subdivided in “dorsal” and “ventral” expression subgroups by computing a d-prime measure (distance normalized by patch size) with respect of the two multimodal neuron clusters found by k-means in the two-photon analysis of Figure 3. Neurons outside the region of low d-prime with patch centroid falling respectively within AL/LI or A/RL respectively were categorized as ventral and dorsal respectively (Fig. 4G and I show the overlay of the binary masks together with their centers as dots over the common atlas).

Quantification and statistical analysis

Statistical details of experiments and analysis are described in figure legends and in the Results section. Details include statistical tests used, sample type and number as well as definition of bar plots and error bars. In figure legends, standard error of the mean (S.E.M.) is specified when plotted as error bars. Paired or unpaired t-tests were

used to assess significance of mean comparisons (implemented by Matlab functions "ttest" and "ttest2" respectively). Normality tests were not performed systematically but individual data points were plotted to visualize distributions. Wilcoxon signed rank test were used to assess significance in paired median comparisons (implemented by Matlab function "signrank"). Wilcoxon rank sum test was used for unpaired median comparisons (implemented by Matlab function "ranksum"). Pearson correlation coefficient was used to compute correlations between two conditions (implemented by Matlab function "corr"). Across all fits reported uncertainties (i.e. confidence intervals) for best-fit parameter values were extracted fit covariance matrices (fitting was performed using Matlab function "nlinfit").

1279

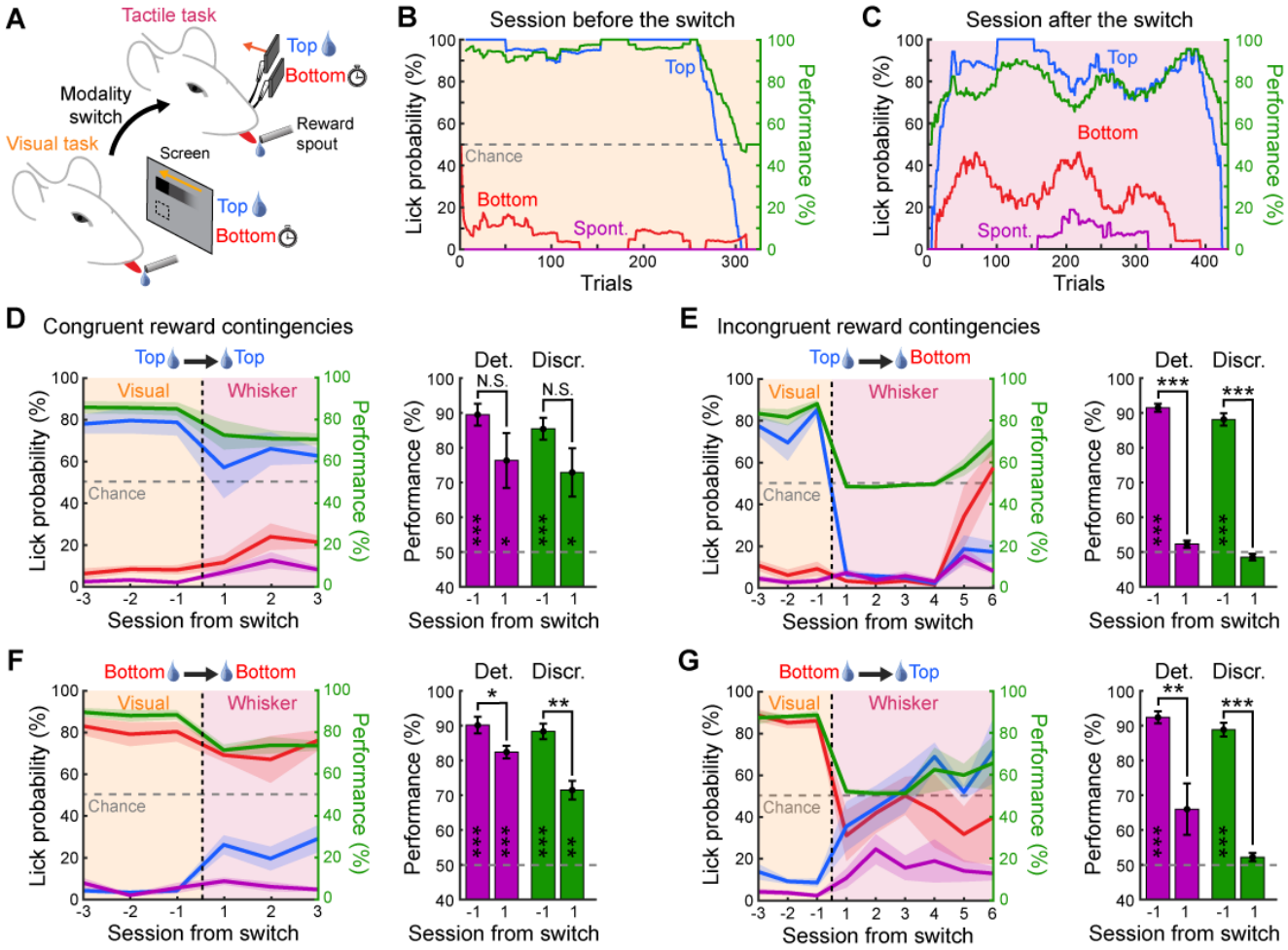

Figure S1

Fig. S1. Cross-modal transfer learning in mice after a switch from a visual discrimination task to a whisker discrimination task.

(**A**) Schematic of behavioral paradigm for cross-modal transfer learning between a visual task to a whisker-based tactile task. (**B**) Example session taking place the last day before modality switch for an expert mouse at the visual discrimination task where the top stimulus is rewarded upon licking. Lick probabilities over trials are shown for the top square (blue), the bottom square (red) and in absence of stimuli (pink). Performance computed as the percentage of correct discrimination trials (see Methods) is shown in green. Performance chance level is indicated with a gray dashed line. Traces shown were computed using a sliding window of 60 trials. (**C**) Same as **B** for the first session after modality switch to a whisker discrimination task with congruent reward contingencies. **(D)** Average daily task performance and lick rates across days for mice population (N=5 mice) switching from the visual task to the whisker task with congruent reward contingencies where the top stimulus is rewarded. Shaded area: S.E.M. Black dashed line indicates the switch between modalities. Color code as in **B**. Histograms in the right indicate the detection (purple) and discrimination (green) performance distribution the day before and after the switch (paired two-sided t-test comparing days, *p<0.05, **p<0.01, ***p<0.001, N.S. Not significant). Performances are also tested against chance level (two-sided t-test, *p<0.05, **p<0.01, ***p<0.001, Blank: Not significant). **(E)** Same as **D** but for a modality switch with incongruent reward contingencies between the visual and whisker task where the top visual stimulus was rewarded (detection after switch, p=0.098; discrimination after switch, p=0.196, N=5 mice). **(F)** Same as **D** but for modality switches where bottom visual stimuli are rewarded. **(G)** Same as **E** but for modality switches where bottom

visual stimuli are rewarded (detection after switch, p=0.096; discrimination after switch, p=0.143, N=5 mice).

1306

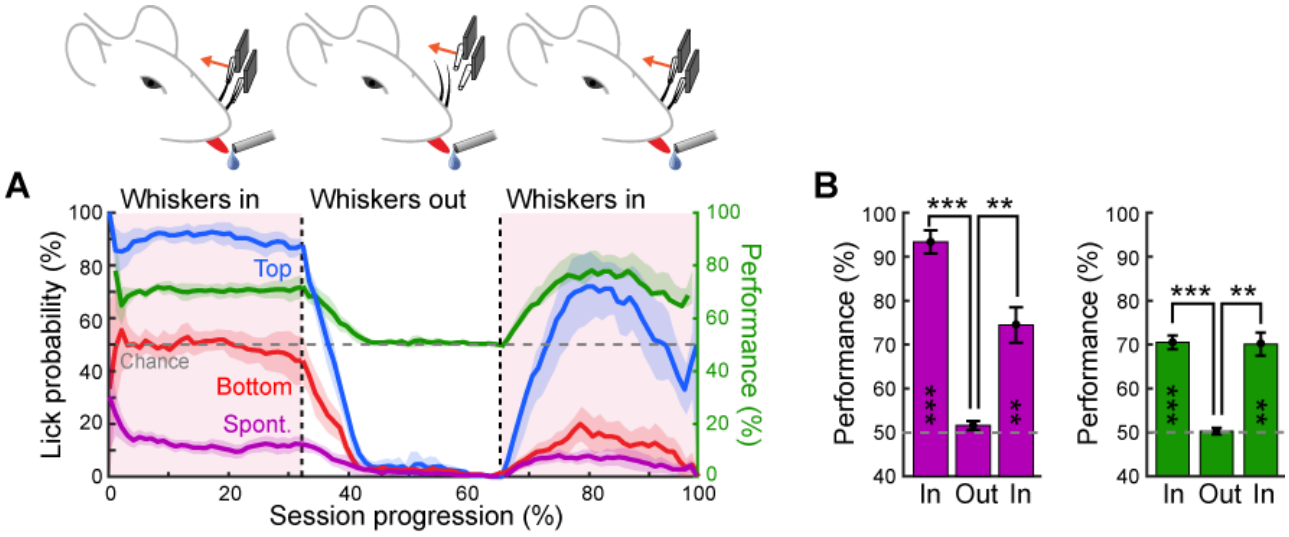

Figure S2

Fig. S2. Mice performance at the whisker discrimination task with and without whiskers in the capillary tubes used to deliver tactile stimulations.

(**A**) Behavioral responses during the whisker discrimination task averaged across expert mice (N=5 mice). Sessions were normalized between beginning (0%) to end (100%) before average. Color code is the same as in Figure 1G. In these sessions, mice were performing the task for about a third of the session before whiskers were removed from the capillary tubes which however remained in their position. The task was then continued for the same duration before reinserting the whiskers back in the tubes to control that mice are still motivated to perform the task.

(**B**) Comparison of detection (purple) and discrimination (green) performance during the 3 different phases of the session described in A. Comparison between consecutive phases (paired two-sided t-test, ***p<0.001,**p<0.01,*p<0.05) and with chance level (two-sided t-test with p=0.188 for detection and p=0.710 for discrimination during the Out phase) are performed. Mice are not able to perform the task in absence of direct whisker stimulations.

1323

1324

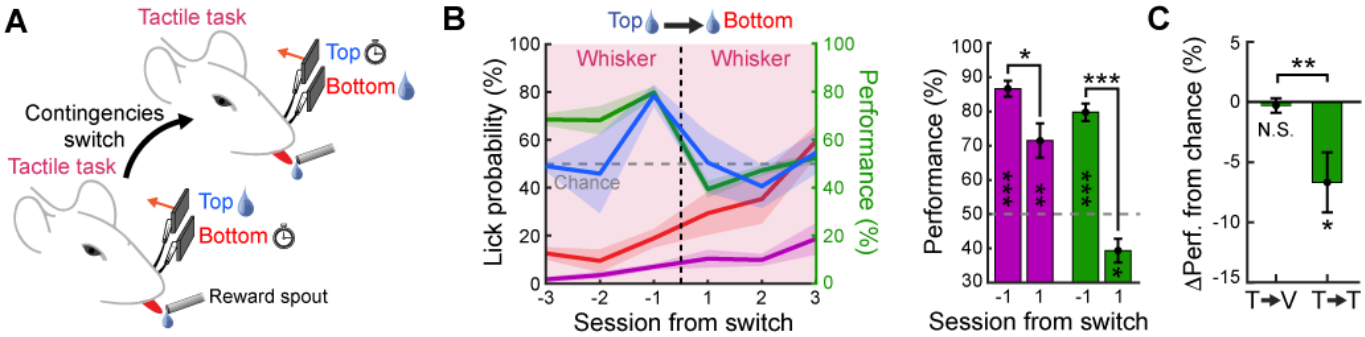

Figure S3

Fig. S3. Relearning to perform the whisker discrimination task after reward contingencies switch within the same modality.

(**A**) Schematic of behavioral paradigm for reward contingencies switch within the same whisker modality. Once mice become expert at the whisker discrimination task where the top whisker is rewarded upon licking, the reward contingencies are reversed so that the bottom whisker is now associated to a reward. (**B**) Task performance averaged over whole sessions across days for mice population (N=7 mice) with reversed reward contingencies. Shaded area: S.E.M. Black dashed line indicates the transition from one task to another. Histograms in the right indicate the detection (purple) and discrimination (green) performance distribution the day before and after the switch (paired two-sided t-test comparing days, *p<0.05, **p<0.01, ***p<0.001, N.S. Not significant). Performances are also tested against chance level (two-sided t-test, *p<0.05, **p<0.01, ***p<0.001, Blank: Not significant). (**C**) Comparison of performance change relative to chance level after reward contingencies switch between modalities (tactile to visual) or within whisker task (N=10 mice for visuo-tactile and N=7 mice for tactile only, unpaired two-sided t-test, *$p=0.668*10^{-3}$). Mice continue to respond to whisker stimulations significantly above chance level but perform the discrimination significantly below chance level. This indicates that they exhibit less flexibility or response suppression for reverse reward contingencies within the same modality. This could be cause to a stronger habitual component on the previous rewarded whisker or a pre-existing prior on the relevance of the whisker stimulations.

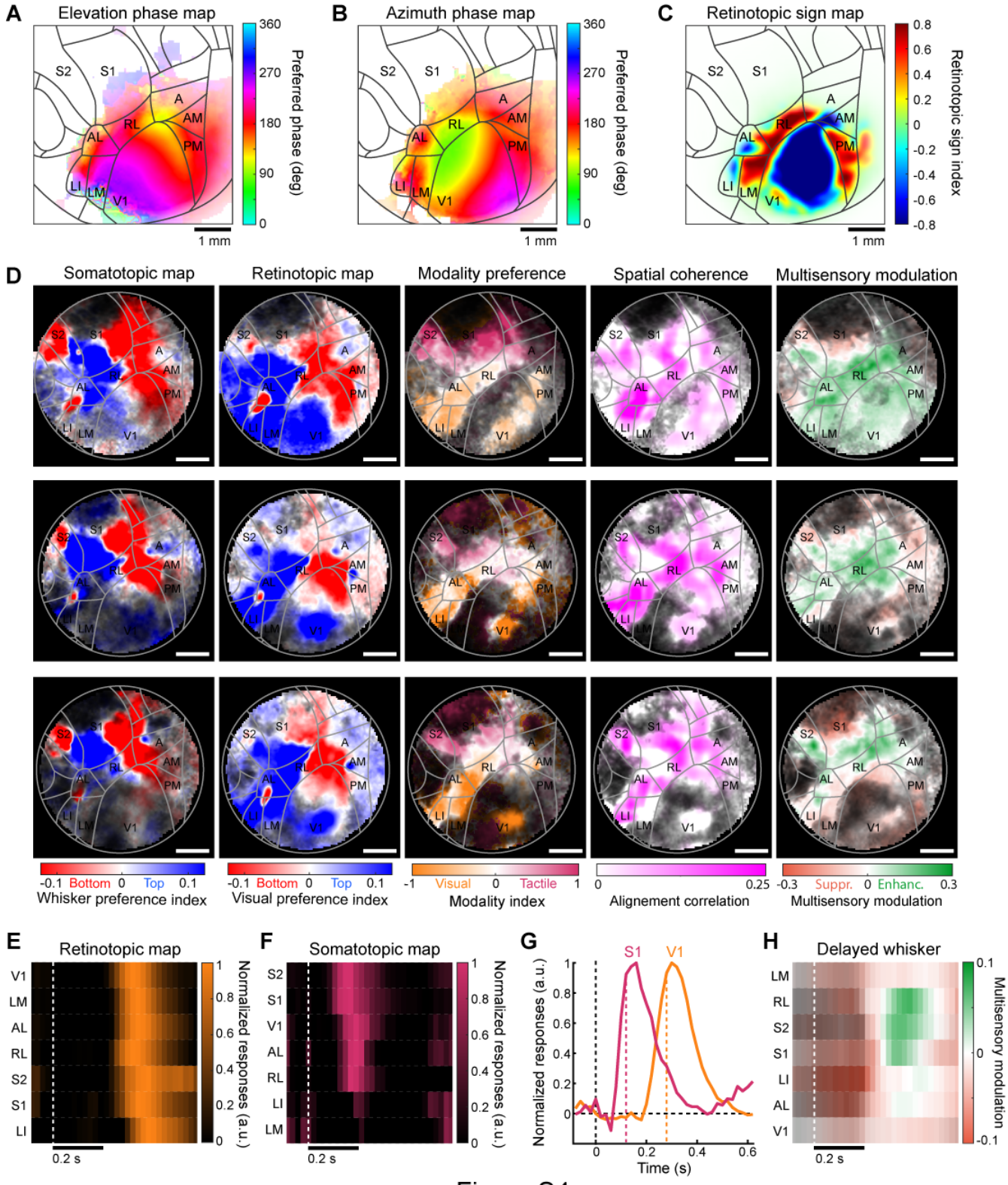

Figure S4

Fig. S4. Average phase maps for retinotopic mapping and examples of dynamic visuo-tactile maps obtained with wide-field imaging.

(**A**) Elevation phase map average over mice (N=50 mice). Atlas and corresponding areas labels are overlaid for reference. Pixels with low magnitude modulations are displayed in white. (**B**) Azimuth phase map average over the same mice as in **A**. (**C**) Retinotopic sign map computed based on the elevation and azimuth phase maps of **A** and **B**. (**D**) Examples of visuo-tactile functional maps for 3 mice (rows). From left to right: 1. Whisker preference map used to describe somatotopy in the vertical space with red colors indicating preference for bottom whisker and blue colors indicating preference for top whisker. 2. Visual position preference map used to describe retinotopy in the vertical space. 3. Modality preference maps describing the dominant modality in each pixel. 4. Spatial phase coherence maps between retinotopic and somatotopic maps. 5. Multisensory modulation maps indicating areas with responses enhancement or suppression for visuo-tactile stimuli. (**E**) Average activation time course for each visually responsive area following stimulus onset (N=29 mice). (**F**) Same as **E** for somatosensory responses in cortical areas evoked by whisker stimuli. (**G**) Average response time course in the primary visual and whisker somatosensory cortex highlighting an offset of about 150ms between the two modalities. (**H**) Multisensory modulation computed over time in visuo-tactile areas in for responses that are evoked simultaneously in the visual and somatosensory cortex.

1370

1371

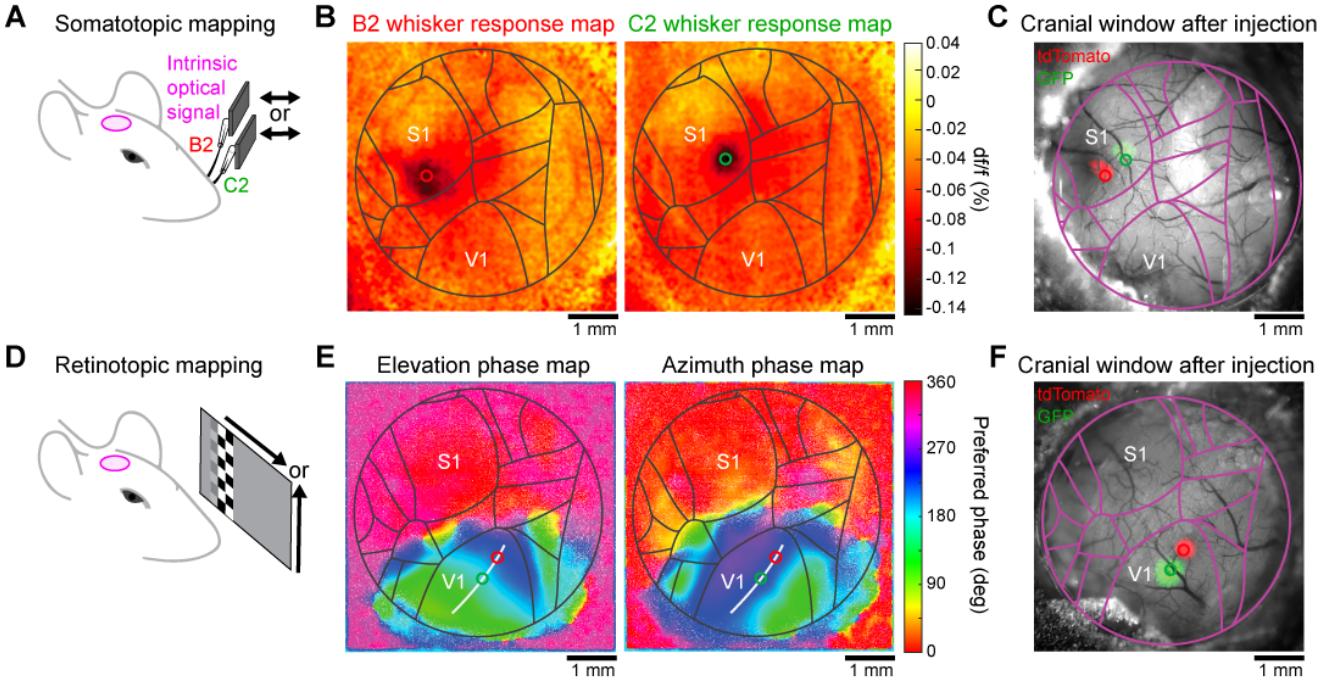

Figure S5

Fig. S5. Mapping two distinct vertical positions in the primary visual or whisker somatosensory cortex.

(**A**) Schematic of the protocol for mapping whisker-responding areas in the dorsal cortex. One of two whiskers is stimulated repeatedly during intrinsic optical signal imaging. (**B**) Change in red light reflectance on the surface of the cortex measured after averaging several trials of whisker stimulations relative to baseline. The left plot reveals the barrel location for B2 whisker stimulations and the right plot for C2 whisker stimulation. (**C**) Image of the cranial window several weeks after injection of viral vectors to express GFP and tdTomato in two distinct parts of the whisker somatotopic map. Locations of B2 and C2 barrels in S1 are indicated with circles overlaid on top of the fluorescent signal where injections were done. (**D**) Schematic of the protocol for retinotopic mapping in areas of the dorsal cortex. Horizontal or vertical bars drift across the screen and they are textured with a flickering checkerboard pattern to elicit strong cortical responses at different phases corresponding to the retinotopic preference. (**E**) Retinotopic phase maps for elevation and azimuth on the dorsal cortex. These maps are then used to identify a gradient across vertical dimension in an iso-profile along the azimuth dimension. Two retinotopically distinct locations along the vertical axis are identified and labeled for viral injection. (**F**) Image of the cranial window several weeks after injection of viral vectors to express GFP and tdTomato in two distinct part of the primary cortex retinotopic map. Locations where stereotaxic injections were performed are indicated with circles overlaid on top of the fluorescent signal where injections were done.

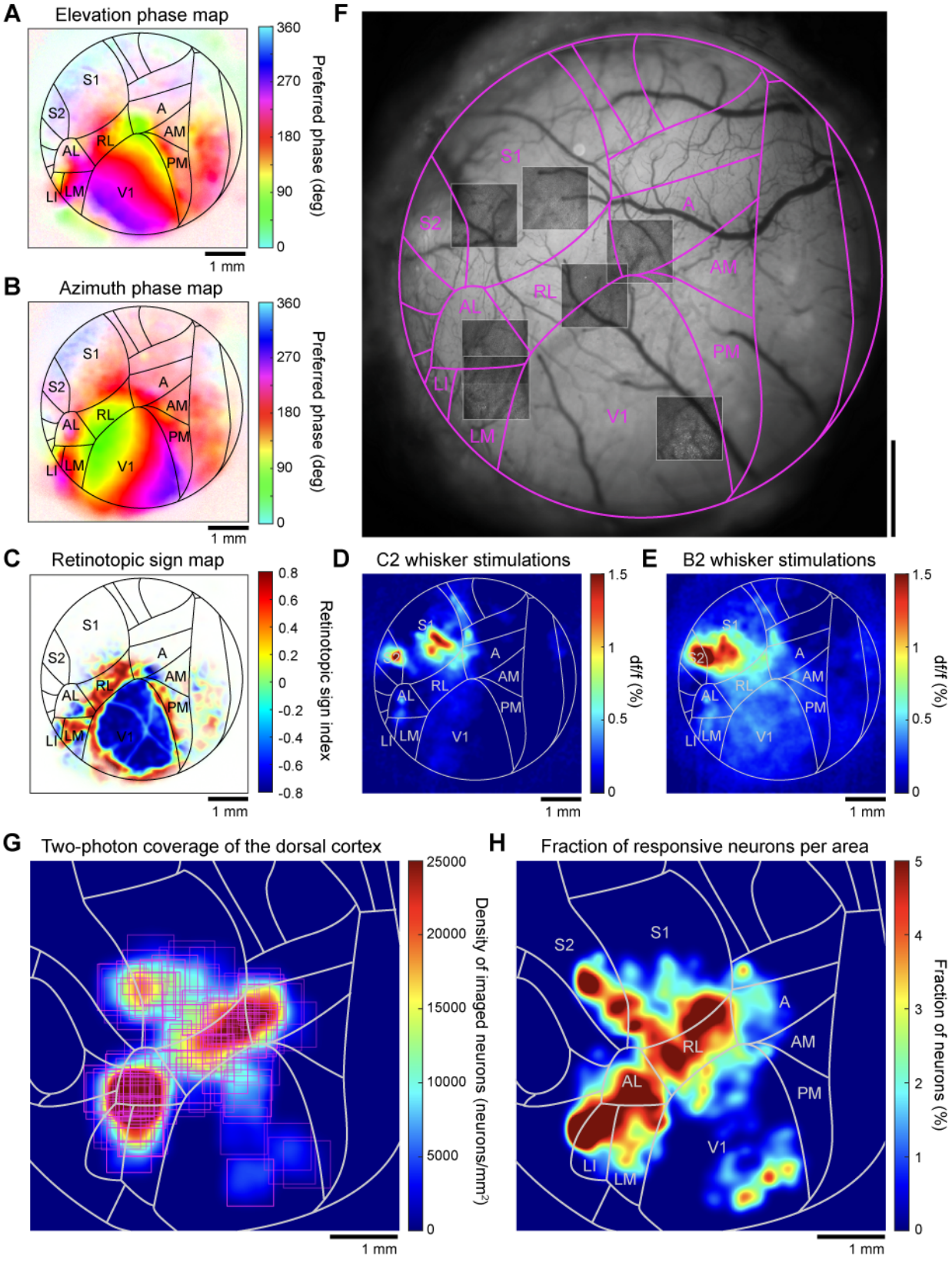

Figure S6

1395

Fig. S6. Coverage of large portions of the cranial window for characterizing visuo-tactile responses at single-cell level using two-photon microscopy.

(**A-E**) Functional maps are first obtained to fit the Allen mouse atlas to precisely localize neurons in their respective areas and to realign with a reference atlas. Visual stimulation protocols are used to obtain phase maps for elevation (**A**) and azimuth (**B**) in order to compute the retinotopic sign map (**C**). Additionally, whisker stimulations provide response maps for C2 (**D**) and B2 (**E**) whiskers. (**F**) Image obtained with wide-field imaging where blood vessels are well distinguished. Atlas fitted with functional maps is overlaid on the top in purple. For this mouse, 7 fields-of-view have been imaged with two-photon imaging and placed on the surface of the cortex using blood vessels pattern. (**G**) Coverage of visuo-tactile areas of the dorsal cortex with two-photon imaging across mice. Using the realigned atlas and information about all neuron's location, we can recreate a large-scale population using single-cell information. We could collect N=134 fields-of-view across N=25 mice. (**H**) Percentage of neurons imaged with two-photon microscopy that were responsive to at least one stimulus condition. The areas of high responsivity are in accordance with the wide-field imaging.

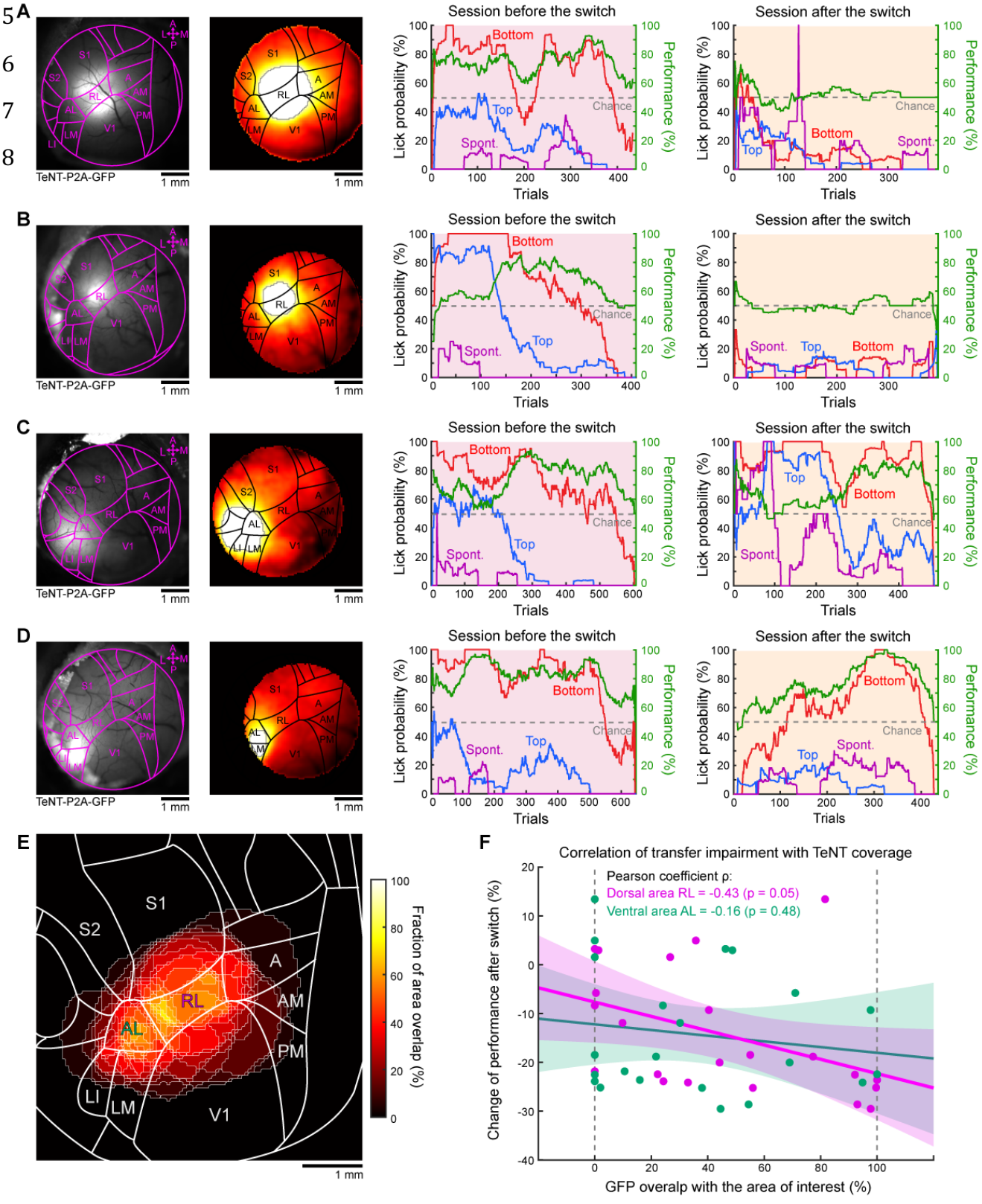

Figure S7

Fig. S7. Transfer learning impairment in relation with TeNT expression in different associative areas of the dorsal cortex.

(**A**) Left: Example of TeNT-P2A-GFP expression pattern over the cranial window and corresponding estimate of expression coverage after blood vessels subtraction. Right: Effect of cortical silencing on transfer learning from tactile task to visual task. Here the discrimination performance dropped at chance level after the switch. (**B-D**) Same as **A** but for other example mice with injections in different locations of the cortex and corresponding effect on transfer learning. (**E**) Overall coverage across all experiments performed with this protocol. (**F**) Example of correlation between the level of GFP overlap with a specific area (here RL or AL) and the corresponding effect on the performance after switch. To obtain a more nuanced effect for the correlation, we computed the performance changes over 3 consecutive days after switch (N=22 mice, Pearson coefficient: -0.43, p=0.05 for RL; Pearson coefficient: -0.16, p=0.48 for AL).