Charting the single cell transcriptional landscape governing visual imprinting.

Vincenzo Lagani[1,2,#], Lela Chitadze[2], Ana Cecilia Gonzalez Alvarez[1], Teimuraz Bokuchava[2], Giulia Sansone[1], Veriko Bokuchava[2], Lia Tsverava[2,3], Aragorn Jones[4], Steven Lisgo[4], Xabier Martinez De Morentin[1], Robert Lehmann[1], Leena Ali Ibrahim[1], Jesper Tegner[1,5,6,7], Brian J. McCabe[8], Zaza Khuchua[2], David Gomez Cabrero[1,9,#], and Revaz Solomonia[2,3,#]

[1]Division of Biomedical Sciences, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia.

[2]School of Natural Sciences and Medicine, Institute of Chemical Biology, Ilia State University, Tbilisi, Georgia

[3]Iv. Beritashvili Centre of Experimental Biomedicine, Tbilisi, Georgia

[4]Biosciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom

[5]Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia.

[6]Department of Medicine, Center for Molecular Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden.

[7]Science for Life Laboratory, Solna, Sweden

[8]Department of Zoology, University of Cambridge, Cambridge, United Kingdom

[9]Translational Bioinformatics Unit, Navarrabiomed, Fundacion Miguel Servet, Universidad Publica de Navarra (UPNA), IdiSNA, Pamplona, Spain

[#]Correspondence: vincenzo.lagani@kaust.edu.sa, david.gomezcabrero@kaust.edu.sa, revaz_solomonia@iliauni.edu.ge

Key words: visual imprinting, learning and memory, Single-cell transcriptomic, IMM cellular composition, long non-coding RNAs, LUC7L, FOXP2, RORA, ROBO1

Abstract

Memory-related transcriptional events in brain remain poorly understood. Visual imprinting is a form of learning in which young animals develop preferences through early exposure to specific stimuli. In chicks, visual imprinting memory is stored in the intermediate medial mesopallium (IMM) of the forebrain. To investigate learning-associated molecular changes, we performed single-nucleus RNA sequencing of the left IMM in strongly imprinted chicks and untrained controls. This analysis generated the first classification of cells composing the IMM, identifying as a result over 30 cell clusters with distinct transcriptional differences putatively linked to memory formation, nearly half of them in long non-coding RNAs (lncRNAs). Follow-up analysis on selected genes, confirmed that the gene expression levels of two lncRNAs and protein levels of FOXP2, RORA (transcription factors), LUC7L (splicing factor), and ROBO1 (axon guidance molecule) correlate with memory strength, reflecting either innate learning potential or imprinting experience. Additionally, among the confirmed lncRNAs, the brain- and avian-specific lncRNA ENSGALG00010007489 is enriched in the nuclei of specific glutamatergic clusters and its association with imprinting was further confirmed through quantitative multi-probe in situ hybridization. These findings offer the first single-cell resolution map of transcriptional changes underlying memory formation in the avian brain.

Introduction

Memory is the capacity to retain learned information and is of fundamental importance to animals. While different types of memory have distinct operational characteristics, they share certain molecular mechanisms, particularly those subserving synaptic plasticity[1]. Study of a given type of memory may therefore elucidate not only the specific molecular details of that particular process, but also extend understanding of generally applicable molecular mechanisms of memory[2]. The specific gene expression programs activated in distinct brain cell subtypes during memory encoding, consolidation, and retrieval remain poorly characterized, particularly at single-cell resolution. Understanding the neural mechanisms of learning and memory is crucial for both scientific advancement and practical applications. In particular, the rational treatment of mental illness, especially the amnesias, requires detailed knowledge of how the brain stores information[3].

A particularly suitable type of learning for memory studies is visual imprinting in chicks. Visual imprinting is a form of learning that occurs early in life, in which the young animal rapidly forms a strong and lasting attachment to a visual stimulus after brief exposure to it[3–7]. Visual imprinting is characterized by highly specific recognition memory, making it a powerful tool for studying the biological basis of memory. In domestic chicks, this form of learning offers key advantages for memory research: (i) imprinting may be studied soon after hatching when chicks do not require food, ensuring standardized nutritional conditions; (ii) the sensory environment during incubation and rearing can be tightly controlled, minimizing extraneous sensory input and allowing for clearer detection of learning-related neural changes; and (iii) newly hatched chicks have a diverse behavioral repertoire, enabling precise behavioral assessments of learning. A commonly used measure of imprinting strength, and thus memory, is the preference score, which quantifies the chick's approach toward the training stimulus relative to a novel, unfamiliar stimulus[3,6,7].

Beyond these operational advantages, a specific brain region has been identified that plays a critical role in the learning and memory processes of chick visual imprinting [3,6,8]. This region is the *intermediate medial mesopallium* (IMM), previously known as the *intermediate and medial hyperstriatum ventrale* (IMHV)[9], which has been identified as a site of memory of the imprinting stimulus. Using regression analysis, it is possible to assess whether molecular changes in the IMM (and other brain regions) are attributable to memory formation consequent on learning, rather than to side effects of the training procedure. Criteria for identifying learning-related changes using linear modeling and variance partitioning have

been documented[10]. Variance that is unexplained by covariation with learning (i.e., residual variance from a regression with preference score) provides information enabling discrimination between a neural change resulting from learning and learning ability existing before training (a predisposition to learn)[10].

Visual imprinting in chicks is associated with hemispheric asymmetry at behavioral, molecular, and electrophysiological levels[3,6,7,11]. Learning-related molecular changes are typically more strongly expressed in the left IMM, especially 24 h after training[3,6,7,10,12–17]. For this reason, the present study focuses on the left IMM.

We have previously demonstrated a progression of learning-related molecular changes in the IMM over the 24 h after training[7]. At least some of these changes are not uniformly distributed among IMM cells and occur at different times after training[18–22]. For example, learning-related up-regulation of Fos protein in parvalbumin-containing, but not calbindin-containing, GABA neurons occurs 1 h after the end of training in the IMM but not subsequently[22]. In contrast, a learning-related change in NMDA receptor binding occurs several hours later[18,19]. To further understand the molecular mechanisms of memory, it is important to know which molecular changes are occurring and in which cell types they take place.

Single cells/nuclei omics technologies have revolutionized molecular biology, allowing investigation of the heterogeneity of living tissues at the cellular level[23]. Publicly available atlases[24,25] of brain cell transcriptomics profiles, together with appropriate technology, enable detailed study of neural plasticity and memory at the cellular level[26,27], increasing understanding of fundamental mechanisms and thus establishing a rational basis for the successful treatment of neurological disorders[28,29]. No previous study has investigated imprinting-associated molecular changes in the chick IMM with single-cell RNA-seq technology.

In the present study, we compared single-nucleus transcriptome profiles of the left IMM between untrained chicks and good learners – those that achieved high preference scores during training (see Methods). We collected tissue 24 h after the end of the training, a time point when numerous long-term, learning-dependent changes occur in the IMM[7,10,12–17,30]. For the first time, our findings have enabled the identification of cell-type-specific gene expression differences - both coding and non-coding - between good learners and untrained chicks.

Further analyses have revealed that for one long non-coding RNA (lncRNA), ENSGALG00010007489 (here identified as GLUBK89) at the transcript level, and three protein coding genes – LUC7L, FOXP2, and RORA at the protein level – display strong correlations between preference score and expression level, indicating changes attributable to learning during training. In contrast, changes in lncRNA ENSGALG00010026609 (lncRNA6609) and ROBO1 likely reflect learning capacity, suggesting a predisposition to learn existing prior to training. These validation experiments were performed in four brain regions (left and right IMM, left and right posterior pole of neostriatum, PPN, a control brain region not involved in imprinting). Moreover, the results indicate that the magnitude of learning-related gene-expression changes within specific cellular clusters is sufficiently large to be detected at the whole-tissue level.

Finally, fluorescence *in situ* hybridization (FISH) experiments confirmed that GLUBK89 is localized in the nuclei of putative glutamatergic neurons in the IMM. Expression of GLUBK89 in these neurons was significantly higher in good learner chicks compared to untrained controls, with poor learners exhibiting intermediate expression levels. These findings indicate that this lncRNA contributes to memory processing within a subset of glutamatergic neurons.

Our results provide insight into memory mechanisms of imprinting and are potentially of general importance in understanding the molecular basis of memory.

Results

IMM cellular composition suggests homology with deep layers of the mammalian association cortex

We followed an established protocol for training and testing[21] (see Methods) to assess imprinting behavior. In short, 24 hours after hatching, chicks were placed in a running wheel and exposed to a training stimulus consisting of a cuboid rotating red light (Figure 1a). Ten minutes after training, a preference test was conducted to measure memory strength. Chicks with preference scores above 80 were considered good learners, having mostly approached the original training stimulus rather than the alternative one (a rotating blue light, Figure 1a). Twenty-four hours after the preference test, the chicks were decapitated, and the left IMM removed. Untrained chicks, which were never exposed to the imprinting stimulus (Figure 1a), served as controls.

We processed left IMM samples from a total of 7 chicks (3 good learners and 4 untrained chicks, Supplementary Figure S1) using single-nucleus RNA-seq technology (see Methods), obtaining transcriptomic profiles for a total of 54,763 nuclei, which were grouped into 36 clusters (Supplementary Table S1 and Supplementary Figure S2).

We contrasted our snRNA-seq data with a recently published single-cell chicken brain scRNA-seq atlas[31], to annotate our clusters on the basis of its classifications (see Methods and Supplementary Table S2). Major cell-types could be identified confidently, as confirmed by well-established marker genes[32–34] (Figure 1b,c): GABAergic neurons (n=8,999; GAD1, GAD2, SLC32A1), glutamatergic neurons (n=37,582; ARPP21, SLC17A6, SV2B, SATB2), astrocytes (n=4,472; AQP4, GLI3), and oligodendrocyte precursors (OPC; n=2,251; PDGFRA and SOX10).

Beyond these major cell types, some clusters were not readily identified by this analysis. Cluster 32 (C32) expresses markers characteristic of vasculature (n=284; INPP5D, FLT1, and ADGRL4) while C35 is characterized by markers of the immune system (n=76; B2M, GPX1, SLC4A1). Cluster C23 (n = 1,099) presents a distinct marker signature, including the LRRIQ1, LRGUK and NEK10 genes, which are associated with “cilium movement involved in cell motility” and related gene ontology biological processes (Supplementary Table S3). We hypothesize that this group corresponds to ependymal cells.

166 With the overall cellular composition of the IMM defined, we next focused on finer-grained
167 classification. Recent works[31,35] have highlighted that GABAergic inhibitory neurons are
168 largely conserved across amniotes. In contrast, no clear unequivocal homology has been
169 established yet between avian and mammalian excitatory neuron subtypes.

170 Consistent with these observations, label transfer identifies known GABAergic subtypes in
171 our data (Supplementary Table S2). LAMP5 inhibitory neurons form a distinct cluster,
172 clearly separate from the other cells (Figure 1d). Other notable inhibitory subtypes include
173 VIP- and MEIS-expressing neurons, as well as striatal-like inhibitory neurons (STR)[36] and
174 somatostatin (SST) cells[37] (Figure 1d). Expression of key markers provides additional
175 evidence for this classification (Supplementary Figures S3 – S14).

176 We next examined the organization of glutamatergic neurons. Our glutamatergic clusters
177 present a high degree of transcriptional similarity with the excitatory subgroups identified in
178 the chicken brain single-cell atlas (see Methods and Supplementary Table S2), and we
179 therefore adopt the corresponding classification. Notably, a substantial fraction of
180 glutamatergic neurons maps to subtypes resembling excitatory neurons of the deep layers of
181 the mammalian cortex. Specifically, 9,269 (24.66%) neurons are classified as GLU-7, whose
182 transcriptomics are most similar to layer 6 (L6) excitatory neurons, while 8,495 (22.60%)
183 cells are classified as GLU-2, corresponding most closely to layer 4 and layer 5 (L4/L5)
184 cortical glutamatergic neurons.

185 Finally, we applied self-assembling manifold mapping[38] (SAMap), an algorithm that exploits
186 similarities between protein sequences to match scRNA-seq profiles across species
187 (Methods), as a confirmatory analysis of our findings. We used a single-cell transcriptomics
188 atlas of the mouse cortex and hippocampus as a reference[32]. This analysis confirmed several
189 results obtained through label transfer from the chicken brain cell atlas (Supplementary Table
190 S2 and S4).

191 In summary, we annotated cells at two different resolutions: at a coarser level, we confidently
192 identified three major cell types (glutamatergic neurons, GABAergic neurons, and non-
193 neuronal cells), each spanning multiple clusters. At a finer level, we identified 21 cell
194 subtypes, as summarized in Supplementary Table S2.

195 Together, these results provide the first detailed classification of the transcriptomic profiles of
196 cells composing the IMM, a brain region strongly implicated in imprinting and passive
197 avoidance memory in chicks[3].

Good learners and untrained chicks exhibit distinct transcriptomic profiles.

We first examined cell-type representation in good learners and untrained chicks, to disentangle transcriptional changes associated with learning from potential differences in cellular composition.

Cells from both experimental groups were detected across all annotated subtypes (Figure 2a). Major cell types were robustly represented across all samples and did not exhibit significant differences in cell proportions between good learners and untrained chicks (Supplementary Figure S15).

In contrast, eight cell subtypes contained fewer than 50 cells in at least one of the seven samples: EXC GLU-6, INH 1, INH 2, INH LAMP5, INH STR, Blood and Immune System, Microglia/Endothelial/Vascular, and INH MEIS2/COL12A1. We therefore refrained from drawing conclusions for these subtypes, either with respect to cell proportion changes or differential gene expression, as differences cannot be reliably interpreted in the presence of low per-sample cell counts[39]. The remaining subtypes showed no significant differences in overall cell abundance between experimental groups (Supplementary Figure S15).

Next, we performed a pseudo-bulk differential expression analysis between conditions (see Methods), both for major cell types and for each subtype. Several differentially expressed transcripts were identified across cell types (log2 fold change > 1 in absolute value, FDR < 0.05, Supplementary File S1). Interestingly, roughly half of these changes were in long non-coding transcripts (lncRNAs, Figure 2b). This represents a statistically significant enrichment relative to expectation ($p < 10^{-3}$), with lncRNAs being at least three times more likely to be differentially expressed than other types of transcripts (Supplementary Table S5).

A similar pattern was observed at the subtype level. All analyzed cell subtypes exhibited transcripts with statistically significant differences between good learner and untrained conditions, ranging from 283 transcripts in EXC GLU-2 down to 6 in EXC IMN and ependymal cells. Consistent with the major cell-type analysis, lncRNAs were significantly overrepresented (FDR < 0.05) among differentially expressed transcripts in 9 subtypes out of the 13 considered for differential analysis (see Supplementary Table S6).

Numerous studies have identified molecular changes in the chick IMM associated with imprinting[7,10,12–22,30]. Our results recapitulate previous findings[40–43] and associate them with specific cell types or clusters: the CHRM2 and CHRM3 genes, encoding muscarinic

receptors, were differentially expressed in several glutamatergic subtypes, while the GABRG3 and GABBR2 genes, encoding GABA-A and GABA-B receptors, exhibited increased expression in good learners within EXC GLU-7, EXC GLU-3, and INH SST (Supplementary File S1).

We further examined whether the observed changes arise jointly across cell types or subtypes. Differential expression results showed a high correlation between log2 fold changes across all major cell types, with excitatory and inhibitory neurons showing the highest level of concordance (Spearman correlation coefficient $\rho = 0.85$, $p < 0.01$), while correlations with non-neuronal cells were slightly lower ($\rho = 0.69$ and 0.72 for glutamatergic and GABAergic cells, respectively, $p < 0.01$ in both cases). High correlation between log2 fold changes persisted at the subtype level (Figure 2c), suggesting that transcriptomic changes occur in a coordinated manner. In particular, glutamatergic and inhibitory subtypes form their own distinct clusters, with the exception of immature excitatory neurons (EXC IMN) which exhibited a partially divergent differential expression pattern (Figure 2c).

We then investigated whether transcriptomic changes are cluster-specific. Figure 2d reports the log2 fold-change for the twenty-five transcripts with the largest expression increases across all glutamatergic cells ($|\log2\ FC| > 1$, $FDR < 0.05$). When examined at the cluster level, they exhibit statistically significant changes ($FDR < 0.05$) only in specific subtypes, indicating that these subtypes are the primary drivers of the overexpression of these transcripts in the broader excitatory neuron population. Furthermore, Supplementary Table S7 lists the 257 transcripts that are differentially expressed in individual glutamatergic or inhibitory subtypes but not in their corresponding cell types, supporting the hypothesis that changes linked to imprinting may occur, or at least be more prevalent, in isolated clusters.

Taken together, our findings (a) highlight an overrepresentation of lncRNAs among differentially expressed transcripts, (b) provide insights into the specific cell types underlying the imprinting-associated molecular alterations reported in prior studies[17,40–43], and (c) demonstrate that, although imprinting-related transcriptomic changes are broadly coordinated across cell types and subtypes, selected gene expression changes are specific to or predominantly driven by individual clusters.

Imprinting is associated with biological processes related to neural plasticity

We next characterized biological processes (BPs) associated with imprinting through functional enrichment analysis across major cell types and neuronal subtypes (Supplementary File S2, Figure 3a). Across all major populations, cell–cell adhesion and cell junction organization were consistently enriched, indicating coordinated structural remodeling of the mesopallium circuitry during learning and corroborating previous findings[17].

Glutamatergic and GABAergic neurons exhibited convergent enrichment for synaptic membrane adhesion, postsynaptic signaling, and axon or neuron projection guidance, together with biological processes directly associated with higher cognitive functions, including GO:0007612 ("learning"), GO:0008306 ("associative memory"), and GO:0007613 ("memory") (Figure 3b). Despite this convergence, the two neuronal classes displayed distinct functional specializations. Glutamatergic neurons were characterized by enrichment for excitatory synaptic transmission and cyclic nucleotide–dependent pathways (Figure 3c), consistent with potentiation and refinement of excitatory synapses. In contrast, GABAergic neurons showed pronounced enrichment for regulation of membrane potential, action potential, and ion transport processes, suggesting modulation of inhibitory neuronal excitability during learning.

Non-neuronal cell types displayed markedly different enrichment profiles. Astrocytes were primarily enriched for developmental and morphogenetic processes, consistent with a supportive and regulatory role in the learning-associated neural environment. Similarly, oligodendrocyte precursors were enriched for developmental patterning, cell fate commitment, and morphogenetic pathways, suggesting modulation of glial developmental programs accompanying imprinting.

We next examined imprinting-associated BPs at the neuronal subtype level. Among glutamatergic subtypes, EXC GLU-7 displayed a distinctive enrichment profile combining biological processes directly linked to learning and adult behavior with sensory-associated programs and regulation of neuronal excitability. The concurrent enrichment for sensory-linked processes, action potential generation, potassium transport, and adenylate cyclase–modulating GPCR/cyclic nucleotide signaling suggests that this deep-layer–like population integrates sensory-driven input with neuro-modulatory gating mechanisms relevant to imprinting-related circuit refinement.

291 In contrast, EXC GLU-2 was characterized by enrichment for dendritic spine development
292 and vesicle fusion alongside prominent cyclic nucleotide signaling, indicating active
293 structural and functional synaptic remodeling. Notably, GLU-2 uniquely showed strong
294 enrichment for miRNA biogenesis and regulatory processes, pointing to an additional post-
295 transcriptional regulatory layer that may contribute to the stabilization and persistence of
296 learning-associated transcriptional states.

297 Across inhibitory neurons, the three analyzed subtypes shared a common backbone of
298 synaptic adhesion and assembly processes but diverged along distinct functional axes. The
299 PVALB population uniquely coupled contact remodeling with prominent adenylate cyclase–
300 coupled GPCR/cyclic nucleotide signaling and regulation of pre- and postsynaptic membrane
301 potential. In contrast, SST interneurons were most strongly defined by coordinated synapse
302 assembly and intrinsic excitability programs, while VIP interneurons combined synaptic
303 remodeling with the strongest axon or neuron projection guidance signatures and a clear
304 neuro-modulatory GPCR/cAMP component.

305 Together, these enrichment analyses indicate that imprinting engages convergent but cell-
306 type- and subtype-specific biological processes related to synaptic remodeling, excitability,
307 and circuit reorganization.

308 lncRNAs associated with imprinting map to cell-type-specific co-
309 expression modules linked to synaptic and developmental processes

310 As significant changes in expression between good learners and untrained chicks appear to
311 preferentially involve lncRNAs rather than protein-coding genes, we aimed to better
312 understand the functions of these differentially expressed lncRNAs. To do so, we clustered
313 all transcripts—including both protein-coding genes and lncRNAs – using a methodology
314 designed for single-cell/nuclei RNA-seq (hdWGCNA[44], see Methods). This analysis
315 generated cell-type-specific transcript modules, encompassing 2,846 transcripts across 17
316 modules for glutamatergic neurons, 921 transcripts in 6 modules for GABAergic neurons,
317 3,476 transcripts in 6 modules for astrocytes, and 50 transcripts in a single module for OPC
318 (see Supplementary Figure S16 and Supplementary File S3).

319 We then performed enrichment analysis on each module, considering only genes included in
320 Gene Ontology BPs (Supplementary File S4). This approach allows assignment of biological
321 functions to each module and, by extension, to the differentially expressed lncRNAs within
322 them[45,46]. Differentially expressed lncRNAs associated with memory formation in

323 glutamatergic chick neurons are distributed across multiple hdWGCNA modules, indicating
324 functional heterogeneity. The largest contribution maps to a synapse-related module enriched
325 for synapse assembly and regulation of postsynaptic membrane potential, suggesting lncRNA
326 involvement in synaptic organization and excitability[47]. Additional lncRNAs associate with
327 modules linked to dendritic morphogenesis, extracellular matrix proteoglycan metabolism,
328 and developmental or regeneration-like neuronal programs, consistent with structural and
329 circuit-level remodeling during learning[48]. A single DE lncRNA maps to a large
330 mitochondrial/translation module, indicating a possible association with neuronal
331 bioenergetic and proteostasis pathways. In astrocytes, three differentially expressed lncRNAs
332 mapped to a single hdWGCNA module enriched for guidance- and adhesion-related
333 processes. This module is characterized by genes involved in autonomic nervous system
334 development and heterophilic cell–cell adhesion, including semaphorin–plexin–neuropilin
335 signaling and synapse-associated adhesion molecules such as neurexins and neuroligins.
336 These enrichments may point to astrocyte-mediated regulation of cell–cell interactions and
337 circuit organization, consistent with an active role of astrocytes in synaptic remodeling during
338 memory formation[49].

339 Together, these findings indicate that both protein-coding genes and lncRNAs differentially
340 expressed in good learners converge on cell-type-specific biological programs, prominently
341 involving the same previously identified pathways on synaptic function, neural development,
342 and circuit remodeling.

343 Selected differential transcripts exhibit specific learning-related 344 changes.

345 Good learners show learning-induced molecular changes in the IMM, but they also differ
346 from untrained chicks in other aspects, such as exposure to visual stimuli and levels of motor
347 activity. Therefore, transcripts with prominent transcriptomic changes were further examined
348 in experiments designed to isolate their specific role in memory, while controlling for
349 potential confounding effects unrelated to learning. The following sections discuss the
350 selection of the transcripts undergoing validation, as well as their respective results.

351 **LUC7L, FOXP2, ROBO1, and RORA in learning and memory.**

352 Long-term memory formation involves specific and sustained transcriptional changes,
353 mediated by transcription factors and other nuclear processes such as alternative splicing[50,51].
354 Based on statistical significance, fold change magnitude, and biological relevance, four

differentially expressed protein-coding genes were selected for further analysis: two transcription factors (FOXP2 and RORA), a splicing-associated protein (LUC7L), and an axon guidance receptor (ROBO1) [52,53] . For all four genes, learning-related quantitative correlations were studied at the level of protein expression (the final product of gene expression).

The results of the immunoblotting experiments show that the levels of all four proteins were significantly ($p < 0.05$) correlated with preference score solely in the left IMM, except for FOXP2, for which a significant correlation was also observed in the left PPN (Figures 4 and 5, Supplementary Tables S8 – S11, Supplementary Figures S17 – S20). LUC7L, FOXP2, and RORA changes cannot be explained as a predisposition; they are rather attributable to learning during training. A similar finding holds for FOXP2 changes in the left PPN. Training had an additional effect on RORA expression, which was significantly depressed at preference score 50, indicating a suppressive effect of training unrelated to learning. ROBO1's association with memory formation appears to be due to a predisposition, as the residual variance from the regression plot is significantly lower than the variance in untrained animals ($P < 0.0005$). The association with preference score is also weak, since there is no significant difference between the mean concentration of the protein in untrained chicks and the corresponding mean value in trained chicks with preference score 100 or 50.

Selected lncRNAs in learning and memory.

Nearly half of the differentially expressed transcripts identified across all cell clusters corresponded to lncRNAs. On this basis, two lncRNAs were selected for further validation.

The first, ENSGALG00010007489, was the most strongly differentially expressed lncRNA in glutamatergic neurons. Notably, this effect was largely driven by a single subtype, EXC GLU-7, whose transcriptional profile resembles that of layer 6 (L6) neurons in the mammalian cortex, suggesting a highly cell-type–specific association. By contrast, the second selected lncRNA, ENSGALG00010026609 (hereafter referred to as lncRNA6609), displayed statistically significant differential expression across multiple major cell types, including excitatory neurons, inhibitory neurons, and oligodendrocyte precursor cells (Supplementary File S1).

We next examined the cellular context of ENSGALG00010007489 expression in greater detail. This lncRNA was predominantly expressed in cluster 2, one of the 5 clusters classified as EXC GLU-7 subtype (Supplementary Table S2), where it was detected in 22% of cells,

387 compared with only 1.5% of cells across other glutamatergic clusters. Strikingly, 91% of
388 cells in cluster 2 also expressed KCNMB2, compared with 21% in other glutamatergic
389 neurons, representing a statistically significant enrichment (FDR < 0.001).

390 KCNMB2 encodes the β2 auxiliary subunit of large-conductance calcium- and voltage-
391 activated potassium (BK) channels, which play key roles in regulating synaptic plasticity,
392 neuronal excitability, and memory[54]. Based on this co-expression, we hereafter refer to
393 ENSGALG00010007489 as GLUBK89.

394 In addition to BK2 channels (KCNMB2) and the GLUBK89 marker, the expression of
395 several other genes distinguishes cluster 2 glutamatergic neurons from other glutamatergic
396 populations (Supplementary File S5). Among the most significantly differentially expressed
397 genes are MDFIC2, the transcription factors EST1 and ATOH8, and the contractile protein
398 ACTN2. These gene expression differences contribute to alterations in multiple biological
399 processes (Supplementary File S5). Notably, the most significantly affected biological
400 process in cluster 2 neurons compared to other glutamatergic cells is regulation of
401 endocytosis. This finding aligns with our previous work and that of others, which reported
402 learning-related upregulation of proteins involved in endocytosis[16,55].

403 We applied the same training protocol as for Western immunoblotting studies and
404 standardized lncRNA (GLUBK89 or LNCRNA6609) expression levels were measured across
405 all samples using qRT-PCR. Preference scores were recorded for all trained chicks (see
406 Methods for details).

407 The results, presented in Figure 6A and Supplementary Table S12, show a significant
408 correlation between GLUBK89 expression and preference score (Pearson's $r = 0.64$, $p =$
409 0.03, Fig. 6A). The intercept at maximum preference is significantly higher than the mean
410 expression in untrained chicks ($p = 0.019$). In contrast, the intercept at 50% preference (no
411 learning) does not differ significantly from the untrained mean. Additionally, the residual
412 variance from the regression is not significantly different from the variance in untrained
413 chicks.

414 These findings indicate that the observed positive correlation between GLUBK89 expression
415 and preference score directly results from learning during training (see Methods). No
416 significant correlations were found in other brain regions, suggesting a localized effect of
417 learning in the left IMM (Supplementary Table S12 and Supplementary Figure S21).

418 The correlation between the standardized amount of LNCRNA6609 and the preference score
419 is significant (r = 0.63, p = 0.037; Figure 6B, Supplementary Table S13). However, the
420 differences between the mean untrained value and the intercepts for preference scores 50 and
421 100 were not significantly different. Together with the residual variance from the regression
422 line being significantly lower than the variance of untrained chicks (p = 0.003), these results
423 suggest that higher LNCRNA6609 expression may be associated with an inherent
424 predisposition for learning rather than an effect of imprinting training (see Methods
425 subchapter "Interpretation of a correlation between a molecular quantity and preference
426 score"). No significant learning-related changes were evident in other brain regions
427 (Supplementary Table S13 and Supplementary Figure S22).

428 **GLUBK89 is a brain and nucleus specific transcript.**

429 In addition to their nuclear localization, lncRNAs can also reside in other cellular
430 compartments, where they perform distinct functions. To determine the subcellular
431 localization of GLUBK89 and LNCRNA6609, IMM tissue was fractionated into cytoplasmic
432 and nuclear components, followed by RNA extraction and qRT-PCR analysis. GLUBK89
433 levels were found to be over 100-fold higher in the nuclear fraction, indicating predominant
434 nuclear localization. LNCRNA6609 is present both in the nuclear and cytoplasmic fractions
435 (Supplementary Figure S23).

436 Finally, the expression of both lncRNAs was analyzed in chick heart, muscle, liver, lung, and
437 brain using quantitative RT-PCR. GLUBK89 expression was detected only in brain. In
438 contrast, LNCRNA6609 was detected across all studied tissues, with the highest expression
439 levels observed in the lung and liver (Supplementary Figure S24).

440 **GLUBK89 is an avian-specific transcript, while LNCRNA6609 shows broad sequence**
441 **conservation in mammals.**

442 To further characterize these lncRNAs, we examined their evolutionary conservation by
443 identifying potential orthologs of GLUBK89 and lncRNA6609 in mammals.

444 Synteny analysis of protein-coding genes flanking the GLUBK89 locus on chicken
445 chromosome 1 revealed a high degree of homology between chicken and turkey; however,
446 chromosomal and regional conservation with the golden eagle, and especially with humans,
447 was markedly reduced.

448 A similar analysis of the lncRNA6609 locus on chromosome 7 demonstrated strong
449 conservation of gene content among chicken, turkey, and golden eagle, whereas conservation
450 with the human genome was substantially lower (Supplementary Figure S25 and
451 Supplementary File S6).

452 Further searches in the RNAcentral database revealed that GLUBK89 is conserved at the
453 sequence level among avian species but lacks identifiable orthologs in mammals.

454 In contrast, lncRNA6609 shows sequence similarity with 89 regions in the human genome
455 and 30 in the mouse genome. It shares more than 70% sequence identity with several human
456 lncRNAs of unknown function, including HSALNT0096019 and SH3TC2-DT. Similarly, the
457 mouse genome contains regions sharing over 65% identity with NONMMUT029999.2,
458 NONMMUT009705.2, and NONMMUT032357.2, none of which encode proteins or have
459 well-defined functions.

460 FISH analysis confirms GLUBK89 localization in glutamatergic 461 neurons and its association with memory.

462 We further validated the cellular localization of GLUBK89 using triple fluorescence *in situ*
463 hybridization (FISH). Specifically, we selected a good learner (preference score > 85%), a
464 poor learner (preference score < 65%), and an untrained chick from each of four different
465 batches, totaling 12 chicks. Six sections were collected from each brain. Five sections were
466 triple-stained for GLUBK89, PPIB (a housekeeping gene), and SLC17A6 (a glutamatergic
467 neuron marker). In the remaining section, PPIB was replaced with GAD2, a GABAergic
468 neuron marker (see Supplementary Figure S26 for the experimental design). Regions
469 corresponding to the left and right IMM and nidopallium were manually identified (see
470 Methods for details). RNAscope images were processed using QuPath[56] to obtain average
471 fluorescence values for each channel within each detected nucleus (see Methods for
472 additional details).

473 Figure 7 provides confocal FISH images of left IMM cells from a trained chick. Notably, the
474 GLUBK89 fluorescent signal closely overlaps the glutamatergic marker SLC17A6 (panels A
475 and C), while the same lncRNA is virtually absent from the GABAergic neurons marked with
476 GAD2 (panels B and C).

To assess whether GLUBK89 expression in glutamatergic neurons is associated with memory of visual imprinting, we quantitatively analyzed sections stained with SLC17A6, GAD2, and GLUBK89 (Supplementary Figure S27) in different experimental groups of chicks. Nuclei of putative glutamatergic and GABAergic cells were selected as having above-threshold SLC17A6 (>1700) and GAD2 (>650) fluorescence levels, respectively. GLUBK89 fluorescence levels were consistently and significantly (two-tailed t-test $p < 0.001$) higher in glutamatergic than in GABAergic nuclei (Supplementary Figure S27).

In the IMM, there was a significant effect of experimental group ($F_{2,6} = 6.42$, $P = 0.032$); no significant effect of side and no significant interaction between experimental group and side were observed ($P > 0.25$). The mean value for good learners was significantly greater than that of untrained chicks ($P = 0.031$, adjusted by Tukey's method for a family of three means). No other pairs of means differed significantly. The rank order of mean normalized values for lncRNA in SLC-positive nuclei was good learners > poor learners > untrained chicks (5.49, 4.76 and 2.68, respectively). This rank order is consistent with the result shown in Figure 6A, which was obtained with a larger sample and thus greater statistical precision.

The same analysis of variance model applied to data from the nidopallium showed no significant effects (Supplementary Figure S28).

Thus, the triple FISH experiments confirm that: (i) GLUBK89 is localized exclusively within the nuclei of glutamatergic neurons; (ii) its expression is non-uniform, and the low-frequency distribution indicates the presence of a distinct subpopulation of GLUBK89-expressing glutamatergic neurons; and (iii) elevated GLUBK89 expression within this subcluster is associated with learning. These observations are fully consistent with our snRNA-seq data.

500 # Discussion

501 The present study, for the first time, (i) characterizes the cellular and transcriptomic
502 organization of the avian intermediate medial mesopallium (IMM), a brain region critical for
503 learning and memory, at single-nucleus resolution, and reveals transcriptomic homologies
504 with deep layers of the mammalian cortex; (ii) provides evidence for extensive involvement
505 of long non-coding RNAs (lncRNAs) in long-term memory formation; (iii) identifies specific
506 cell types and subcellular compartments underlying learning-related differential gene
507 expression during imprinting; (iv) defines the biological processes associated with these
508 differentially expressed genes; and (v) uncovers molecular signatures of visual imprinting in
509 chicks at single-cell resolution. These signatures encompass both imprinting-specific
510 mechanisms and molecular pathways shared with other forms of memory.

511 **IMM homology with mammalian brain structures and cell-type identification.**

512 The IMM is a key brain region involved in memory encoding during visual imprinting in
513 chicks and is also essential for passive avoidance learning[57]. Visual imprinting involves
514 recognition of a sensory stimulus, and analysis of its neural basis in chicks offers a powerful
515 framework for studying recognition memory across species. Moreover, alterations in
516 imprinting-related behavior in chicks have been proposed as a model for aspects of autism
517 spectrum disorders[58] and for long-term memory studies in general[3,6,7].

518 Despite the widespread occurrence of imprinting-like processes in mammals and the
519 fundamental importance of recognition memory, the correspondence between the avian IMM
520 and mammalian brain regions has remained unclear. Early studies proposed that the IMM is
521 homologous to the mammalian association cortex[8]. Subsequent work suggested similarities to
522 neocortical layers 2 and 3 based on gene expression patterns[59,60], while later analyses of
523 transcription factor expression indicated that mesopallial excitatory neurons may resemble
524 mammalian neocortical neurons located within layers 2, 3, 5, and 6[61]. In agreement with
525 these earlier results, a recently published scRNA-seq atlas of the chicken brain draws
526 parallels between mesopallium glutamatergic neurons and mammalian deep-layer cortical
527 excitatory neurons using scRNA-seq, scATAC-seq, and regulatory data[31]. When the same
528 atlas was used for classifying cells in our study, we found that 47.26% of excitatory neurons
529 were classified with high confidence either as EXC GLU-7 or GLU-2, whose transcriptomic
530 profiles are most similar to layer 6 (L6) and layers 4, 5 (L4/L5) of the mouse cortex,
531 respectively.

532 Finally, a previous study reported that a substantial proportion of glutamatergic (VGLUT2+)
533 neurons in the chick mesopallium co-express SATB2 and BCL11A, markers whose co-
534 expression is characteristic of intra-telencephalic (IT) neurons in mammals[61]. Consistent with
535 this, we identified co-expression of SATB2 and BCL11A in 20,714 IMM excitatory neurons
536 (37.8%), constituting the majority of cells in several glutamatergic clusters (Supplementary
537 Table S14; Supplementary Figure S29).

538 These findings further support the hypothesis that a substantial subset of IMM glutamatergic
539 neurons shares molecular features with mammalian cortical excitatory neuron classes,
540 including intra-telencephalic (IT) neurons.

541 **The IMM and learning-related molecular correlations**

542 Visual imprinting is accompanied by behavioral and physiological side-effects, such as
543 increased arousal and motor activity, which are not essential for learning or memory but may
544 nonetheless be linked to biochemical changes. Such effects can confound the identification of
545 molecular correlates that are specifically attributable to learning. An analytical framework
546 has been developed to address this issue and to distinguish molecular changes arising directly
547 from learning from those reflecting a pre-existing capacity or predisposition to learn[10].

548 Lesion experiments[11] have shown that, 26 h after training, a brain region outside the IMM
549 (termed S′) can support retention of a preference for the training stimulus. Lesioning the right
550 IMM approximately 3 h after training prevents the formation of S′, although the precise
551 anatomical location of this region remains uncertain[3,6]. Importantly, despite disruption of S′
552 formation, chicks retain the acquired imprinting memory, indicating that the left IMM is
553 required for memory retention at 24 h[62].

554 Consistent with this interpretation, neurons in the intact IMM increase their selective
555 responses to the familiar training stimulus (e.g., a red box or a rotating blue cylinder) over the
556 24 h period after training [63–66] A novel stimulus evoked only a baseline response in IMM
557 neurons throughout the experiment. If chicks were allowed to sleep undisturbed 5-11 h after
558 the start of training, at ~20 h the chicks were strongly imprinted and neuronal responsiveness
559 to the familiar imprinting stimulus was more than twice the pre-training level. If sleep was
560 disturbed during the same critical time window, preference score showed no evidence of
561 imprinting and IMM neuronal responsiveness collapsed to the pre-training level[65].
562 Manipulation of sleep over the day following imprinting training thus reveals a close
563 association between learned responsiveness in IMM neurons and preference score. Learning-

564 related increases in neural cell adhesion molecules (NCAMs) and clathrin heavy chain are
565 observed at 24 h but not at 9.5 h after training[17,55]. Together with numerous other molecular
566 changes reported at this time point in the left IMM[7], these findings support the existence of a
567 time-dependent cascade of biochemical events culminating in stable recognition memory.
568 The functional roles of these lasting quantitative associations with memory, and recognition
569 of the imprinting stimulus by IMM neurons, remain to be elucidated.

570 **Transcriptomic fingerprints of learning in the IMM.**

571 Single-nucleus transcriptomic analysis of the IMM in good learner and untrained chicks
572 identified 319 differentially expressed genes in glutamatergic neurons of the left IMM (|log2
573 FC| > 1, FDR < 0.05). Remarkably, 180 of these transcripts were lncRNAs, highlighting their
574 potential functional relevance in imprinting-related memory formation.

575 Several biological processes impacted by imprinting in the present study confirm results from
576 our previous work. Cell adhesion processes, pathways of chemical synaptic transmission,
577 dendritic spine organization, regulation of miRNA transcription, among others, are
578 upregulated in the left IMM of good learner chicks. Specifically, our previous data indicate:
579 (i) upregulation of NCAMs[17]; (ii) increased glutamate release[67]; (iii) elevated synaptic vesicle
580 turnover (clathrin heavy chains, dynamin) [16,55]; (iv) increased membrane bound MARCKS
581 protein[68] and (v) changes in mi-RNA spectrum[10].

582 Although once considered transcriptional noise, lncRNAs are now recognized as key
583 regulators of chromatin remodeling, transcriptional control, and post-transcriptional
584 regulation [69–71]. They function as molecular guides, scaffolds, and decoys, modulating
585 chromatin architecture and gene expression through interactions with chromatin-modifying
586 proteins and RNA-binding complexes, thereby contributing to cellular identity and genome
587 organization.

588 Based on neuronal relevance and involvement in synaptic processes, we selected four
589 protein-coding genes, namely LUC7L, FOXP2, ROBO1, and RORA, and two lncRNAs,
590 GLUBK89 and lncRNA6609, for downstream validation.

591 Behavioral analyses combined with molecular measurements and established criteria for
592 learning-relatedness revealed that GLUBK89, LUC7L, FOXP2, and RORA are upregulated
593 as a consequence of learning at the protein level in the left IMM. In contrast, lncRNA6609
594 and ROBO1 were associated with a predisposition or capacity to learn rather than training-
595 induced changes.

596 Detailed analysis of GLUBK89 using quantitative multi-probe *in-situ* hybridization
597 (RNAscope), supported by RNA-seq data, demonstrated that its learning-induced
598 upregulation is restricted to the nuclei of glutamatergic neurons. Although primarily
599 expressed in a single cluster, GLUBK89 ranked highest in statistical significance in pseudo-
600 bulk analyses across all glutamatergic neurons (Supplementary File S1). qRT-PCR analyses
601 confirmed a learning-driven correlation exclusively in the left IMM, with no corresponding
602 positive correlation in the right IMM or in either hemisphere of the PPN. GLUBK89 is
603 further distinguished by its brain-specific, nuclear localization and apparent restriction to
604 avian species. FISH analysis confirmed its localization to a small subset of glutamatergic
605 neurons, supporting its role as an imprinting-specific regulator.

606 Given its co-expression with BK channels, known modulators of synaptic plasticity[72,73], and
607 its enrichment in a cluster characterized by KCNMB2 expression and endocytosis-related
608 genes, we propose that GLUBK89 participates in coordinated gene regulation underlying
609 learning and memory in this specific glutamatergic neuron population.

610 lncRNA6609, in contrast, is more broadly overexpressed across neurons and oligodendrocyte
611 precursors. Unlike GLUBK89, lncRNA6609 is also localized in the cytoplasm and expressed
612 in tissues beyond the brain. Its levels correlate with preference score exclusively in the left
613 IMM, reflecting an inherent learning capacity rather than a training-induced effect, as
614 inferred from variance partitioning criteria. Its cytoplasmic localization raises the possibility
615 that lncRNA6609 functions through translation or as a scaffold for protein complexes, similar
616 to previously described lncRNAs associated with four glycolytic payoff phase enzymes[74].

617 LUC7-like proteins are components of U1 small nuclear ribonucleoproteins and regulate
618 alternative splicing. Of the three vertebrate paralogs, LUC7L, LUC7L2, and LUC7L3[75,76],
619 only LUC7L showed learning-related differential expression. LUC7L expression was
620 significantly higher in good learner chicks compared to the untrained group in almost all
621 glutamatergic subtypes (EXC GLU-1, GLU-2, GLU-4, GLU-5, GLU-7), with a borderline
622 significant increase observed in oligodendrocyte precursors. No significant changes were
623 found for LUC7L2 or LUC7L3. Furthermore, we demonstrated that LUC7L protein levels
624 positively correlate with learning strength specifically in the left IMM, while no such
625 correlation was observed in other brain regions. This relationship is driven by the learning
626 process itself rather than an inherent predisposition to learn.

627 LUC7L, LUC7L2, and LUC7L3 differentially regulate two major 5′ splice sites (5′SS)
628 classes. The learning-related upregulation of specific splicing factor LUC7L suggests
629 increased splice events during memory formation. This represents, to our knowledge, the
630 first demonstration of memory-related regulation of LUC7L, consistent with emerging
631 evidence implicating alternative splicing in synapse development and memory formation [77].

632 Retinoid-related orphan receptor-alpha (RORA) was upregulated in astrocytes and
633 glutamatergic neurons of good learner chicks, with protein levels correlating positively with
634 learning strength in the left IMM. This correlation was driven by the learning process rather
635 than an inherent learning capacity.

636 In addition to this correlation, RORA expression at a preference score of 50 is significantly
637 lower than in untrained chicks (Figure 4, Supplementary Table S10), indicating that the
638 training procedure suppresses RORA expression in the absence of learning. This reduction
639 could be attributed to various side-effects of training, such as locomotor activity, visual
640 stimulation, or stress. In contrast, chicks that successfully learned during training exhibited
641 increased RORA expression. Notably, at the highest preference scores (indicative of strong
642 learning), the training-induced suppression of RORA was fully reversed. Given RORA's
643 versatility, it is perhaps unsurprising that its expression responds in opposite directions under
644 different behavioral conditions, as illustrated in Figure 4.

645 RORA is a nuclear receptor family member that functions as a constitutive transcriptional
646 activator[78] and is predicted to regulate over 2,700 genes[79]. These include genes associated
647 with autism spectrum disorders, neuronal differentiation, adhesion, synaptogenesis, synaptic
648 transmission, cognition, memory, and spatial learning. Recent studies utilizing single-nucleus
649 and spatial transcriptomics have identified neurons resilient to neurodegeneration in the
650 human brain during Alzheimer's disease progression. Some of these resilient neurons have
651 been reported to be located in layer 4 of the association cortex and, early in the disease,
652 exhibit upregulation of genes involved in synapse maintenance, synaptic plasticity, calcium
653 homeostasis, and neuroprotection, including RORA and NCAM2[80]. Notably, the EXC GLU-
654 2 subtype shares similarities with cortical layers 4/5 in mammals, and in our data these cells
655 show upregulation associated with learning of both RORA (FDR = 0.007) and NCAM2
656 (nominal p-value = 0.012). To the best of our knowledge, this is the first demonstration of
657 learning-related upregulation of RORA in a chick brain region known to be involved in
658 memory.

659 Forkhead box protein P2 (FOXP2), a transcriptional repressor implicated in speech and
660 cognitive functions[81], was upregulated in GABAergic neurons, astrocytes, and
661 oligodendrocytes of good learner chicks. FOXP2 may directly bind up to 400 promoters and
662 regulate the transcription of at least 1800 genes in different tissues [82]. Our snRNA-seq data
663 reports the significant upregulation of FOXP2 in oligodendrocyte precursors (FDR = 0.02),
664 as well as in astrocytes (FDR = 0.056). Further immunoblotting experiments revealed
665 significant positive correlations between FOXP2 and the strength of learning and memory in
666 the left IMM and left PPN. Both changes fulfill criteria of learning-relatedness[10], and in both
667 cases the up-regulation is attributable to training and rather than a predisposition to learn
668 well. We speculate that learning-related changes in myelination may extend to the left PPN
669 region. In all cases, the data obtained indicate long-term, memory-related changes of FOXP2-
670 regulated biological processes.

671 This is the first time a change satisfying our criteria for learning-relatedness [10] (see Methods)
672 has been detected in the PPN. The probability of all these criteria being satisfied by chance is
673 low, raising the possibility that consequences of learning radiate from the IMM throughout
674 the telencephalon in the hours following imprinting training. Indeed, regions outside the
675 IMM can support memory for an imprinting stimulus under the control of the right IMM, in
676 an area termed S'[8,11], and several candidates for the location of S' have been proposed [83–85].
677 Whereas the critical role played by the IMM in the learning and memory of visual imprinting
678 and passive avoidance memory is well established [6,8], there is no evidence that the PPN has
679 any specific role in imprinting, as opposed to passively reflecting changes originating in the
680 IMM. ROBO1 (roundabout guidance receptor 1) binds SLIT2 guidance molecules and is
681 involved in axon outgrowth and orientation[86,87]. Importantly, in the avian brain differential
682 regulation of SLIT–ROBO components has been linked to specialized forebrain circuits
683 underlying learned vocal behavior, supporting a broader role for this pathway in experience-
684 dependent circuit specialization[88,89]. Beyond development, ROBO1 expression has been
685 detected in astrocytes and shown to increase in remodeling contexts such as forebrain
686 ischemia, suggesting potential involvement in glia-mediated structural plasticity[90]. In our
687 scRNA-seq data, the expression of ROBO1 is increased in good learner chicks in non-neural
688 and GABAergic cells (FDR=0.066 and 0.087, respectively). ROBO1 protein amount strongly
689 correlates with the strength of learning and memory only in the left IMM. Importantly, this
690 correlation reflects a predisposition or capacity to learn rather than a result of training.

691 Our snRNA-seq and DEG analysis revealed a set of candidate transcripts, potentially
692 involved in memory formation process. Of particular interest is the implication of lncRNAs.
693 Some of the candidates have been analyzed in further detail, controlling for confounding
694 influences by comparing untrained chicks with trained chicks that display behavioral
695 evidence of particular levels of memory. This behavioral specificity was extended to the
696 cellular level with GLUBK89, locating the learning-related changes in GLUBK89 in the
697 nuclei of glutamatergic neurons in the IMM. Cellular analysis in conjunction with behavioral
698 measurements is thus a powerful means of elucidating neural mechanisms of memory.

699 **Visual imprinting involves both specific and general molecular mechanisms of memory.**
700 Our findings demonstrate that avian imprinting engages both specialized, species-specific
701 molecular pathways and conserved mechanisms shared across vertebrates. GLUBK89
702 represents an avian- and brain-specific lncRNA with a highly restricted cellular and
703 subcellular expression pattern, suggesting a dedicated role in imprinting. In contrast,
704 lncRNA6609 shows broad expression, mammalian homology, and correlation with learning
705 capacity, indicating involvement in more general memory processes. This pattern suggests its
706 potential involvement in broader forms of memory beyond imprinting and across species.
707 These findings illuminate imprinting-specific molecular pathways while also contributing to
708 our understanding of the general molecular architecture underlying memory.

709 **Limitations and further directions.**

710 This study lacks mechanistic insight into the specific roles of lncRNAs in memory formation.
711 The poor sequence conservation of lncRNAs across species hampers the identification of
712 mammalian orthologs and limits functional predictions. Moreover, current computational
713 models for inferring lncRNA function are constrained by the scarcity of avian-specific data[91].
714 To elucidate the biological processes and molecular mechanisms involving these lncRNAs,
715 genetically modified chicks with targeted gain- or loss-of-function perturbations would be
716 highly informative. However, generating transgenic chicken models currently remains
717 technically challenging. As an alternative, adenoviral-mediated delivery of lncRNA
718 constructs could provide a viable approach. Complementary strategies, such as CRISPR-
719 based genome editing or optogenetic manipulation of specific neuronal subpopulations may
720 also help dissect the functional contributions of these lncRNAs to memory processes.

721

Materials and Methods

Chick training and apparatus

Fertile eggs (Cobb 500) were obtained from Sabudara poultry, Tbilisi, Georgia. In total 49 batches of eggs were incubated and hatched in darkness. Chicks were reared in isolation in darkness and at 22–28 h post-hatch trained for 1 h. During training each chick was exposed in a running wheel (1 revolution = 94 cm) to a training stimulus, a cuboidal red box rotating about a vertical axis. The box contained a light surrounded by a red filter (Lee Filters 106 Primary Red); the larger two sides of the box (18 × 18 cm) were translucent and vertical and the remaining sides (18 × 9 cm) were black. During training, the stimulus was rotating and illuminated for 50 seconds and then switched off for 10 seconds each minute. The maternal call (70 –75 dB) of a hen was played while the stimulus was on. This procedure strengthens imprinting to a visual stimulus[92]. As a chick attempted to approach the training stimulus, it rotated the running wheel and revolutions of the wheel were counted to provide a measure of approach activity ("training approach"). A preference test was performed without the maternal call 10 min after the 1 h training period ended. During the preference test, each chick in a running wheel was shown sequentially the training stimulus and an alternative stimulus that the chick had not previously seen, in the following order: training/alternative/alternative/training. Each exposure period during the test lasted 4 min and successive periods were separated by 4 min in darkness making 8 min exposure to each stimulus. The alternative stimulus was a right circular cylinder (height 18 cm and diameter 15 cm) with a translucent wall and vertical axis, rotating about this axis at 28 revolutions per minute. The cylinder contained a light surrounded by a blue filter (Lee filters HT 118 Brilliant Blue). See Horn (1989) for illustrations of the training and alternative stimuli[93]. A preference score was used to measure the strength of imprinting (i.e., learning), computed as (approach to training stimulus during test × 100/total approach during test). A preference score of ∼50 indicates poor learning whereas a score of ∼100 indicates strong learning.

Chicks from: (i) 3 batches were used for single-nuclei RNA experiments; (ii)10 batches for LUC7L measurements; (iii) 11 batches for FOXP2 studies; (iv) 10 batches for lncRNA ENSGAL000100007489 experiments; (v)10 batches for lncRNA ENSGALG00010026609; (vi) 10 batches for RORA measurements; (vii) 10 batches for ROBO1 and (viii) 4 batches for double in situ staining studies. Eight batches of chicks were shared for Luc7L and FOXP2 experiments, 5 batches were shared between the lncRNA experiments, 8 batches were shared

754 between ROBO1 and RORa experiments. Thus, 49 batches of chicks were used for the
755 reported data. For the single nuclei RNA-SEQ experiments, each batch consisted of one good
756 learner and one untrained chick, except for one batch in which two untrained chicks were
757 included. For all other experiments in each batch, there were up to three trained chicks and a
758 control chick from the same hatch.

759 Chicks were decapitated 24 h after the end of training. For single-nuclei RNA experiments
760 only left IMM was sampled, whereas for protein and RNA measurements IMM and PPN
761 from the left and right hemispheres were collected. The locations of the IMM and PPN are
762 described in Solomonia et al. (2013)[21]. Details of the IMM removal procedure are provided in
763 Davies et al. (1985)[94] and details of PPN removal in Solomonia et al. (1998)[17]. For double in-
764 situ experiments whole forebrains were removed (see below). Samples were coded after
765 collection and all further procedures were performed blind.

766 Preparation of nuclear suspension

767 The nuclear fraction for RNA-SEQ was performed by a combination of methods described by
768 Bakken et al., 2018 and Zhou et al., 2020. Frozen IMM from the left hemisphere was
769 homogenized in a Nuclei PURE Lysis Buffer, (NUC-201, Sigma, formulated for solid
770 tissues) supplemented with DTT, Triton X-100, protease inhibitors and 0.5% RNasin Plus
771 RNase inhibitor (Promega). The procedure was carried out in a 1 ml Dounce homogenizer
772 (Wheaton) using 10 strokes of the loose Dounce pestle followed by 10 strokes of the tight
773 pestle to liberate nuclei. Homogenate was strained through a 30 µm cell strainer (Miltenyi
774 Biotech) and centrifuged at 900 x g for 10 minutes to pellet nuclei. The supernatant was
775 carefully removed. The pellet was resuspended in homogenisation buffer and strained
776 through a 20 µm cell strainer. The filtered suspension was centrifuged at 900 x g for 10
777 minutes to pellet nuclei. Further separation of nuclei from myelin debris was carried our as
778 described by Zhou et al, (2020). Nuclei pellets were resuspended in 500 uL Nuclei Wash
779 buffer (1% BSA in PBS with 0.2 U/uL RNasin (Promega) and 900 uL 1.8 M Sucrose. This
780 combined 1400 uL mixture was carefully layered on 500 uL 1.8 M sucrose and centrifuged at
781 13,000 x g for 45 min at 4 ^{o}C. The nuclear pellet obtained was strained through a 30 µm cell
782 strainer, resuspended in Foetal bovine serum + DMSO (90% +10%) and stored for further
783 experiments at -70 ^{0}C.

Single-nuclei RNA-seq library preparation and sequencing.

Single cell transcriptome profiling was performed at SciLifeLab, Stockholm, Sweden. Frozen nuclei suspensions were thawed, washed in a small volume of PBS/BSA/RNase inhibitor at 500 g centrifugation and strained again. Libraries were prepared with a Chromium Controller and the Chromium Single Cell 3p RNA library v3 chemistry (10X Genomics, USA), according to the manufacturer's instructions. The quality of the libraries was verified on a 2100 Bioanalyzer (Agilent Technologies, USA). Paired-end sequencing (read 1–28 bp, read 2–91 bp) was run on a NovaSeq 6000 (S2 flow cell) sequencer (Illumina, USA).

Single-nuclei RNA-seq preprocessing

The 10x application CellRanger[95] (version 7.1.0) was used for alignment and quantification of the raw sequencing files into count matrices. The option "--include-introns" was set to "true", as recommended for single nuclei experiment, while the ENSEMBL GRCg7b.108 genome was used as reference. The resulting count matrices were then processed using several tools. DIEM (Debris Identification using Expectation Maximization)[96] was used in order to identify and eliminate debris, meaning empty droplets most likely to contain ambient mRNA and no cell. DoubletFinder[97] was then applied for removing droplets possibly containing more than one single cell. Finally, cells marked as debris by CellBender[98] were also eliminated. The standard Seurat[99] workflow was then used for removing cell with more than 5% of mitochondrial mRNA, as well as for normalizing and scaling the data. During scaling, we further normalize the data by regressing out of the expression value the amount of mitochondrial mRNA, the total number of reads and measured features. Integration across samples was performed using an "anchor" based approach[99], using the 2000 most variable genes. The Leiden algorithm[100] was finally employed on the integrated data for clustering the cells according to their transcriptomics profiles (the resolution value was set to 0.8).

Cell type identification

The expression of known cellular markers was used in order to assign clusters to major cell types: ARPP21, SLC17A6, SV2B, SATB2 (glutamatergic neurons), SLC6A1, GAD1, GAD2, SLC32A1 (GABAergic neurons), OLIG2, SOX10 (oligodendrocytes), AQP4, GLI3, SLC30A12, and LRRIQ1 (astrocytes and ependymal). The Seurat R package was then sued for refining the cell type classification through label transfer[99] from a reference atlas[31]. Seurat's label transfer is based on the concept of "anchors", which are pairs of analogous cells between a well-annotated reference dataset and a query dataset; these anchors are found

816 by identifying mutual nearest neighbors in a shared low-dimensional space, typically derived
817 from PCA. Once anchors are selected, Seurat projects metadata (e.g., cell type labels) from
818 the reference onto the query by weighting contributions from reference cells associated
819 through the anchors.

820 The SAMap algorithm[38] was then used to confirm the cell classification, using reference data
821 from the mouse cortex and hippocampus[32]. In short, given two scRNA-seq datasets from
822 different species, SAMap first identifies putative homologous genes through reciprocal
823 BLAST searches. The homologous genes are then used for assessing the similarity, in terms
824 of expression profiles, between cells across the two datasets. In this way, it is possible to
825 transfer cell type annotations from one dataset to the other. In our analysis, we used SAMap
826 default settings, performing BLAST searches on the proteome sequences. Only transferred
827 annotations with SAMap score higher than 0.65 were considered.

828 Single-nuclei RNA-seq statistical analysis

829 Possible imbalances in the proportion of cells originating from good learner and untrained
830 chicks were assessed separately for major cell types and clusters, using a permutation
831 approach[101]. Results were deemed significant for permutation-based FDR < 0.05 and ratio
832 between the groups above 1.5.

833 Differential expression analysis between good learner and untrained chicks, corrected for sex,
834 was performed on pseudo-bulk profiles. Given a specific cluster or cell type, pseudo-bulk
835 profiles are derived by summing up for each gene the count values measured across the cells
836 contained in that cluster / cell type. In this way a single expression profile is obtained for
837 each sample (chick) out of several cells, simulating the same conditions obtained with bulk
838 RNA-seq (hence the name). It has been found that pseudo bulk profiles significantly reduce
839 false positive findings in differential expression analysis for single cell experiments[102]. We
840 used the method based on the negative binomial distribution implemented in the DESeq2 R
841 package[103] to perform pseudo-bulk differential expression analysis separately for each cell
842 type and cluster. Notably, DESeq2 forgo testing transcript that do not match specific criteria,
843 for example if all expression values are equal to zero[103]. The differential analysis was forgone
844 altogether if less than 50 cells were available for any chick for that specific cluster / cell type.
845 Gene set enrichment analysis[104] was performed separately for each cluster and cell type on
846 the basis of the results of the differential expression analysis, using the clusterProfiler R
847 package[105]. Gene Ontology (GO) Biological Processes (BP)[106] were used as gene sets. The

848 GO annotations for chicken have been known to be not as well curated as the annotations of
849 other organisms[107]. Thus, we converted chicken genes to their mouse orthologs and used the
850 GO mouse annotations for the enrichment analysis[108,109]. The simplifyEnrichment
851 Bioconductor package[110] was used for summarizing enrichment results across several
852 analyses.

853 Clustering snRNA-seq transcript with hdWGCNA

854 The hdWGCNA Bioconductor package allows to cluster transcripts from single cell and
855 single nuclei RNA-seq experiments[44]. In short, hdWGCNA first groups individual cells
856 within metacells[111], which are supposed to represent cluster of cells sharing the same
857 identical state. A single expression profile is then computed for each metacell, an operation
858 that usually alleviates the drop-out problem common in single cell data[112]. Metacells were
859 computed within each major cell type (Astrocytes, Oligodendrocytes, GABAergic and
860 Glutamatergic neurons) and sample in our analysis. A matrix of transcript-to-transcript
861 associations is then computed through a formula involving linear correlations elevated to a
862 power coefficient. This formula is devised so that to emphasize strong correlations, and the
863 power coefficient is identified automatically in a data-driven way. Finally, a dynamic tree cut
864 algorithm[113] is used to define strongly connected transcript modules. Transcripts that do not
865 show strong connections are grouped in a default cluster, namely the "grey" module.

866 BLAST analysis on lncRNA

867 We used BLAST (Basic Local Alignment Search Tool) to identify genes with significant
868 sequence homologies to the Ensembl canonical transcripts of the lncRNA genes
869 ENSGALG00010007489 and ENSGALG00010026609. BLAST analyses were conducted
870 against chicken transcripts (cDNAs) and genomic sequences using default parameters. All
871 hits were consolidated into a single list, and duplicates were removed.

872 SDS Electrophoresis and Western Blotting

873 Brain tissue samples were rapidly homogenized in standard Tris-HC buffer containing
874 phosphatase inhibitor and protease inhibitor cocktails. Sodium dodecylsulphate (SDS)
875 solution was added up to a final concentration of 5% and the mixture was incubated at 95 ^{o}C
876 for 3min. Protein concentrations were determined in quadruplicate using a micro
877 bicinchoninic acid protein assay kit (Pierce). Aliquots containing 30 µg of protein in 30 µl of
878 standard sample buffer were subjected to SDS gel electrophoresis and Western blotting[14].

879 The proteins were transferred onto nitrocellulose membranes and stained with 0.1% w/v
880 Ponceau S solution. The stained membranes were digitalized and analyzed for equal total
881 protein loading with LabWorks 4.0 (UVP). The nitrocellulose membranes were then treated
882 with PBS + 0.05% Tween 20 solution and a standard immunochemical staining procedure
883 was carried out with peroxidase-labelled secondary antibodies and the Super-Signal West
884 Pico Chemiluminescent substrate (Pierce, 34580). The following primary antibodies were
885 used: anti-LUC7L (ABIN6745218, antibodies-online), anti-FOXP2 (ab1307, Abcam), anti-
886 RORA (Invitrogen PA1-812), and anti-ROBO1 (Invitrogen PA5-29917). For each antibody,
887 the detected protein band corresponded to the expected molecular weight: LUC7L 44 kDa,
888 FOXP2 80 kDa, ROBO1 250 kDa, and RORA 60 kDa.

889 The optical densities of the target immunostained protein bands were measured with
890 LabWorks 4.0 (UVP). Within each of the gel four internal protein standards (15, 30, 45, and
891 60 µg of total protein) were used for the calibration of autoradiographs. These protein
892 standards were from IMM homogenate fraction isolated from control chick brains. We used
893 the same proteins standards in all of the immunoblotting experiments performed. Optical
894 densities of the immunostained internal standard bands increased proportionally with the
895 amount of loaded protein: for all four proteins, plotting optical density against protein amount
896 (15 – 60 µg) showed a virtually perfect fit to a straight line (Supplementary Figure S30).

897 The relative protein amount for each sample (e.g., LUC7L) was calculated by dividing the
898 optical density of the sample band by the optical density corresponding to 30 µg of protein
899 from the calibration curve. The values obtained by this normalization procedure are termed as
900 the "relative amount" [14] of, e.g., FOXP2. There are differences between the batches resulting
901 in variation in the protein amounts between them. To remove these variations the batch mean
902 was subtracted from the relative amount of protein for each sample and this difference then
903 added to the overall mean of all batches. This quantity is referred to as "standardized relative
904 amount of protein".

905 The densities of the target protein bands were not adjusted to the content of any cellular
906 housekeeping proteins, such as glyceraldehyde 3-phosphate dehydrogenase or beta-actin, since
907 it is not guaranteed that such reference proteins will be unaltered in our experiments[114–117]
908 Instead, we consistently verified equal loading of sample proteins by Ponceau S staining.

909 ## RNA isolation and RT-PCR

910 Total RNA from chick brain samples were isolated with "Nucleospin RNA: DNA, RNA and
911 Protein purification (740955 Macherey-Nagel). Complementary DNA (cDNA) was
912 synthesized by High-Capacity cDNA Reverse Transcription Kit (4368814, Thermofisher
913 Scientific). Relative ENSGAL000100007489 and ENSGALG00010026609 cDNA copy
914 numbers were determined by real-time PCR using the Step One Real-Time PCR System
915 (Applied Biosystems) with the SYBR Green detection method and was normalized to
916 HPRT1[118]. The comparative CT (ddCT) method was used to determine the relative target
917 quantity in samples[119]. The same fraction of IMM RNA isolated from the IMMs of untrained
918 chicks was used in all experiments as a reference sample. Amplicons from randomly chosen
919 samples were sequenced at BGI BIO Solutions Co., Ltd. (Hong-Kong, China). The primer
920 sequences are provided in Table 1.

921

922 Table 1: Primer sequences for RT-PCR

Gene name	Primer Sequences (5'-3') (forward/reverse)
ENSGAL000100007489 (GLUBK89)	GGTGCTCAAATTTGGACTTCTG/ TTTCTCACATGCAAGGCTTATTC
ENSGALG00010026609 (LNCRNA6609)	GCTTACCTCCAGGGTCAAAT/ TGGTGCCAACTCTTCTCTTC
HPRT1	ACCTGGACTTGTTCTG CATAC/ TCTCTCTGTCCTGTCC ATGAT

923

924 ## Subcellular localization of lncRNAs

925 Subcellular distribution of GLUBK89 and LNCRNA6609 lncRNAs was studied by a
926 cytoplasmic and nuclear purification kit. (Cat. No. 21000 Norgen Biotek Corporation). The
927 total RNA fraction was obtained separately from the nuclear and cytoplasmic fractions of the
928 IMM and RT-PCR performed as described above.

929 ## Interpretation of a correlation between a molecular quantity and
930 preference score

931 A significant association between a molecular quantity (e.g., protein abundance, gene
932 expression level, etc.) and preference score may indicate (i) that the change in the quantity is
933 induced by the learning process, or that (ii) the quantity may facilitate / impede the learning
934 process, i.e., it provides some type of predisposition. In order to distinguish between these

935 possibilities, we use a design comprising two groups of chicks: a trained group, where both
936 the molecular quantity under examination and the preference score are measured, and an
937 untrained group, where only the molecular quantity can be measured. The molecular quantity
938 is regressed over the preference score in the trained group using a linear model.

939 The y-intercept at preference score 50 is an estimate of the molecular quantity when no
940 learning occurs in trained chicks. It thus measures the effect of procedures such as handling,
941 locomotor activity, sensory stimulation etc. that occur during training but which are separable
942 from learning (i.e. side-effects). If the intercept at preference score 50 is not significantly
943 different from the mean value in untrained chicks, there is no evidence of side-effects.

944 In addition to a significant correlation with preference score, one would expect learning to
945 give a y-intercept corresponding to the maximum preference score (100) that is significantly
946 greater than the mean value for untrained chicks.

947 If the correlation with preference score is the *result* of learning (see (i) above), the residual
948 variance about the regression line should be *statistically indistinguishable* from the variance
949 in untrained chicks. This is because neither variance has a learning component but they are
950 otherwise samples from the same population.

951 If the correlation is due to a predisposition (see (ii) above), training has no effect on the
952 molecular quantity and the variance of the trained and untrained chicks should be statistically
953 homogeneous. The correlation accounts for some of the variance in trained chicks and the
954 variance about the regression line is correspondingly reduced: a significant *reduction* in
955 residual variance about the regression line, relative to the variance in untrained chicks, is
956 therefore evidence of a *predisposition*[10].

957 Fluorescent in-situ hybridization

958 The experiment was conducted in four batches. Each batch contained a good learner, a poor
959 learner and an untrained chick from the same hatch of eggs, that is 12 samples in total.
960 Hatching, rearing and training were performed as described for the other experiments. After
961 killing, brain tissue was removed and fixed in 4% paraformaldehyde solution for 3 days. The
962 brains were then transferred to a freshly prepared methacarn solution consisting of 60%
963 (vol/vol) absolute methanol, 30% chloroform, and 10% glacial acetic acid[120]. Dehydration
964 was performed through graded ethanol, followed by xylene, then paraffin wax for
965 embedding. Samples were sectioned at 7 µm onto superfrost glass slides. For each brain, two
966 coronal sections were selected corresponding to each of the coordinates A7.0, A7.6 and A8.0

967 of the chick brain atlas by Kuenzel and Masson[121]. Measurements of cellular fluorescence
968 were made on each section within a rectangular sampling frame measuring 0.3 mm x 0.9 mm
969 in the IMM and a square sampling frame 0.3 mm x 0.3 mm in the adjacent nidopallium
970 (Supplementary Figure S31). Fluorescence was measured at each of the three anteroposterior
971 levels for LOC121109133-O1, SLC and PPIB, and nuclei were stained with DAPI. One of
972 the sections from each chick was stained for GAD instead of PPIB.

973 In any one batch, the sections stained for GAD were at the same anteroposterior level in all
974 three experimental groups. There were three GAD sections from the posterior IMM in batch
975 1, three sections from the intermediate IMM in each of batches 2 and 3 and three sections
976 from the anterior IMM in batch 4.

977 RNAscope was conducted to manufacturer specifications using probes Gg-GAD2-C1
978 (reference number 1300861-C1, target region 448-1421, 20 pairs), Gg-PPIB-C1 (reference
979 number 453371, target region 67-876, 16 pairs), Gg-SLC17A6-C2 (reference number
980 1300871-C1, target region 591-1864, 20 pairs), Gg- LOC121109133-O1-C3 (reference
981 number 1300881-C3, target region 91-656, 11 pairs), and Opal fluorophores 520, 570, and
982 650. Samples were scanned on a Zeiss Axioscan 7 at x20 and images were imported into
983 QuPath[56]. Cell detection in QuPath was used, with the following parameters. Detection
984 channel DAPI, requested pixel size 0.5 µm, background radius 8um, use opening by
985 reconstruction yes, median filter radius 0um, sigma 1.5 µm, minimum area 10um², maximum
986 area 400um², threshold 50, cell expansion 5 µm, include cell nucleus yes, smooth boundaries
987 yes, make measurements yes, and visually validated. Subcellular detection in QuPath was
988 then used, with the following parameters, detection threshold SLC 2000, PPIB 400, LNC
989 4000, smooth before detection yes, spilt by intensity yes, split by shape yes, expected spot
990 size 1um², min spot size 0.1 µm², max spot size 4 µm², include clusters yes (still being
991 optimized for all sections), and visually validated.

992 Fluorescent in-situ hybridization data analysis

993 Quantitative measurements on fluorescent intensities provided by QuPath were analyzed
994 using the R Statistical Software[122]. Two main analyses were performed:

995 - *Contrasting GLUBK89 expression values across experimental groups*.
996 For this analysis, only sections stained for SLC17A6, PPIB and GLUBK89 were used
997 (see Supplementary Figure 26 for a schema of the experimental design). Glutamatergic
998 nuclei were identified by thresholding the SLC17A6 marker (> 1700), while a normalized

expression value for GLUBK89 in each nucleus was computed by dividing for its fluorescent intensity for the fluorescent intensity associated to the housekeeping gene PPIB. These normalized expression values at the nucleus level were then averaged across sample, tissue (IMM and nidopallium), and side (left, right side of the brain). The resulting average values were modeled, independently in the IMM and nidopallium, according to the following linear mixed model[123]:

norm_expression_values ~ group * side, random = ~ 1|batch/chick

Which means that we model the normalized expression values depending on experimental groups (good learners, poor learners, untrained), side (left or right), and their interaction, as well as taking into account the variability provided by the nested random effect chick (sample) and batch (see Supplementary Figure S26 for batch definitions). The statistical significance of contrasts between experimental group levels within this model were investigated with the Tukey method[124,125].

- *Contrasting GLUBK89 expression between GABAergic and glutamatergic cells*. Sections stained for SLC17A6, GAD2 and GLUBK89 were considered for this analysis. Glutamatergic and GABAergic nuclei were identified by thresholding the SLC17A6 (> 1700) and GAD2 (> 630) markers, respectively. The expression of GLUBK89 was contrasted between the two cell types using a t-test.

Data and code availability

Data from the scRNA-seq experiments are available in Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE299793), while code for reproducing the results is available in GitHub (https://github.com/vlagani/CHARM-Vis).

Funding

This work has been funded by (i) the European Union's Horizon 2020 research and innovation program (project CHARM-Vis, project ID 867429), (ii) by the Shota Rustaveli National Science Foundation, project number NFR-22-8692, and (iii) by basic funding from Ilia State University and King Abdullah University of Science and Technology.

Author contributions

V.L., R.S., Z.Kh., and D.C.G. conceived the project and designed the experiments. L.C., L.T., T.B., V.B., S.L., and A.J. performed the experiments. A.C.G.A., G.S., X.M.D.M., R.L., B.J.M., V.L. performed bioinformatic analysis and analyzed the data. V.L., R.S., Z.Kh., B.J.M., D.G.C., L.A.I., and J.T. interpreted the results and wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

References

1. Asok, A., Leroy, F., Rayman, J. B. & Kandel, E. R. Molecular Mechanisms of the Memory Trace. *Trends in Neurosciences* vol. 42 Preprint at https://doi.org/10.1016/j.tins.2018.10.005 (2019).

2. Glanzman, D. L. Common Mechanisms of Synaptic Plasticity in Vertebrates and Invertebrates. *Current Biology* vol. 20 Preprint at https://doi.org/10.1016/j.cub.2009.10.023 (2010).

3. Horn, G. Pathways of the past: The imprint of memory. *Nature Reviews Neuroscience* vol. 5 108–120 Preprint at https://doi.org/10.1038/nrn1324 (2004).

4. Bateson, P. P. The characteristics and context of imprinting. *Biol. Rev. Camb. Philos. Soc.* **41**, (1966).

5. Bolhuis, J. J. Mechanisms of Avian Inprinting: a Review. *Biological Reviews* **66**, (1991).

6. McCabe, B. J. Visual imprinting in birds: Behavior, models, and neural mechanisms. *Frontiers in Physiology* vol. 10 658 Preprint at https://doi.org/10.3389/fphys.2019.00658 (2019).

7. Solomonia, R. O. & McCabe, B. J. Molecular mechanisms of memory in imprinting. *Neuroscience and Biobehavioral Reviews* vol. 50 56–69 Preprint at https://doi.org/10.1016/j.neubiorev.2014.09.013 (2015).

8. Horn, G. *Memory, Imprinting, and the Brain*. (Oxford University Press, 1985).

9. Reiner, A. *et al.* Revised Nomenclature for Avian Telencephalon and Some Related Brainstem Nuclei. *Journal of Comparative Neurology* vol. 473 Preprint at https://doi.org/10.1002/cne.20118 (2004).

10. Margvelani, G., Meparishvili, M., Kiguradze, T., McCabe, B. J. & Solomonia, R. Micro-RNAs, their target proteins, predispositions and the memory of filial imprinting. *Sci. Rep.* **8**, 1–12 (2018).

11. Cipolla-Neto, J., Horn, G. & McCabe, B. J. Hemispheric asymmetry and imprinting: The effect of sequential lesions to the hyperstriatum ventrale. *Exp. Brain Res.* **48**, (1982).

12. Margvelani, G., Meparishvili, M., Tevdoradze, E., McCabe, B. J. & Solomonia, R. Mitochondrial fusion and fission proteins and the recognition memory of imprinting in domestic chicks. *Neuroreport* **29**, 128–133 (2018).

13. Chitadze, L., Tevdoradze, E., Kiguradze, T., McCabe, B. J. & Solomonia, R. Filial imprinting in domestic chicks; cytoplasmic polyadenylation element binding protein 3, predispositions and learning. *Neuroreport* **34**, 144–149 (2023).

14. Chitadze, L. *et al.* Src-NADH dehydrogenase subunit 2 complex and recognition memory of imprinting in domestic chicks. *PLoS One* **19**, e0297166 (2024).

15. Chitadze, L. *et al.* Chicken cognin interactome and the memory of filial imprinting. *Neuroreport* **31**, 399–405 (2020).

1075 16. Meparishvili, M., Nozadze, M., Margvelani, G., McCabe, B. J. & Solomonia, R. O. A
1076 Proteomic Study of Memory After Imprinting in the Domestic Chick. *Front. Behav.*
1077 *Neurosci.* **9**, 319 (2015).

1078 17. Solomonia, R. O., McCabe, B. J. & Horn, G. Neural cell adhesion molecules, learning,
1079 and memory in the domestic chick. *Behavioral Neuroscience* **112**, 646–655 (1998).

1080 18. McCabe, B. J. & Horn, G. Learning and memory: Regional changes in N-methyl-D-
1081 aspartate receptors in the chick brain after imprinting. *Proc. Natl. Acad. Sci. U. S. A.*
1082 **85**, 2849–2853 (1988).

1083 19. McCabe, B. J. & Horn, G. Synaptic Transmission and Recognition Memory: Time
1084 Course of Changes in N-Methyl-D-Aspartate Receptors After Imprinting. *Behavioral*
1085 *Neuroscience* **105**, (1991).

1086 20. Harvey, R. J., McCabe, B. J., Solomonia, R. O., Horn, G. & Darlison, M. G.
1087 Expression of the GABAA receptor γ4-subunit gene: Anatomical distribution of the
1088 corresponding mRNA in the domestic chick forebrain and the effect of imprinting
1089 training. *European Journal of Neuroscience* **10**, (1998).

1090 21. Solomonia, R. O. *et al.* AMPA receptor phosphorylation and recognition memory:
1091 Learning-related, time-dependent changes in the chick brain following filial
1092 imprinting. *Exp. Brain Res.* **226**, 297–308 (2013).

1093 22. Ambalavanar, R., McCabe, B. J., Potter, K. N. & Horn, G. Learning-related Fos-like
1094 immunoreactivity in the chick brain: Time-course and co-localization with GABA and
1095 parvalbumin. *Neuroscience* **93**, (1999).

1096 23. Wang, D. & Bodovitz, S. Single cell analysis: The new frontier in 'omics'. *Trends in*
1097 *Biotechnology* vol. 28 Preprint at https://doi.org/10.1016/j.tibtech.2010.03.002 (2010).

1098 24. Regev, A. *et al.* The human cell atlas. *Elife* **6**, (2017).

1099 25. Yao, Z. *et al.* A high-resolution transcriptomic and spatial atlas of cell types in the
1100 whole mouse brain. *Nature* **624**, (2023).

1101 26. Crocker, A., Guan, X. J., Murphy, C. T. & Murthy, M. Cell-Type-Specific
1102 Transcriptome Analysis in the Drosophila Mushroom Body Reveals Memory-Related
1103 Changes in Gene Expression. *Cell Rep.* **15**, (2016).

1104 27. Chen, X. *et al.* Transcriptomic mapping uncovers Purkinje neuron plasticity driving
1105 learning. *Nature* **605**, (2022).

1106 28. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*
1107 **570**, (2019).

1108 29. Morabito, S., Miyoshi, E., Michael, N., Shahin, S. & Swarup, V. Single-nuclei
1109 chromatin accessibility and transcriptomics unravels altered human oligodendrocyte
1110 heterogeneity in Alzheimer's disease. *Alzheimer's & Dementia* **16**, (2020).

1111 30. Meparishvili, M., Chitadze, L., Lagani, V., McCabe, B. & Solomonia, R. Src and
1112 Memory: A Study of Filial Imprinting and Predispositions in the Domestic Chick.
1113 *Front. Physiol.* **12**, 736999 (2021).

1114 31. Hecker, N. *et al.* Enhancer-driven cell type comparison reveals similarities between
1115 the mammalian and bird pallium. *Science (1979).* **387**, (2025).

1116 32. Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and
1117 hippocampal formation. *Cell* **184**, 3222-3241.e26 (2021).

1118 33. Yao, Z. *et al.* A transcriptomic and epigenomic cell atlas of the mouse primary motor
1119 cortex. *Nature* **598**, (2021).

1120 34. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science (1979).* **347**,
1121 (2015).

1122 35. Zaremba, B. *et al.* Developmental origins and evolution of pallial cell types and
1123 structures in birds. *Science (1979).* **387**, (2025).

1124 36. Stachniak, T. J. *et al.* Postmitotic Prox1 Expression Controls the Final Specification of
1125 Cortical VIP Interneuron Subtypes. *The Journal of Neuroscience* **41**, 8150 (2021).

1126 37. Chamessian, A. *et al.* Transcriptional Profiling of Somatostatin Interneurons in the
1127 Spinal Dorsal Horn. *Scientific Reports 2018 8:1* **8**, 1–16 (2018).

1128 38. Tarashansky, A. J. *et al.* Mapping single-cell atlases throughout metazoa unravels cell
1129 type evolution. *Elife* **10**, (2021).

1130 39. Liu, Y. *et al.* Robustness of single-cell RNA-seq for identifying differentially
1131 expressed genes. *BMC Genomics* **24**, (2023).

1132 40. Aoki, N. *et al.* Blockade of muscarinic acetylcholine receptor by scopolamine impairs
1133 the memory formation of filial imprinting in domestic chicks (Gallus Gallus
1134 domesticus). *Behavioural Brain Research* **379**, 112291 (2020).

1135 41. Mori, C. *et al.* Gene expression profiles of the muscarinic acetylcholine receptors in
1136 brain regions relating to filial imprinting of newly-hatched domestic chicks.
1137 *Behavioural Brain Research* **420**, (2022).

1138 42. Aoki, N. *et al.* Subtype-selective contribution of muscarinic acetylcholine receptors for
1139 filial imprinting in newly-hatched domestic chicks. *Behavioural Brain Research* **424**,
1140 (2022).

1141 43. Aoki, N. *et al.* GABA-A and GABA-B receptors in filial imprinting linked with
1142 opening and closing of the sensitive period in domestic chicks (gallus gallus
1143 domesticus). *Front. Physiol.* **9**, (2018).

1144 44. Morabito, S., Reese, F., Rahimzadeh, N., Miyoshi, E. & Swarup, V. hdWGCNA
1145 identifies co-expression networks in high-dimensional transcriptomics data. *Cell*
1146 *Reports Methods* **3**, (2023).

1147 45. Wang, Q., Roy, B. & Dwivedi, Y. Co-expression network modeling identifies key long
1148 non-coding RNA and mRNA modules in altering molecular phenotype to develop
1149 stress-induced depression in rats. *Transl. Psychiatry* **9**, (2019).

1150 46. Mukherjee, S., Banerjee, B., Karasik, D. & Frenkel-Morgenstern, M. Mrna-lncrna co-
1151 expression network analysis reveals the role of lncrnas in immune dysfunction during
1152 severe sars-cov-2 infection. *Viruses* **13**, (2021).

1153 47. Keihani, S. *et al.* The long noncoding RNA neuroLNC regulates presynaptic activity
1154 by interacting with the neurodegeneration-associated protein TDP-43. *Sci. Adv.* **5**,
1155 (2019).

48. Holtmaat, A. & Svoboda, K. Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Reviews Neuroscience* vol. 10 Preprint at https://doi.org/10.1038/nrn2699 (2009).

49. Ota, Y., Zanetti, A. T. & Hallock, R. M. The role of astrocytes in the regulation of synaptic plasticity and memory formation. *Neural Plasticity* vol. 2013 Preprint at https://doi.org/10.1155/2013/185463 (2013).

50. Sengar, A. S. *et al.* Control of Long-Term Synaptic Potentiation and Learning by Alternative Splicing of the NMDA Receptor Subunit GluN1. *Cell Rep.* **29**, (2019).

51. Alberini, C. M. Transcription factors in long-term memory and synaptic plasticity. *Physiological Reviews* vol. 89 Preprint at https://doi.org/10.1152/physrev.00017.2008 (2009).

52. Shen, K. & Cowan, C. W. Guidance molecules in synapse formation and plasticity. *Cold Spring Harbor perspectives in biology* vol. 2 Preprint at https://doi.org/10.1101/cshperspect.a001842 (2010).

53. Wong, L. W., Tann, J. Y., Ibanez, C. F. & Sajikumar, S. The p75 neurotrophin receptor is an essential mediator of impairments in hippocampal-dependent associative plasticity and memory induced by sleep deprivation. *Journal of Neuroscience* **39**, (2019).

54. Yu, M. *et al.* Overexpression of Kcnmb2 in Dorsal CA1 of Offspring Mice Rescues Hippocampal Dysfunction Caused by a Methyl Donor-Rich Paternal Diet. *Front. Cell. Neurosci.* **12**, (2018).

55. Solomonia, R. O., McCabe, B. J., Jackson, A. P. & Horn, G. Clathrin proteins and recognition memory. *Neuroscience* **80**, 59–67 (1997).

56. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**, (2017).

57. Rose, S. P. R. God's organism? The chick as a model system for memory studies. *Learning and Memory* vol. 7 1–17 Preprint at https://doi.org/10.1101/lm.7.1.1 (2000).

58. Matsushima, T., Izumi, T. & Vallortigara, G. The domestic chick as an animal model of autism spectrum disorder: building adaptive social perceptions through prenatally formed predispositions. *Frontiers in Neuroscience* vol. 18 Preprint at https://doi.org/10.3389/fnins.2024.1279947 (2024).

59. Atoji, Y. & Karim, M. R. Homology of the mesopallium in the adult chicken identified by gene expression of the neocortical marker cholecystokinin. *Neurosci. Lett.* **562**, 85–89 (2014).

60. Suzuki, I. K., Kawasaki, T., Gojobori, T. & Hirata, T. The Temporal Sequence of the Mammalian Neocortical Neurogenetic Program Drives Mediolateral Pattern in the Chick Pallium. *Dev. Cell* **22**, (2012).

61. Briscoe, S. D., Albertin, C. B., Rowell, J. J. & Ragsdale, C. W. Neocortical Association Cell Types in the Forebrain of Birds and Alligators. *Current Biology* **28**, 686-696.e6 (2018).

62. Horn, G. & McCabe, B. J. Predispositions and preferences. Effects on imprinting of lesions to the chick brain. *Anim. Behav.* **32**, 288–292 (1984).

1198 63. Horn, G., Nicol, A. U. & Brown, M. W. Tracking memory's trace. *Proc. Natl. Acad.*
1199 *Sci. U. S. A.* **98**, (2001).

1200 64. Brown, M. W. & Horn, G. Learning-related Alterations in the Visual Responsiveness
1201 of Neurons in a Memory System of the Chick Brain. *European Journal of*
1202 *Neuroscience* **6**, (1994).

1203 65. Jackson, C. *et al.* Dynamics of a Memory Trace: Effects of Sleep on Consolidation.
1204 *Current Biology* **18**, (2008).

1205 66. Nicol, A. U., Brown, M. W. & Horn, G. Neurophysiological Investigations of a
1206 Recognition Memory System for Imprinting in the Domestic Chick. *European Journal*
1207 *of Neuroscience* **7**, (1995).

1208 67. Meredith, R. M., Mccabe, B. J., Kendrick, K. M. & Horn, G. Amino acid
1209 neurotransmitter release and learning: A study of visual imprinting. *Neuroscience* **126**,
1210 249–256 (2004).

1211 68. Solomonia, R. O. *et al.* Different forms of MARCKS protein are involved in memory
1212 formation in the learning process of imprinting. *Exp. Brain Res.* **188**, 323–330 (2008).

1213 69. Grinman, E., Espadas, I. & Puthanveettil, S. V. Emerging roles for long noncoding
1214 RNAs in learning, memory and associated disorders. *Neurobiol. Learn. Mem.* **163**,
1215 (2019).

1216 70. Irwin, A. B., Bahabry, R. & Lubin, F. D. A putative role for lncRNAs in epigenetic
1217 regulation of memory: Memory Regulation via lncRNAs. *Neurochemistry*
1218 *International* vol. 150 Preprint at https://doi.org/10.1016/j.neuint.2021.105184 (2021).

1219 71. Kemenes, G., Benjamin, P. R. & Kemenes, I. The role of non-coding RNAs in the
1220 formation of long-term associative memory after single-trial learning in Lymnaea.
1221 *Front. Behav. Neurosci.* **16**, 1005867 (2022).

1222 72. Zhang, J. *et al.* Glutamate-activated BK channel complexes formed with NMDA
1223 receptors. *Proc. Natl. Acad. Sci. U. S. A.* **115**, (2018).

1224 73. Ancatén-González, C. *et al.* BK channels mediate a presynaptic form of mGluR-LTD
1225 in the neonatal hippocampus. *Proc. Natl. Acad. Sci. U. S. A.* **122**, (2025).

1226 74. Zhu, Y. *et al.* The long noncoding RNA glycoLINC assembles a lower glycolytic
1227 metabolon to promote glycolysis. *Mol. Cell* **82**, (2022).

1228 75. Daniels, N. J. *et al.* Functional analyses of human LUC7-like proteins involved in
1229 splicing regulation and myeloid neoplasms. *Cell Rep.* **35**, (2021).

1230 76. Kenny, C. J. *et al.* LUC7 proteins define two major classes of 5′ splice sites in animals
1231 and plants. *Nature Communications 2025 16:1* **16**, 1–15 (2025).

1232 77. Shen, L. *et al.* Loss-of-function mutation in PRMT9 causes abnormal synapse
1233 development by dysregulation of RNA alternative splicing. *Nature Communications*
1234 *2024 15:1* **15**, 1–20 (2024).

1235 78. Jolly, S., Journiac, N., Vernet-Der Garabedian, B. & Mariani, J. RORalpha, a key to
1236 the development and functioning of the brain. *Cerebellum* **11**, 451–452 (2012).

1237 79. Sarachana, T. & Hu, V. W. Genome-wide identification of transcriptional targets of
1238 RORA reveals direct regulation of multiple genes associated with autism spectrum
1239 disorder. *Mol. Autism* **4**, 1–19 (2013).

1240 80. Akila, S. *et al.* Molecular Signatures of Resilience to Alzheimer's Disease in
1241 Neocortical Layer 4 Neurons. *bioRxiv* 2024.11.03.621787 (2024)
1242 doi:10.1101/2024.11.03.621787.

1243 81. Valle-Bautista, R. *et al.* From songbird to humans: The multifaceted roles of FOXP2
1244 in speech and motor learning. *Neurosci. Biobehav. Rev.* **167**, (2024).

1245 82. Li, S., Weidenfeld, J. & Morrisey, E. E. Transcriptional and DNA binding activity of
1246 the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions.
1247 *Mol. Cell. Biol.* **24**, 809–822 (2004).

1248 83. Tiunova, A. A., Komissarova, N. V & Anokhin, K. V. Mapping the Neural Substrates
1249 of Recent and Remote Visual Imprinting Memory in the Chick Brain. *Front. Physiol.*
1250 **10**, (2019).

1251 84. Solomonia, R. O., Kiguradze, T., McCabe, B. J. & Horn, G. Neural cell adhesion
1252 molecules, CaM kinase II and long-term memory in the chick. *Neuroreport* **11**, 3139–
1253 3143 (2000).

1254 85. Behroozi, M. *et al.* Functional MRI of imprinting memory in awake newborn domestic
1255 chicks. *Communications Biology 2024 7:1* **7**, 1–13 (2024).

1256 86. Brose, K. *et al.* Slit proteins bind Robo receptors and have an evolutionarily conserved
1257 role in repulsive axon guidance. *Cell* **96**, 795–806 (1999).

1258 87. Whitford, K. L. *et al.* Regulation of cortical dendrite development by Slit-Robo
1259 interactions. *Neuron* **33**, 47–61 (2002).

1260 88. Pfenning, A. R. *et al.* Convergent transcriptional specializations in the brains of
1261 humans and song-learning birds. *Science (1979).* **346**, (2014).

1262 89. Wang, R. *et al.* Convergent differential regulation of SLIT-ROBO axon guidance
1263 genes in the brains of vocal learners. *Journal of Comparative Neurology* **523**, (2015).

1264 90. Park, J. H., Pak, H. J., Riew, T. R., Shin, Y. J. & Lee, M. Y. Increased expression of
1265 Slit2 and its receptors Robo1 and Robo4 in reactive astrocytes of the rat hippocampus
1266 after transient forebrain ischemia. *Brain Res.* **1634**, (2016).

1267 91. Chen, X. *et al.* Computational models for lncRNA function prediction and functional
1268 similarity calculation. *Brief. Funct. Genomics* **18**, 58–82 (2019).

1269 92. Smith, F. V. & Bird, M. W. The relative attraction for the domestic chick of
1270 combinations of stimuli in different sensory modalities. *Anim. Behav.* **11**, 300–305
1271 (1963).

1272 93. Horn, G. & Johnson, M. H. Memory systems in the chick: Dissociations and neuronal
1273 analysis. *Neuropsychologia* **27**, 1–22 (1989).

1274 94. Davies, D. C., Horn, G. & McCabe, B. J. Noradrenaline and learning: Effects of the
1275 noradrenergic neurotoxin DSP4 on imprinting in the domestic chick. *Behavioral*
1276 *Neuroscience* **99**, 652–660 (1985).

1277 95. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single
1278 cells. *Nat. Commun.* **8**, (2017).

1279 96. Alvarez, M. *et al.* Enhancing droplet-based single-nucleus RNA-seq resolution using
1280 the semi-supervised machine learning classifier DIEM. *Sci. Rep.* **10**, (2020).

1281 97. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in
1282 Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**,
1283 (2019).

1284 98. Fleming, S. J. *et al.* Unsupervised removal of systematic background noise from
1285 droplet-based single-cell experiments using CellBender. *Nat. Methods* **20**, (2023).

1286 99. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, (2019).

1287 100. Waltman, L. & Van Eck, N. J. A smart local moving algorithm for large-scale
1288 modularity-based community detection. *European Physical Journal B* **86**, (2013).

1289 101. Miller, S. A. *et al.* LSD1 and aberrant DNA methylation mediate persistence of
1290 enteroendocrine progenitors that support BRAF-mutant colorectal cancer. *Cancer Res.*
1291 **81**, (2021).

1292 102. Squair, J. W. *et al.* Confronting false discoveries in single-cell differential expression.
1293 *Nat. Commun.* **12**, (2021).

1294 103. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
1295 dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

1296 104. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for
1297 interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**,
1298 15545–50 (2005).

1299 105. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics
1300 data. *Innovation (Cambridge (Mass.))* **2**, (2021).

1301 106. The Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic*
1302 *Acids Res.* **43**, D1049–D1056 (2015).

1303 107. Buza, T. J., Mccarthy, F. M., Wang, N., Bridges, S. M. & Burgess, S. C. Gene
1304 Ontology annotation quality analysis in model eukaryotes. *Nucleic Acids Res.* **36**,
1305 (2008).

1306 108. Kui, H. *et al.* Gene expression profiles of specific chicken skeletal muscles. *Sci. Data*
1307 **9**, (2022).

1308 109. Nie, H. *et al.* Gene expression in chicken reveals correlation with structural genomic
1309 features and conserved patterns of transcription in the terrestrial vertebrates. *PLoS One*
1310 **5**, (2010).

1311 110. Gu, Z. & Hübschmann, D. simplifyEnrichment: A Bioconductor Package for
1312 Clustering and Visualizing Functional Enrichment Results. *Genomics Proteomics*
1313 *Bioinformatics* **21**, (2023).

1314 111. Baran, Y. *et al.* MetaCell: Analysis of single-cell RNA-seq data using K-nn graph
1315 partitions. *Genome Biol.* **20**, (2019).

112. Bilous, M. *et al.* Metacells untangle large and complex single-cell transcriptome networks. *BMC Bioinformatics* **23**, (2022).

113. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics* **24**, (2008).

114. Dittmer, A. & Dittmer, J. β-Actin is not a reliable loading control in Western blot analysis. *Electrophoresis* **27**, (2006).

115. Li, R. & Shen, Y. An old method facing a new challenge: Re-visiting housekeeping proteins as internal reference control for neuroscience research. *Life Sciences* vol. 92 Preprint at https://doi.org/10.1016/j.lfs.2013.02.014 (2013).

116. Ghosh, R., Gilda, J. E. & Gomes, A. V. The necessity of and strategies for improving confidence in the accuracy of western blots. *Expert Review of Proteomics* vol. 11 Preprint at https://doi.org/10.1586/14789450.2014.939635 (2014).

117. Chen, W. & Xu, W. H. β-Actin as a loading control: Less than 2 μg of total protein should be loaded. *Electrophoresis* **36**, (2015).

118. Lenart, J., Kogut, K. & Salinska, E. Lateralization of housekeeping genes in the brain of one-day old chicks. *Gene Expression Patterns* **25–26**, (2017).

119. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* **25**, (2001).

120. Shibutani, M., Uneyama, C., Miyazaki, K., Toyoda, K. & Hirose, M. Methacarn fixation: A novel tool for analysis of gene expressions in paraffin-embedded tissue specimens. *Laboratory Investigation* **80**, (2000).

121. Kuenzel, Wayne J. and Masson, M. *A Stereotaxic Atlas of the Brain of the Chick (Gallus Domesticus)*. *Journal of Comparative Neurology* vol. 181 (1988).

122. R Development Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing* Preprint at http://www.r-project.org (2016).

123. Pinheiro, J. *et al.* Package nlme: Linear and Nonlinear Mixed Effects Models. *R package* (2020).

124. Tukey, J. W. Comparing Individual Means in the Analysis of Variance. *Biometrics* **5**, (1949).

125. Lenth, R. V. emmeans: Estimated Marginal Means, aka Least-Squares Means. *R package version 1.10.2.090002* vol. 34 (2024).

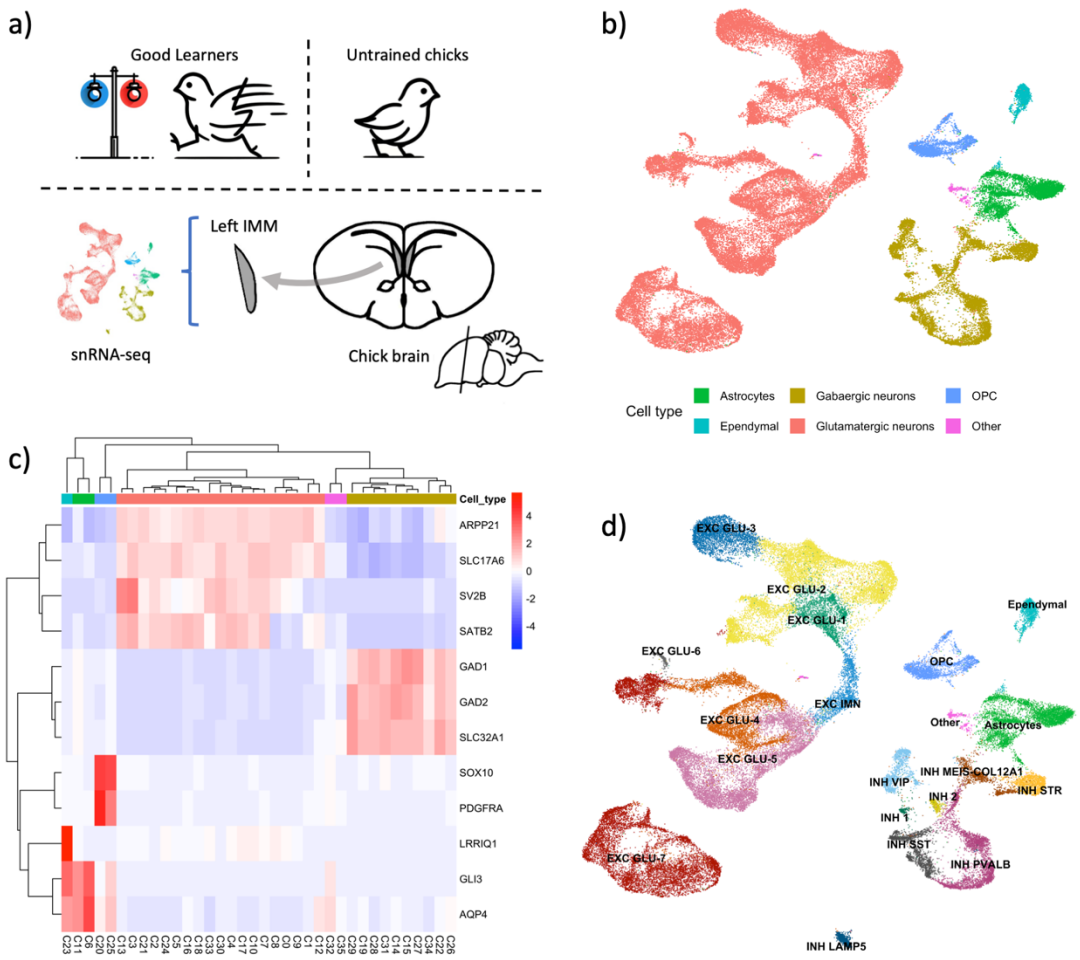

Figure 1: (a) Graphical presentation of the study design. 22-28h old chicks are either subjected to a preference test for assessing their imprinting or left untrained. Trained chicks are selected only if consistently directing themselves towards the training stimulus (good learners; see methods for details). The left portion of the intermediate and medial mesopallium (IMM) is collected from all selected chicks. Single nuclei transcriptomics data are then measured. (b) UMAP visualizing the 54763 IMM cells profiled with snRNA-seq. Color indicates major cell classifications: astrocytes, oligodendrocyte precursors (OPC), as well as GABAergic and glutamatergic neurons. (c) Heatmap showing the expression levels of gene markers for major cell types. (d) Localization of cell subtypes, see main text for details.

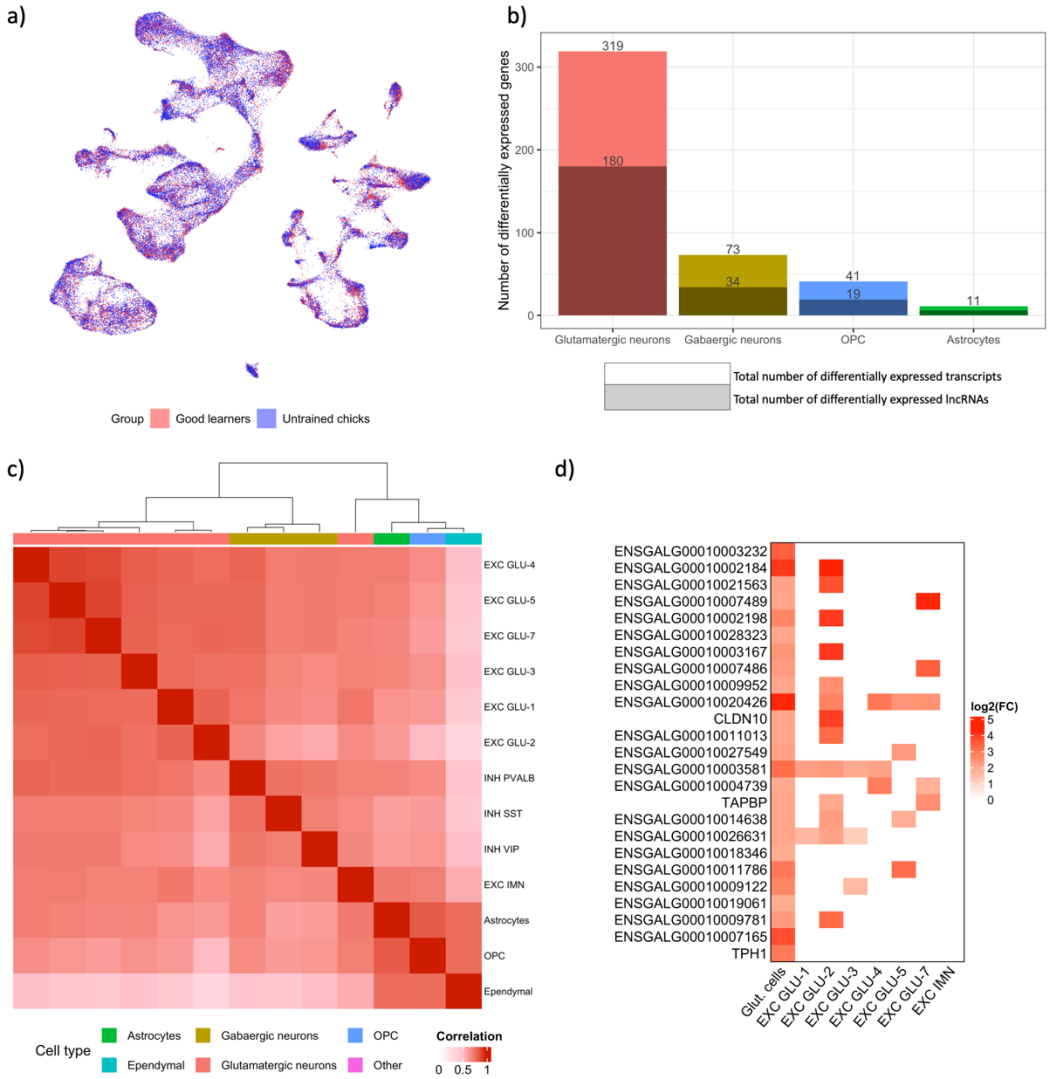

Figure 2: (a) UMAP visualizing IMM cells according to the group of their sample of origin, red good learners and blue untrained chicks. (b) Number of statistically significant transcriptomic changes (log FC > 1 in absolute value, FDR <= 0.05). Darker shaded bars indicate the number of non-coding RNA found differentially expressed. Six lncRNAs were differentially expressed for Astrocytes (c) Correlations among log2 fold changes across different clusters represented as heatmap. The dendrogram on top of the heatmap mostly splits clusters according to their original cell type, i.e., glutamatergic and GABAergic neurons, as well as non-neuronal cells. (d) Log2 FC for the transcripts with the highest increase of expression in glutamatergic cells. The log2 FC is reported for results with FDR <= 0.05, as well as with at least 50 cells in each pseudo-bulk sample (see Methods for details).

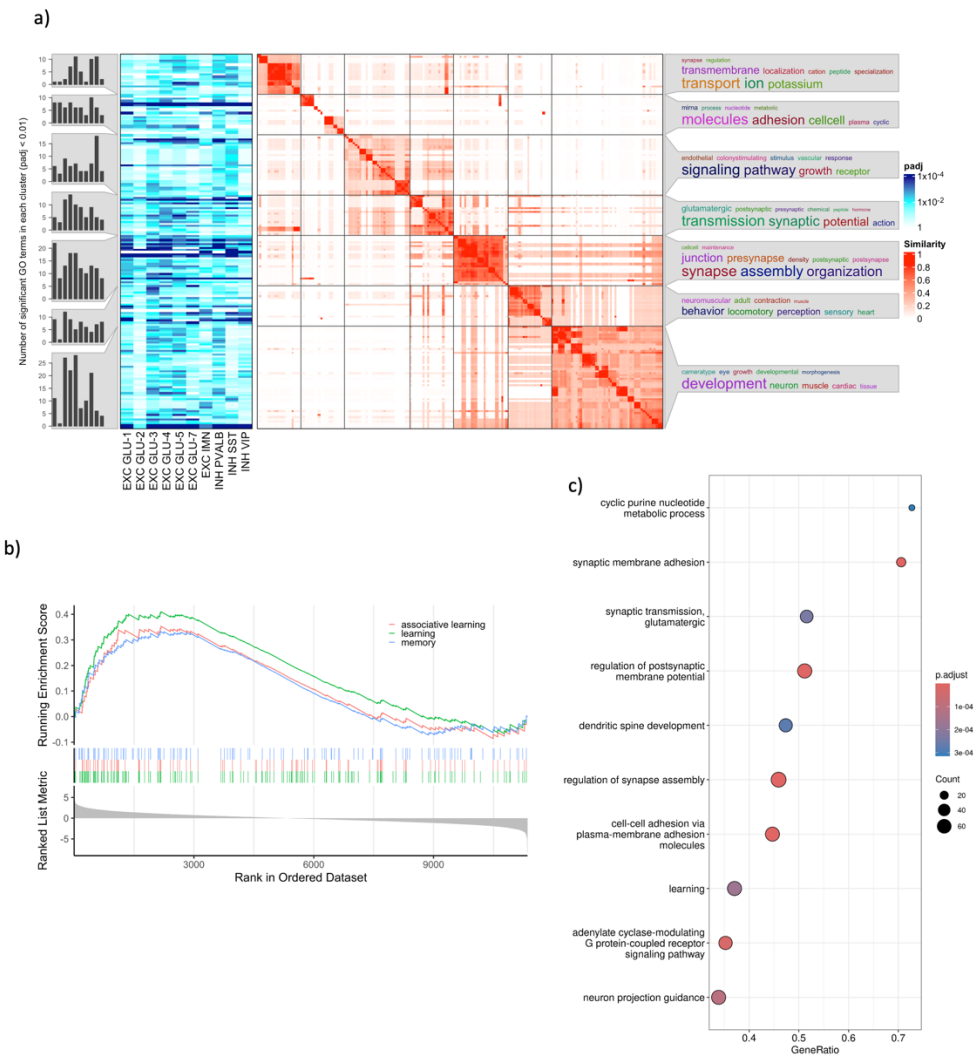

Figure 3: (a) Gene Ontology biological processes (BPs) significantly upregulated (FDR < 0.01) in good learner chicks, across neuronal subtypes, clustered and summarized. Central heatmap reports the BP-to-BP term similarity (see Methods), with BP clusters formed through the DynamicTreeCut algorithm. Left side heatmap reports the significance of each BP in each cell type (darker blue shades indicating higher significance). The word clouds on the right present the frequency of the most common terms within the BPs description. Bar plots on the left indicate the number of significantly upregulated BPs for each cell subtype and BP cluster. (b) Enrichment of relevant BPs in glutamatergic neurons: GO:0007612 learning, GO:0007613 memory, GO:0008306 associative learning. (c) Dot plot presenting the most upregulated BP in glutamatergic neurons; color indicate significance, size the number of upregulated genes within the BP.

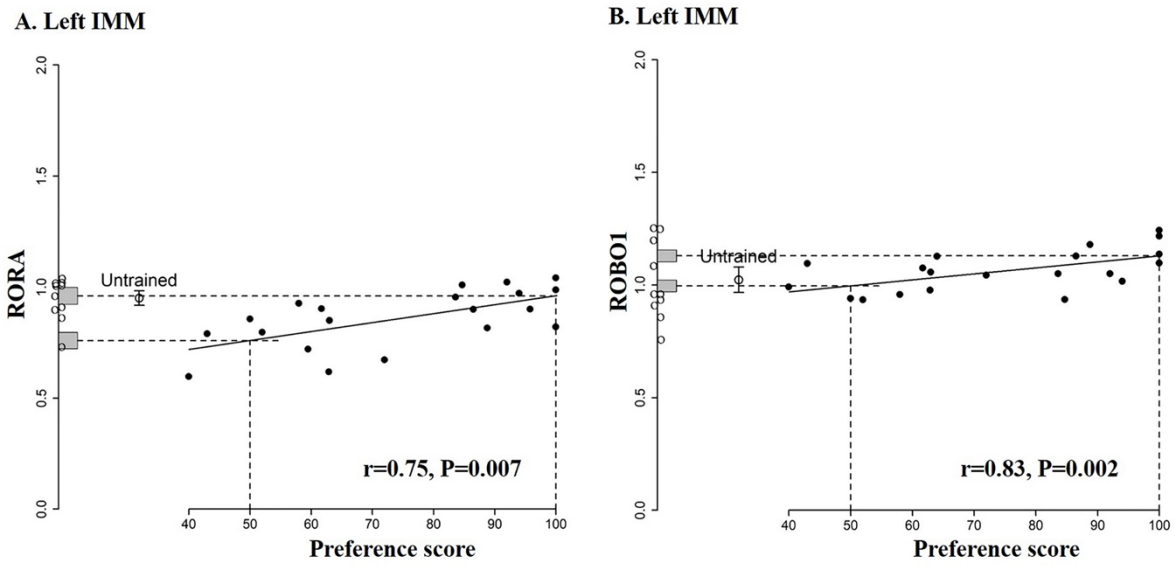

Figure 4. Standardized relative amount of RORA (A) and ROBO1 (B) proteins in the left IMM plotted against preference score. Filled circles, trained chicks. Open circle with error bars, mean level in untrained controls ± SEM; individual values for untrained chicks are shown as open circles along the y-axis. Vertical dashed lines, 'no preference' score 50 (approach to training and alternative stimuli equal/no learning) and 'maximum preference' score 100 (approach to training stimulus only/strong learning); horizontal dashed lines, y intercepts for 'no preference' and 'maximum preference'. Gray bars on y-axis, ± SE of intercept. For both proteins correlations are significant

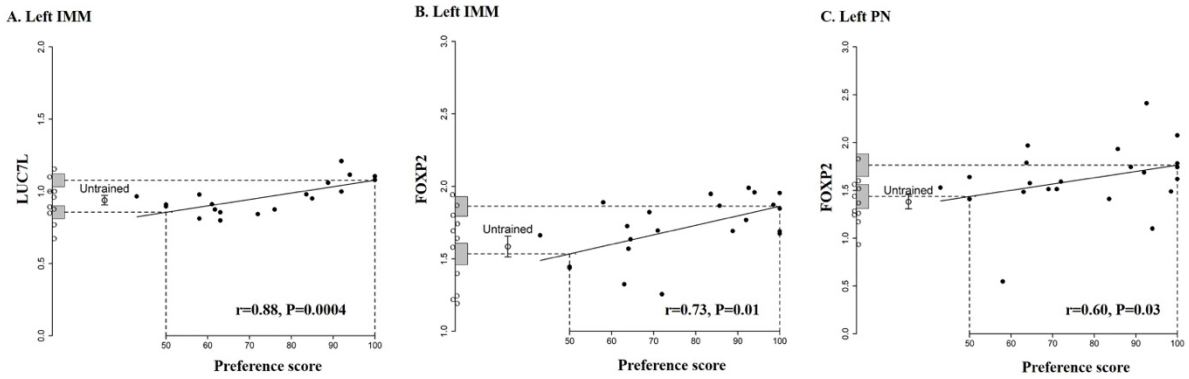

Figure 5. Standardized relative amount of LUC7L (A-left IMM) and FOXP2 (B-left IMM, C-left PPN) plotted against preference score. Filled circles, trained chicks. Open circle with error bars, mean level in untrained controls ± SEM; individual values for untrained chicks are shown as open circles along the y-axis. Vertical dashed lines, 'no preference' score 50 (approach to training and alternative stimuli equal/no learning) and 'maximum preference' score 100 (approach to training stimulus only/strong learning); horizontal dashed lines, y intercepts for 'no preference' and 'maximum preference'. Gray bars on y-axis, ± SE of intercept. For all three cases correlations are significant.

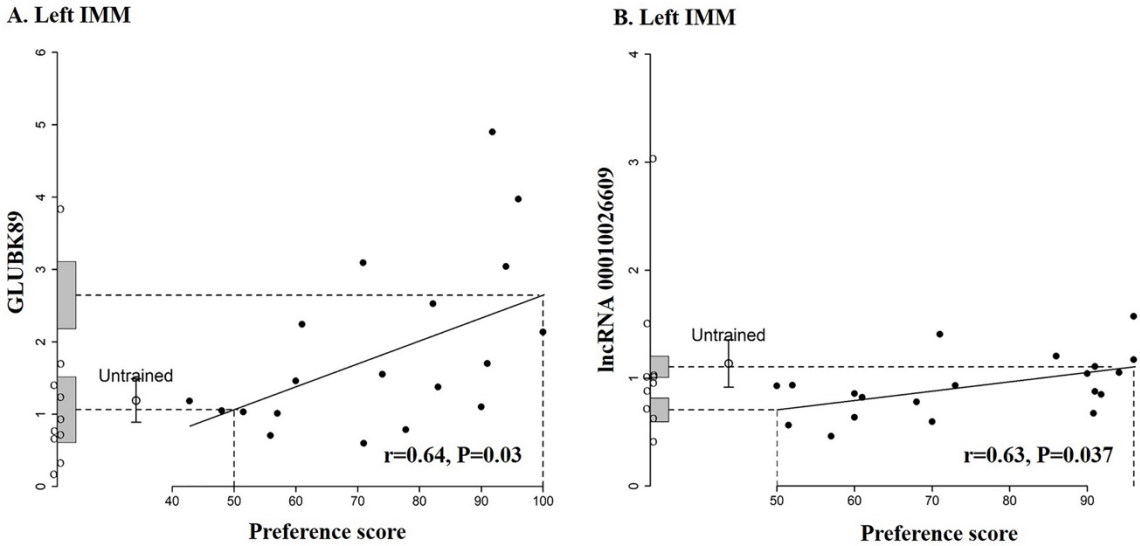

Figure 6. Standardized relative amount of GLUBK89 (A) and lncRNA 6609 (B) in the left IMM plotted against preference score. Filled circles, trained chicks. Open circle with error bars, mean level in untrained controls ± SEM; individual values for untrained chicks are shown as open circles along the y-axis. Vertical dashed lines, 'no preference' score 50 (approach to training and alternative stimuli equal/no learning) and 'maximum preference' score 100 (approach to training stimulus only/strong learning); horizontal dashed lines, y intercepts for 'no preference' and 'maximum preference'. Gray bars on y-axis, ± SE of intercept. For both lncRNA the correlations are significant (P=0.03 and P=0.037 respectively).

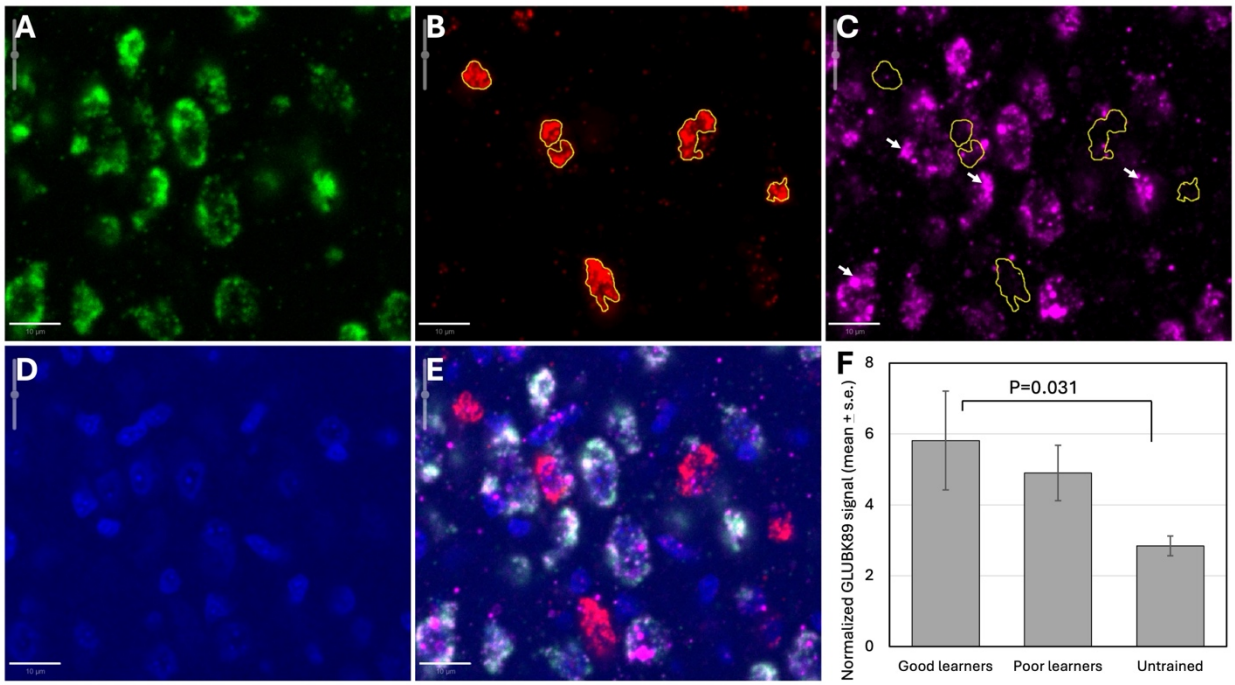

Figure 7. Representative confocal FISH imaging of the left IMM region in the brain of a trained chick. (A) Glutamatergic neurons expressing SLC17A6 are labeled with a green fluorescent marker. (B) GABAergic neurons expressing GAD2 are labeled with a red fluorescent marker, with their boundaries outlined in yellow. (C) GLUBK89 transcripts (purple) are primarily localized within the nuclei of glutamatergic neurons (indicated by arrows), rather than GABAergic neurons (outlined in yellow). (D) Nuclei are stained with DAPI. (E) Merged image. Bar=10 µm. (F) GLUBK89 expression, normalized by PPI, in the left IMM across groups.