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], Jesper Tegner[1,5,6,7], Leena Ali Ibrahim[1], Brian J. McCabe[8], Zaza Khuchua[2], David Gomez Cabrero[1,9,#], and Revaz Solomonia[2,3,#]

[1]Biological and Environmental Sciences and Engineering Division (BESE), 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, Swede

[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 identified over 30 cell clusters with distinct transcriptional differences putatively linked to memory formation, nearly half of them in long non-coding RNAs (lncRNAs). Expression levels of two lncRNAs and four protein-coding genes *FOXP2*, *RORA*, *LUC7L*, and *ROBO1* correlate with memory strength, reflecting either innate learning potential or imprinting experience. Notably, the brain- and avian-specific lncRNA *ENSGALG00010007489* is enriched in the nuclei of specific glutamatergic clusters. These findings offer the first single-cell resolution map of transcriptional changes underlying memory formation in the avian brain.

Introduction

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, even in brain regions known to be a memory store.

One of most suitable phenomena for memory studies is visual imprinting in chicks. Visual imprinting is an experience-dependent form of learning that occurs during the early-life period, in which a young animal rapidly forms a strong and lasting attachment to a visual stimulus after brief exposure [1–3]. 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 occurs 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 exhibit 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[1–3].
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[1,2]. Strong evidence suggests that the *intermediate medial mesopallium* (IMM), previously known as the intermediate and medial *hyperstriatum ventrale* (IMHV)[4], is a storage site for information about the imprinting stimulus. By applying regression analysis, one can assess whether molecular changes in the IMM (as well as in other brain regions that are not involved in imprinting) correlate with a behavioral measure of learning strength, such as the preference score, rather than with unrelated factors. Established criteria for identifying learning-related changes using linear modeling and variance partitioning have been documented[5]. 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)[5].

Visual imprinting in chicks is associated with hemispheric asymmetry at behavioral, molecular, and electrophysiological levels[1–3,6]. Brain lesion experiments implicate both left

79 and right IMM in memory and also indicate that the right IMM is necessary for storage in a
80 region termed S' outside the IMM[1,6]. Learning-related molecular changes are typically more
81 strongly expressed in the left IMM, especially 24 h after training[1–3,5,7–12]. For this reason, the
82 present study focuses on the left IMM. We have previously demonstrated a progression of
83 learning-related molecular changes in the IMM over the 24 h after training[3]. At least some of
84 these changes are not uniformly distributed among IMM cells and occur at different times
85 after training[13–17]. For example, learning-related up-regulation of Fos protein in parvalbumin-
86 containing, but not calbindin-containing GABA neurons, occurs 1 h after the end of training
87 in the IMM but not subsequently[17]. In contrast, a learning-related change in NMDA receptor
88 binding occurs several hours later[13,14]. To further understand the molecular mechanisms of
89 memory, it is important to know which molecular changes are occurring and in which cell
90 types they take place.

91 Single cells/nuclei omics technologies have revolutionized molecular biology, allowing
92 investigation of heterogeneity of living tissues at the cellular level[18]. Publicly available
93 atlases[19,20] of brain cell transcriptomics profiles, together with appropriate technology, enable
94 detailed study of neural plasticity and memory at the cellular level[21,22], increasing
95 understanding of fundamental mechanisms and thus establishing a rational basis for the
96 successful treatment of neurological disorders[23,24]. No study has investigated imprinting-
97 associated molecular changes in the chick IMM with single-cell RNA-seq technology.

98 In this study, we compared single-nuclei transcriptome profiles of the left IMM between
99 untrained chicks and good learners – those that achieved high preference scores during
100 training (see Methods). We collected tissue 24 hours after the end of the training, a time point
101 when numerous long-term, learning-dependent changes occur in the IMM[3,5,7–12,25]. Our
102 findings enable us to: (i) classify IMM cells based on transcriptomic profiles and (ii) identify
103 cell-type-specific gene expression differences between good learners and untrained chicks.

104 We also performed quantitative RT-PCR (qRT-PCR) and immunoblotting validation
105 experiments to determine whether the observed differences were specifically related to
106 memory rather than to side effects of the training procedure. One long non-coding RNA
107 (lncRNA), ENSGALG00010007489 (GLUBK89), and three genes – LUC7L, FOXP2, and
108 RORA showed strong correlations between preference score and their expression levels,
109 indicating changes directly attributable to learning during training. In contrast, lncRNA
110 ENSGALG00010026609 (lncRNA6609) and ROBO1 exhibited associations that likely
111 reflect learning capacity, suggesting a predisposition to learn that existed prior to training.

112 Among the molecules that showed significant correlations in the left IMM, only FOXP2 also
113 displayed a correlation in the posterior pole of the nidopallium (PPN).

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

120

Results

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

We followed an established protocol for training and testing[16] (see Methods) to assess imprinting behavior in chicks. In short, 24 hours after hatching, chicks were placed in a running wheel and exposed to a training stimulus 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) using single nuclei RNA-seq technology (see Methods), obtaining transcriptomic profiles for a total of 54,763 cells (Supplementary Table S1 and Figure S1), which were grouped into 36 clusters. By analyzing those clusters, we identified four major cell-types using well-established marker genes[26–28] (Figure 1b,c; Figure S2): GABAergic neurons (n=8,999; GAD1, GAD2, SLC6A1, SLC32A1), glutamatergic neurons (n=37,582; ARPP21, SLC17A6, SV2B, SATB2), astrocytes (n=5,571; AQP4, GLI3, SLC39A12), and oligodendrocytes (n=2,251; OLIG2 and SOX10). Additionally, two clusters expressed markers characteristic of vasculature (cluster C32, n=284, INPP5D, FLT1, and ADGRL4) and immune system (cluster C35, n=76, B2M, GPX1, SLC4A1). Finally, we sought to identify subpopulations of cells within these major groups.

There is evidence from work on reptiles that major GABAergic neuron types in amniotes have a common ancestor[29]. Indeed, we identified cellular subpopulations of neurons expressing known markers of GABAergic subtypes (Supplementary Table S2). LAMP5 inhibitory neurons form a distinct cluster, clearly separate from the other cells (Figure 1d). Other notable inhibitory subtypes include CHODL, VIPR2, and MEIS expressing neurons. We also identified a subgroup of PROX1+ VIP neurons[30], as well as somatostatin cells characterized by CPNE8 and NPY expression[31] (Figure 1d and Supplementary Figures S3 – S18).

Glutamatergic neurons are typically classified according to the cortical layer they reside in and their projection patterns[26]. This may not apply to excitatory neurons located within the IMM. We nevertheless sought to identify similarities between the expression profiles of known excitatory neuron types and IMM glutamatergic neurons, which may shed light on their functional role.

A previous study showed that a large portion of glutamatergic (VGLUT2+) neurons in the chick mesopallium co-express two markers whose co-presence is specific to intra-telencephalic neurons in mammals, namely SATB2 and BCL11A[32]. We indeed identify co-expression of these two markers in 20,714 (37.8%) of the IMM excitatory neurons, forming most cells in several glutamatergic clusters (Supplementary Table S3 and Supplementary Figure S19). This finding supports the hypothesis that a portion of IMM glutamatergic neurons resemble intra-telencephalic mammal excitatory neurons.

Furthermore, we applied self-assembling manifold mapping[33] (SAMap), an algorithm that exploits similarities between protein sequences for matching scRNA-seq profiles across different species (Methods). We used a single cell transcriptomics atlas of the mouse cortex and hippocampus as a reference[26]. This analysis showed that some IMM excitatory clusters resemble glutamatergic neurons found in deep layers (L4/5/6) of the mouse cortex (Supplementary Table S2 and Supplementary Table S4).

Within the non-neural populations, one of the astrocyte clusters presents a distinct signature of markers, including the LRGUK and NEK10 genes, which are associated with “cilium movement involved in cell motility” and related gene ontology biological processes (Supplementary Table S5). We hypothesize that this astrocyte subgroup may preserve mobility functions characteristic of immature astrocytes.

These results provide the first detailed classification of the transcriptomic profiles of cells composing the IMM, a brain region strongly implicated in imprinting and passive avoidance memory in chicks[1].

Good learners and untrained chicks exhibit distinct transcriptomics profiles.

We performed a differential expression analysis on pseudo-bulk samples (details in Methods). Cells from the good learner and untrained chicks were present in all clusters (Figure 2a). Three clusters (C16, C26, and C35) showed a statistically significant disproportion (ratio > 1.5, FDR < 0.05) in terms of number of cells originating from good

learner and untrained chicks (see Methods and Supplementary Figure S20). These three clusters and eight additional ones (C21, C22, C29 – C34) also count fewer than 50 cells for their least represented sample. We do not report differential analysis results for these 11 clusters, to avoid any misleading findings due to imbalance in cell proportions. In contrast, no major cell types presented a statistically significant cell disproportion between the two experimental group, and no sample contributed fewer than 50 cells.

Several differentially expressed transcripts were identified in all major cell types (log2 fold change > 1 in absolute value, FDR < 0.05, Supplementary File S1). Interestingly, roughly half of all changes were in long non-coding transcripts (lncRNAs, Figure 2b), a statistically significant higher proportion ($p < 10^{-4}$) than the one observed in the tested transcripts (see Methods and Supplementary Table S6).

Numerous other studies have identified molecular changes in the chick IMM associated with imprinting[3,5,7–17,25]. Our results recapitulate previous findings, and associate them with specific cell types or clusters: the neural cell adhesion molecule 2 (NCAM2) gene[12], is overexpressed in good learners within Cluster 0, the predominant glutamatergic group (log2FC = 0.77, FDR = 0.056). The GABRG3 and GABBR2 genes, encoding GABA-A and GABA-B receptors[34], exhibited increased expression in good learners across various clusters, while the CHRM2 and CHRM3 genes encoding muscarinic receptors[35–37], were differentially expressed in glutamatergic neurons (Supplementary File S1).

We further examined whether the observed changes arise jointly across cell types or subtypes (clusters). Differential expression results exhibited a high correlation between log2 fold changes across all major cell types, (minimum Pearson correlation coefficient 0.63, $p < 0.01$), however the association between excitatory and inhibitory neurons, and between non-neuronal cells, was even more marked (Pearson correlation coefficient 0.85 and 0.83, respectively, $p < 0.01$ in both cases, Supplementary Figure S21). This pattern persisted at the subtype level (Figure 2c), suggesting that transcriptomic changes occur in a coordinated manner.

We also investigated whether transcriptomic changes can be cluster-specific. Supplementary Table S7 lists all transcripts (85 in total) that are differentially expressed in individual clusters but not in their corresponding cell types, indeed supporting the hypothesis that changes linked to imprinting may occur in isolated clusters.

Finally, Figure 2d reports the log2 fold-change for the five transcripts with the greatest expression increases across all glutamatergic cells (base average expression > 100, FDR < 0.05). When their expression is examined at the cluster level, they exhibit statistically significant changes (FDR < 0.05) only in specific glutamatergic clusters (see Figure 2d), indicating that these clusters are the primary drivers for the overexpression of these transcripts in the general excitatory neuron population.

Taken together, our findings (a) offer insights into the particular cell types underlying the imprinting-associated molecular alterations reported in prior studies[12,34–37], and (b) demonstrate that although imprinting-related transcriptomic changes are broadly coordinated across cell types and clusters, certain gene expression changes are specific to or predominantly driven by individual clusters.

Imprinting is associated with biological processes related to neural plasticity

Next, we aimed to characterize through functional analysis those biological processes (BPs) upregulated in good learners across different cell types (Figure 3a, Supplementary File S2) and clusters (Figure S22). BPs related to cell-to-cell adhesion molecules were upregulated across all major cell types (Figure 3a), further corroborating previous findings[12]. Processes related to synapse assembly, as well as axon and neuron projection guidance, were primarily upregulated in both excitatory and inhibitory neurons (Figure 3a, 3b). In contrast, developmental and morphogenesis-related processes were mainly enriched in non-neuronal cells (Figure 3a). Notably, BPs associated with higher cognitive functions—GO:0007612 ("learning"), GO:0008306 ("associative memory"), and GO:0007613 ("memory")—were upregulated in both glutamatergic and GABAergic neurons (Figure 3c, FDR < 0.1). Enrichment results at the cluster level showed similar trends.

Finally, we aimed to better understand the functions of the numerous lncRNAs differentially expressed between good learners and untrained chicks. To do so, we clustered all transcripts—including both protein-coding genes and lncRNAs – using a methodology designed for single-cell/nuclei RNA-seq (hdWGCNA[38], see Methods). This analysis generated cell-type-specific transcript modules, encompassing 2,846 transcripts across 17 modules for glutamatergic neurons, 921 transcripts in 6 modules for GABAergic neurons,

245 3,476 transcripts in 6 modules for astrocytes, and 50 transcripts in a single module for
246 oligodendrocytes (see Supplementary Figure S23 and Supplementary File S3).

247 We then performed enrichment analysis on each module, considering only genes included in
248 Gene Ontology BPs. This approach allows to assign biological functions to each module and,
249 by extension, to the differentially expressed lncRNAs within them[39,40].

250 Following this procedure, the lncRNAs differentially expressed in glutamatergic cells were
251 mostly associated with regulation of dendrite development, synapse assembly, neuron
252 migration, and regulation of dendrite morphogenesis. Regulation of monoatomic ion and
253 cation transmembrane transport, as well as synapse assembly were the BPs most associated
254 with lncRNAs showing significant changes in GABAergic neurons. Finally, axoneme
255 assembly and various BPs related to cilium movement and motility were associated, among
256 others, with lncRNAs differentially expressed in astrocytes (for details see Supplementary
257 File S3 and S4).

258 Together, these findings highlight that both protein-coding genes and lncRNAs differentially
259 expressed in good learners converge on cell-type-specific biological processes, many of
260 which are linked to synaptic function, and neural development.

261 Selected differential transcripts exhibit specific learning-related 262 changes.

263 Good learners show learning-induced molecular changes in the IMM, but they also differ
264 from untrained chicks in other aspects, such as exposure to visual stimuli and levels of motor
265 activity. Therefore, genes with prominent transcriptomic changes were further examined in
266 experiments designed to isolate their specific role in memory, controlling for potential
267 confounding effects unrelated to learning.

268 LUC7L, FOXP2, ROBO1, and RORA in learning and memory.

269 Long-term memory formation involves specific and sustained transcriptional changes,
270 mediated by transcription factors and other nuclear processes such as alternative splicing[41,42].
271 Based on this, we selected two transcription factors—FOXP2 and RORA—and one
272 alternative splicing-associated protein, LUC7L, for further study. Additionally, axon
273 guidance molecules and their receptors are known to contribute to synaptic remodeling and
274 long-term memory[43,44]. Therefore, the fourth differentially expressed gene chosen for
275 validation was ROBO1, an axon guidance receptor.

For all proteins the molecular weights of the stained protein bands corresponded to expected values: antibodies against LUC7L reacted with a protein band of molecular weight of 80 kDa, antibodies against FOXP2 reacted with a protein band of 44 kDa, antibodies against ROBO1 with a protein band of 250 kDa whereas antibodies against RORA with a protein band of 60 kDa. Moreover, for all four proteins, plotting the optical densities of the internal standards against the amount of protein (15, 30, 45 and 60 μg protein of IMM homogenate fraction) showed a virtually perfect fit to a straight line (Supplementary Figure S24).

The results of the immunoblotting experiments show that the levels of all four proteins were significantly correlated ($p < 0.05$) with a preference score solely in the left IMM, except for FOXP2, for which a significant correlation was found also in the left PPN (Figures 4 and 5, Supplementary Tables S8 – S11, Supplementary Figures S25 – S28). 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 an additional, suppressive, effect of training unrelated to learning. ROBO1's association with memory formation seems to be due to a predisposition, as 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. Also, the residual variance from the regression plot is significantly lower than the variance in untrained animals ($P < 0.0005$).

Selected lncRNAs in learning and memory.

Two lncRNAs were selected for validation. The first, ENSGALG00010007489 was the most differentially expressed in glutamatergic neurons (see Supplementary File S1). Notably, this differential expression was predominantly driven by a single cluster, C2, suggesting a highly localized and potentially cluster-specific function.

In contrast, the second selected lncRNA, ENSGALG00010026609 (hereafter referred as LNCRNA6609), exhibits a statistically significant difference in expression between good learners and untrained chicks (FDR < 0.05) in all major cell types.

ENSGALG00010007489 is a lncRNA overexpressed in good learners specifically within cluster 2, a sub-population of glutamatergic neurons. It is also highly expressed in at least one additional glutamatergic cluster (cluster 30) and one GABAergic cluster (cluster 29, identified as LAMP5 neurons). According to SAMap analysis, cluster 2 closely resembles L6

308 CT CTX neurons and is distinguished by high expression of the KCNMB2 gene.
309 Remarkably, 91% of cells in cluster 2 express KCNMB2, compared to only 21% in other
310 glutamatergic neurons, representing a statistically significant overexpression (FDR < 0.001)
311 relative to other glutamatergic clusters.

312 KCNMB2 encodes the β2 auxiliary subunit of large-conductance calcium- and voltage-
313 activated potassium (BK) channels, which play key roles in regulating synaptic plasticity,
314 neuronal excitability, and memory[45]. Based on these functional associations, we hereafter
315 refer to ENSGALG00010007489 as GLUBK89.

316 In addition to the BK2 channels (KCNMB2) and the GLUBK89 marker, the expression of
317 several other genes distinguishes cluster 2 glutamatergic neurons from other glutamatergic
318 populations (Supplementary File S5). Among the most significantly differentially expressed
319 genes are MDFIC2, the transcription factors EST1 and ATOH8, and the contractile protein
320 ACTN2. These gene expression differences contribute to alterations in multiple biological
321 processes (Supplementary File S5). Notably, the most significantly affected biological
322 process in cluster-2 neurons compared to other glutamatergic cells is regulation of
323 endocytosis. This finding aligns with our previous work and that of others, which reported
324 learning-related upregulation of proteins involved in endocytosis[11,46].

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

329 The results, presented in Figure 6A and Supplementary Table S12, show a significant
330 correlation between GLUBK89 expression and preference score (Pearson's correlation =
331 0.64, p = 0.03, Fig. 6A). The intercept at maximum preference is significantly higher than the
332 mean expression in untrained chicks (p = 0.019, Supplementary Table S12). In contrast, the
333 intercept at 50% preference (no learning) does not differ significantly from the untrained
334 mean. Additionally, the residual variance from the regression is not significantly different
335 from the variance in untrained chicks (Supplementary Table S12).

336 These findings indicate that the observed positive correlation between GLUBK89 expression
337 and preference score directly results from learning during training (see Methods). No

338 significant correlations were found in other brain regions, suggesting a localized effect of
339 learning in the left IMM (Supplementary Table S12, and Supplementary Figure S29).

340 The correlation between the standardized amount of LNCRNA6609 and the preference score
341 is significant ($r = 0.63$, $P = 0.037$; Figure 6B, Supplementary Table S13). However, the
342 differences between the mean untrained value and the intercepts for preference scores 50 and
343 100 were not significantly different. Together with the residual variance from the regression
344 line being significantly lower than the variance of untrained chicks ($p = 0.003$,
345 Supplementary Table S13), these results suggest that chicks with higher expression of
346 LNCRNA6609 expression may be associated with an inherent predisposition for learning
347 rather than an effect of imprinting training (Please see the Methods subchapter "Interpretation
348 of a correlation between a molecular quantity and preference score"). No significant learning-
349 related changes were evident in other brain regions (Supplementary Table S13 and
350 Supplementary Figure S30).

351 GLUBK89 is a brain and nucleus specific transcript.

352 To determine the subcellular localization of GLUBK89 and LNCRNA6609, IMM tissue was
353 fractionated into cytoplasmic and nuclear components, followed by RNA extraction and
354 qRT-PCR analysis. GLUBK89 levels were found to be over 100-fold higher in the nuclear
355 fraction, indicating predominant nuclear localization. LNCRNA6609 is present both in the
356 nuclear and cytoplasmic fractions (Supplementary Figure S31).

357 Finally, the expression of both lncRNAs was analyzed in chick heart, muscle, liver, lung and
358 brain using quantitative RT-PCR. GLUBK89 expression was detected only in brain. In
359 contrast, LNCRNA6609 is detected across all studied tissues, with the highest expression
360 levels observed in the lung and liver (Supplementary Figure S32).

361 GLUBK89 is avian-specific transcript while LNCRNA6609 shows broad
362 sequence conservation in mammals.

363 To further explore the biological roles of these lncRNAs, we examined their evolutionary
364 conservation by identifying potential orthologs of GLUBK89 and LNCRNA6609 in
365 mammals. A search in the RNA family database RNAcentral[47] revealed that GLUBK89 is
366 conserved among avian species but lacks identifiable orthologs in mammals.

367 In contrast, LNCRNA6609 exhibits sequence similarity with 89 regions in the human
368 genome and 30 in the mouse genome. It shares over 70% sequence identity with several

369 human lncRNAs of unknown function, including HSALNT0096019 and SH3TC2-DT.
370 Similarly, the mouse genome shares over 65% sequence identity with NONMMUT029999.2,
371 NONMMUT009705.2, and NONMMUT032357.2, none of which encode proteins or have
372 well-defined functions.

373 FISH analysis confirms GLUBK89 localization in glutamatergic 374 neurons and its association with memory.

375 We further validated the cellular localization of GLUBK89 using triple fluorescence *in situ*
376 hybridization (FISH). Specifically, we selected a good learner (preference score > 85%), a
377 poor learner (preference score < 65%), and an untrained chick from each of four different
378 batches, totaling 12 chicks. Six sections were collected from each brain. Five sections were
379 triple-stained for GLUBK89, PPIB (a housekeeping gene), and SLC17A6 (a glutamatergic
380 neuron marker). In contrast, in the remaining section, PPIB was replaced with GAD2, a
381 GABAergic neuron marker (see Supplementary Figure S33 for the experimental design).
382 Regions corresponding to the left and right IMM and nidopallium were manually identified
383 (see Methods for details). RNAscope images were processed using QuPath[48] to obtain
384 average fluorescence values for each channel within each detected nucleus (see Methods for
385 additional details).

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

390 To assess whether GLUBK89 expression in glutamatergic neurons is associated with memory
391 of visual imprinting, we quantitatively analyzed sections stained with SLC17A6, GAD2, and
392 GLUBK89 (Supplementary Figure S33) in different experimental groups of chicks. Nuclei of
393 putative glutamatergic and GABAergic cells were selected as having above-threshold
394 SLC17A6 (>1700) and GAD2 (>650) fluorescence levels, respectively. GLUBK89
395 fluorescence levels were consistently and significantly (two tailed t-test $p < 0.001$) higher in
396 glutamatergic than in GABAergic nuclei (Supplementary Figure S34).

397 In the IMM, there was a significant effect of experimental group ($F_{2,6} = 6.42$, $P = 0.032$), no
398 significant effect of side and no significant interaction between experimental group and side
399 were observed ($P > 0.25$). The mean value for good learners was significantly greater than
400 that of untrained chicks ($P = 0.031$, adjusted by Tukey's method for a family of three means).

401 No other pairs of means differed significantly. The rank order of mean normalized values for
402 lncRNA in SLC-positive nuclei was good learners > poor learners > untrained chicks (5.49,
403 4.76 and 2.68, respectively). This rank order is consistent with the result shown in Figure 6A,
404 which was obtained with a larger sample and thus greater statistical precision.

405 The same analysis of variance model applied to data from the nidopallium showed no
406 significant effects (Supplementary Figure S35).

407 Thus, the triple FISH experiments unequivocally confirm: (i) the localization of GLUBK89
408 in specific cluster of glutamatergic cells and (ii) increase associated with memory.

409

410

411

412 # Discussion

413 The present study, for the first time, (i) identifies a homology between the avian intermediate
414 medial mesopallium (IMM)—a brain region critical for learning and memory—and
415 mammalian deep cortical layers; (ii) provides compelling evidence for a previously
416 unrecognized role of long non-coding RNAs (lncRNAs) in long-term memory; and (iii)
417 uncovers a molecular signature of visual imprinting in chicks at single-cell resolution.

418 **IMM homology with mammalian brain structures and cell type identification.** The
419 intermediate medial mesopallium (IMM) is a key brain region involved in memory encoding
420 during the visual imprinting process in chicks and is also important in passive avoidance
421 learning[49]. Imprinting involves recognition of a sensory stimulus and analysis of its neural
422 basis in chicks potentially offers insights into the mechanism of recognition memory in many
423 species. Modification of behavior associated with visual imprinting in chicks has been
424 proposed as a model for autistic spectrum disorder[50]. In spite of IMM's importance, the
425 widespread occurrence of imprinting in mammals and the functional significance of
426 recognition memory, the correspondence of the IMM to regions in the mammalian brain
427 remains unclear. Horn proposed that the IMM may be homologous to the mammalian
428 association cortex[51]. Subsequent studies have supported and expanded upon this idea, with
429 some suggesting that the IMM is analogous to neocortical layers 2 and 3 based on gene
430 expression patterns[52,53]. Other research, examining the expression of six transcription factors,
431 has proposed that excitatory neurons in the mesopallium resemble mammalian neocortical
432 neurons, typically located in layers 2, 3, 5, and 6 [32]. To further investigate this potential
433 homology, we applied single-nucleus transcriptomics to determine whether any excitatory
434 neuron clusters in the chick IMM exhibit similarities to known excitatory neuron types in the
435 mouse brain. Our analysis (Supplementary Table S2) reveals that 7 out of 19 excitatory
436 clusters in the chick IMM display transcriptomic profiles closely resembling those of neurons
437 in layers 4, 5, and 6 of the mouse cortex. These findings are consistent with a recent study,
438 which also demonstrated similarities between mesopallium glutamatergic neurons and deep-
439 layer cortical neurons in mice using scRNA-seq, scATAC-seq, and regulatory data[54]. For the
440 remaining 12 excitatory clusters, we were unable to establish a clear mammalian counterpart,
441 suggesting these populations may be specific to the avian forebrain.

442 **Transcriptomic fingerprints of learning in IMM.** Single-nucleus transcriptomic analysis of
443 the IMM in GL and UNT chicks identified 319 differentially expressed genes in

444 glutamatergic neurons of the left IMM (logFC ≥ 1, FDR ≤ 0.05). Notably, 180 of these
445 transcripts were lncRNAs, indicating their potential functional role in imprinting. While
446 previously dismissed as transcriptional noise, lncRNAs are now recognized as critical
447 regulators in numerous biological processes, including chromatin remodeling, transcriptional
448 control, and post-transcriptional regulation [55–57]. These molecules function as molecular
449 guides, scaffolds, and decoys, modulating chromatin architecture and gene expression by
450 recruiting chromatin-modifying proteins to specific genomic loci. Their roles are essential for
451 maintaining cellular identity and proper genome organization.

452 Based on the involvement of differentially expressed genes in neuronal processes and
453 synapse formation, we selected four protein-coding genes—LUC7L, FOXP2, ROBO1, and
454 RORA for downstream validation analysis. In addition, two lncRNAs, GLUBK89 and
455 lncRNA6609, which were significantly upregulated in the left IMM of GL chicks, were also
456 selected for validation.

457 Behavioral experiments assessing learning strength, combined with molecular analyses and
458 criteria formulated for learning-relatedness, revealed that GLUBK89, LUC7L, FOXP2, and
459 RORA are modulated and increased by learning. In contrast, lncRNA6609 and FOXP2
460 appear to be associated with a predisposition for effective learning. Further investigation of
461 GLUBK89 using quantitative multi-probe in situ hybridization, supported by RNA-seq data,
462 demonstrated that its learning-induced up-regulation occurs specifically within the nuclei of
463 glutamatergic neurons.

464 GLUBK89 lncRNA is uniquely positioned among the hundreds of differentially expressed
465 non-coding transcripts following imprinting. Although primarily expressed in a single cluster,
466 GLUBK89 ranks first in terms of statistical significance in the pseudo-bulk analysis
467 conducted across all glutamatergic neurons (Supplementary File S1).

468 Quantitative RT-PCR analysis of GLUBK89 revealed a significant correlation only in the left
469 IMM, driven by the effects of training rather than a predisposition to learn. While a positive
470 correlation trend was observed in the right IMM, it was negative in both the left and right
471 PPN.

472 Further investigations highlighted the importance of GLUBK89: (i) it is exclusively
473 expressed in the brain, (ii) localized in the nucleus, and (iii) specific to avian species. FISH
474 analysis confirmed its nuclear localization within a small subset of glutamatergic neurons,

475 while quantitative measurements supported the qRT-PCR findings, demonstrating memory-
476 related upregulation of GLUBK89.

477 Based on these findings and co-expression with BK channels to be known by their
478 involvement in synaptic plasticity[58,59], we propose that GLUBK89 plays a role in the
479 coordinated regulation of gene expression essential for learning and memory in a specific
480 cluster of glutamatergic neurons, a cluster characterized by both a BK2 channel marker
481 (KCNMB2) and increased expression of genes connected to endocytosis regulation.
482 lncRNA6609, another non-coding transcript, is more ubiquitously expressed than GLUBK89
483 and is significantly elevated in glutamatergic neurons, astrocytes, and oligodendrocytes of
484 good learner chicks compared to untrained controls. Unlike GLUBK89, lncRNA6609
485 exhibits a broader tissue distribution beyond the brain and is localized also in the cytoplasm.
486 Its levels significantly correlate with the preference score exclusively in the left IMM.
487 However, unlike GLUBK89, this correlation is not influenced by training but reflects an
488 inherent capacity or predisposition for learning.

489 The presence of lncRNA6609 in the cytoplasm raises the intriguing possibility that it acts
490 through (i) its translation and/or (ii) its role as a scaffold for protein complexes involved in
491 specific cellular functions. Similar complexes were initially identified for c-Myc-responsive
492 long noncoding RNA, which was found to associate with four glycolytic payoff phase
493 enzymes[60].

494 LUC7-like proteins are integral components of U1 small nuclear ribonucleoproteins and play
495 a key role in regulating alternative splicing. In vertebrates, this protein family includes three
496 paralogs: LUC7L, LUC7L2, and LUC7L3[61,62]. Our snRNA-Seq analysis detected the
497 expression of all three paralogs, but only LUC7L exhibited significant differences between
498 good learners and untrained chicks. Specifically, LUC7L expression was markedly higher in
499 GL chicks compared to the UNT group in nearly two-thirds of glutamatergic neuron clusters
500 (Clusters: 0, 2, 3, 4, 5, 8, 9, 10, 13, 17, 18, and 24), with a borderline significant increase
501 observed in oligodendrocytes. No significant changes were found for LUC7L2 or LUC7L3.
502 Furthermore, we demonstrated that LUC7L protein levels positively correlate with learning
503 strength specifically in the left IMM, while no such correlation was observed in other brain
504 regions. This relationship is driven by the learning process itself rather than an inherent
505 predisposition to learn.

506 LUC7L, LUC7L2, and LUC7L3 differentially regulate two major 5′ splice sites (5′SS)
507 classes. The learning-related upregulation of specific splicing factor LUC7L3 suggests
508 increased defined splice events during memory formation. To our knowledge, this is the first
509 demonstration of memory-related changes in the LUC7L protein. However, recent findings
510 have shown that alternative splicing is essential for synapse development and memory
511 formation[63].

512 The snRNA-seq experiment revealed an upregulation of retinoid-related orphan receptor-
513 alpha (RORA) in astrocytes and glutamatergic neurons of good learners compared to
514 untrained chicks. Quantitative analysis of RORA protein levels showed a significant positive
515 correlation with preference score exclusively in the left IMM. This correlation was driven by
516 the learning process rather than an inherent learning capacity.

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

526 RORA is a nuclear receptor family member, functions as a constitutive transcriptional
527 activator[64] and is predicted to regulate over 2,700 genes[65]. These include genes associated
528 with autism spectrum disorders, neuronal differentiation, adhesion, synaptogenesis, synaptic
529 transmission, cognition, memory, and spatial learning[66].

530 Recent studies utilizing single-nucleus and spatial transcriptomics have identified neurons
531 resilient to neurodegeneration in the human brain during Alzheimer's disease progression.
532 Some of these resilient neurons are located in layer 4 of the association cortex and, early in
533 the disease, exhibit upregulation of genes involved in synapse maintenance, synaptic
534 plasticity, calcium homeostasis, and neuroprotection, including RORA and NCAM2[67].

535 Notably, our data suggest that IMM cells in chicks share similarities with cortical layers 4/5/6
536 in mammals. With learning, these cells also show upregulation of RORA and NCAM2. To

537 the best of our knowledge, this is the first demonstration of learning-related upregulation of
538 RORA in a chick brain region known to be involved in memory.

539 Forkhead box protein P2 (FOXP2) is a transcription factor and has been implicated mainly in
540 speech development[68]. FOXP2 acts preferentially as a transcriptional repressor, may directly
541 bind around up to 400 promoters and regulate the transcription of at least 1800 genes in
542 different tissues[69]. Our RNA-seq data reports the significant upregulation of FOXP2 in
543 astrocytes and oligodendrocytes of good learner chicks. Further immunoblotting experiments
544 revealed significant positive correlations between FOXP2 and the strength of learning and
545 memory in the left IMM and left PPN. Both changes fulfill criteria of learning-relatedness[5]
546 and in both cases the up-regulation is attributable to training and rather than a predisposition
547 to learn well. According to our RNA-seq data, upregulation occurs in oligodendrocytes and
548 we speculate that changes in the myelination process reach the left PPN region. In all cases,
549 the data obtained indicate long-term, memory-related changes of FOXP2-regulated biological
550 processes.

551 This is the first time a change satisfying our criteria for learning-relatedness[5] (see Methods)
552 has been detected in the PPN. The probability of all these criteria being satisfied by chance is
553 low, raising the possibility that consequences of learning radiate from the IMM throughout
554 the telencephalon in the hours following imprinting training. Indeed, regions outside the
555 IMM can support memory for an imprinting stimulus under the control of the right IMM, in
556 an area termed S'[6,51], and several candidates for the location of S' have been proposed[70–72].
557 Whereas the critical role played by the IMM in the learning and memory of visual imprinting
558 and passive avoidance memory is well established[2,51], there is no evidence that the PPN has
559 any specific role in imprinting, as opposed to passively reflecting changes originating in the
560 IMM.

561 ROBO1 (roundabout guidance receptor 1) binds SLIT2 guidance molecules and is involved
562 in axon outgrowth and orientation[73,74]. The expression of ROBO1 is increased in astrocytes
563 and in cluster 3 of glutamatergic neurons of the good learner chicks. The amount of ROBO1
564 protein strongly correlates with the strength of learning and memory only in the left IMM.
565 Importantly, this correlation reflects a predisposition/capacity to learn and rather than a result
566 of training.

567 This study has systematically revealed various candidate transcripts for roles in memory. Of
568 particular interest is the implication of lncRNA. Some of the candidates have been analyzed

569 in further detail, controlling for confounding influences by comparing untrained chicks with
570 trained chicks that display behavioral evidence of particular levels of memory. This
571 behavioral specificity was extended to the cellular level with GLUBK89, locating the
572 learning-related changes in GLUBK89 in the nuclei of glutamatergic neurons in the IMM.
573 Cellular analysis in conjunction with behavioral measurements is thus a powerful means of
574 elucidating neural mechanisms of memory.

575 **Limitations and further plans.** A major limitation of our study is its primarily descriptive
576 nature, lacking mechanistic insights into the specific roles of lncRNAs in memory formation.
577 The poor sequence conservation of lncRNAs across species hampers the identification of
578 mammalian orthologs and limits functional predictions. Moreover, current computational
579 models for inferring lncRNA function are constrained by the scarcity of avian-specific data [75].

580 To elucidate the biological processes and molecular mechanisms involving these lncRNAs,
581 genetically modified chicks with targeted gain- or loss-of-function perturbations would be
582 highly informative. However, generating transgenic chicken models remains technically
583 challenging. As an alternative, adenoviral-mediated delivery of lncRNA constructs could
584 provide a viable approach. Complementary strategies—such as CRISPR-based genome
585 editing or optogenetic manipulation of specific neuronal subpopulations—may also help
586 dissect the functional contributions of these lncRNAs to memory processes.

587

588

589

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[76]. 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, along with 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[77]. 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 stating studies. Eight batches of chicks were shared for Luc7L and FOXP2

experiments, 5 batches were shared between the lncRNA experiments, 8 batches were shared between ROBO1 and RORa experiments. Thus, 49 batches of chicks were used in the reported data to sum up. For the single nuclei RNA-SEQ experiments, each batch consisted of one good learner and one untrained chick, except for one batch in which two untrained chicks were included. For all other experiments in each batch, there were up to three trained chicks and a control chick from the same hatch.

Chicks were decapitated 24 h after the end of training. For single-nuclei RNA experiments only left IMM was sampled, whereas for protein and RNA measurements IMM and PPN from the left and right hemispheres were collected. The locations of the IMM and PPN are described in Solomonia et al. (2013)[16]. Details of the IMM removal procedure are provided in Davies et al. (1985)[78] and details of PPN removal in Solomonia et al. (1998)[12]. For double in-situ experiments whole forebrains were removed (see below). Samples were coded after collection and all further procedures were performed blindly.

Preparation of nuclear suspension

The nuclear fraction for RNA-SEQ was performed by a combination of methods described by Bakken et al., 2018 and Zhou et al., 2020. Frozen IMM from the left hemisphere was homogenized in a Nuclei PURE Lysis Buffer, (NUC-201, Sigma, formulated for solid tissues) supplemented with DTT, Triton X-100, protease inhibitors and 0.5% RNasin Plus RNase inhibitor (Promega). The procedure was carried out in a 1 ml Dounce homogenizer (Wheaton) using 10 strokes of the loose Dounce pestle followed by 10 strokes of the tight pestle to liberate nuclei. Homogenate was strained through a 30 µm cell strainer (Miltenyi Biotech) and centrifuged at 900 x g for 10 minutes to pellet nuclei. The supernatant was carefully removed. The pellet was resuspended in homogenisation buffer and strained through a 20 µm cell strainer. The filtered suspension was centrifuged at 900 x g for 10 minutes to pellet nuclei. Further separation of nuclei from myelin debris was carried our as described by Zhou et al, (2020). Nuclei pellets were resuspended in 500 uL Nuclei Wash buffer (1% BSA in PBS with 0.2 U/uL RNasin (Promega) and 900 uL 1.8 M Sucrose. This combined 1400 uL mixture was carefully layered on 500 uL 1.8 M sucrose and centrifuged at 13,000 x g for 45 min at 4 ^{o}C. The nuclear pellet obtained was strained through a 30 µm cell strainer, resuspended in Foetal bovine serum + DMSO (90% +10%) and stored for further 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[79] (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)[80] was used in order to identify and eliminate debris, meaning empty droplets most likely to contain ambient mRNA and no cell. DoubletFinder[81] was then applied for removing droplets possibly containing more than one single cell. Finally, cells marked as debris by CellBender[82] were also eliminated. The standard Seurat[83] 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[83], using the 2000 most variable genes. The Leiden algorithm[84] 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 the correct cell type. Major cell types were identified using the markers ARPP21, SLC17A6, SV2B, SATB2 (glutamatergic neurons), SLC6A1, GAD1, GAD2, SLC32A1 (GABAergic neurons), OLIG2, SOX10 (oligodendrocytes), and AQP4, GLI3, SLC30A12 (astrocytes). Subclasses for each cluster were identified, when possible, using the markers reported in Supplementary Table 2 and Supplementary Figures 3 – 18.

684 The SAMap algorithm[33] was then used to refine the cell classification, by identifying cell
685 types from the mouse cortex and hippocampus[26] whose transcriptomics profiles resemble the
686 expression profiles of clusters in our data. In simple terms, given two scRNA-seq datasets
687 from different species, SAMap first identifies putative homologous genes through reciprocal
688 BLAST searches. The homologous genes are then used for assessing the similarity, in terms
689 of expression profiles, between cells across the two datasets. In this way, it is possible to
690 transfer cell type annotations from one dataset to the other. In our analysis, we used SAMap
691 default settings, performing BLAST searches on the proteome sequences. Only transferred
692 annotations with SAMap score higher than 0.65 were considered.

693 Single-nuclei RNA-seq statistical analysis

694 Possible imbalances in the proportion of cells originating from good learner and untrained
695 chicks were assessed separately for major cell types and clusters, using a permutation
696 approach[85]. Results were deemed significant for permutation-based FDR < 0.05 and ratio
697 between the groups above 1.5.

698 Differential expression analysis between good learner and untrained chicks, corrected for sex,
699 was performed on pseudo-bulk profiles. Given a specific cluster or cell type, pseudo-bulk
700 profiles are derived by summing up for each gene the count values measured across the cells
701 contained in that cluster / cell type. In this way a single expression profile is obtained for
702 each sample (chick) out of several cells, simulating the same conditions obtained with bulk
703 RNA-seq (hence the name). It has been found that pseudo bulk profiles significantly reduce
704 false positive findings in differential expression analysis for single cell experiments[86]. We
705 used the method based on the negative binomial distribution implemented in the DESeq2 R
706 package[87] to perform pseudo-bulk differential expression analysis separately for each cell
707 type and cluster. Notably, DESeq2 forgo testing transcript that do not match specific criteria,
708 for example if all expression values are equal to zero[87]. The differential analysis was forgone
709 altogether if less than 50 cells were available for any chick for that specific cluster / cell type.
710 Gene set enrichment analysis[88] was performed separately for each cluster and cell type on the
711 basis of the results of the differential expression analysis, using the clusterProfiler R
712 package[89]. Gene Ontology (GO) Biological Processes (BP)[90] were used as gene sets. The GO
713 annotations for chicken have been known to be not as well curated as the annotations of other
714 organisms[91]. Thus, we converted chicken genes to their mouse orthologs and used the GO
715 mouse annotations for the enrichment analysis[92,93]. The simplifyEnrichment Bioconductor
716 package[94] was used for summarizing enrichment results across several analyses.

717 ## Clustering snRNA-seq transcript with hdWGCNA

718 The hdWGCNA Bioconductor package allows to cluster transcripts from single cell and
719 single nuclei RNA-seq experiments[38]. In short, hdWGCNA first groups individual cells
720 within metacells[95], which are supposed to represent cluster of cells sharing the same identical
721 state. A single expression profile is then computed for each metacell, an operation that
722 usually alleviates the drop-out problem common in single cell data[96]. Metacells were
723 computed within each major cell type (Astrocytes, Oligodendrocytes, GABAergic and
724 Glutamatergic neurons) and sample in our analysis. A matrix of transcript-to-transcript
725 associations is then computed through a formula involving linear correlations elevated to a
726 power coefficient. This formula is devised so that to emphasize strong correlations, and the
727 power coefficient is identified automatically in a data-driven way. Finally, a dynamic tree cut
728 algorithm[97] is used to define strongly connected transcript modules. Transcripts that do not
729 show strong connections are grouped in a default cluster, namely the "grey" module.

730 ## BLAST analysis on lncRNA

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

736 ## SDS Electrophoresis and Western Blotting

737 Brain tissue samples were rapidly homogenized in standard Tris-HC buffer containing
738 phosphatase inhibitor and protease inhibitor cocktails. Sodium dodecylsulphate (SDS)
739 solution was added up to a final concentration of 5% and the mixture was incubated at 95 ^{o}C
740 for 3min. Protein concentrations were determined in quadruplicate using a micro
741 bicinchoninic acid protein assay kit (Pierce). Aliquots containing 30 µg of protein in 30 µl of
742 standard sample buffer were subjected to SDS gel electrophoresis and Western blotting[9]. The
743 proteins were transferred onto nitrocellulose membranes and stained with 0.1% w/v Ponceau
744 S solution. The stained membranes were digitalized and analyzed for equal total protein
745 loading with LabWorks 4.0 (UVP). The nitrocellulose membranes were then treated with
746 PBS + 0.05% Tween 20 solution and a standard immunochemical staining procedure was
747 carried out with peroxidase-labelled secondary antibodies and the Super-Signal West Pico
748 Chemiluminescent substrate (Pierce, 34580). The standard immunochemical staining

749 procedure was carried out with peroxidase-labeled secondary antibodies and the Super-Signal
750 West Pico Chemiluminescent substrate (Pierce, 34580). The following primary antibodies
751 were used: anti-Luc7L (ABIN6745218, antibodies online), anti-FOXP2 (ab1307, Abcam),
752 anti-RORA (Invitrogen # PA1-812) and anti-ROBO1 (Invitrogen PA5-29917). The optical
753 densities of the target immunostained protein bands were measured with LabWorks 4.0
754 (UVP). Within each of the gel four internal protein standards (15, 30, 45, and 60 µg of total
755 protein) were used for the calibration of autoradiographs. These protein standards were from
756 IMM homogenate fraction isolated from control chick brains. We used the same proteins
757 standards in all of the immunoblotting experiments performed. Optical densities of
758 corresponding immunostained internal standard protein bands were proportional to the
759 amount of loaded protein (Supplementary Figure S28). The respective data were obtained by
760 dividing the individual band optical densities of a given sample (e.g., Luc7L) by the optical
761 density, which from the autoradiograph calibration plot corresponded to 30 µg of total protein
762 in the standard. The values obtained by this normalization procedure are termed as the
763 "relative amount" [9] of, e.g., FOXP2. There are differences between the batches resulting in
764 variation in the protein amounts between them. To remove these variations the batch mean
765 was subtracted from the relative amount of protein for each sample and this difference then
766 added to the overall mean of all batches. This quantity is referred to as "standardized relative
767 amount of protein".

768 The densities of the target protein bands were not adjusted to the content of any cellular
769 housekeeping proteins, such as glyceraldehyde 3-phosphate dehydrogenase or beta-actin, since
770 it is not guaranteed that such reference proteins will be unaltered in our experiments[98–101]
771 Instead, we consistently verified equal loading of sample proteins by Ponceau S staining.

772 RNA isolation and RT-PCR

773 Total RNA from chick brain samples were isolated with "Nucleospin RNA: DNA, RNA and
774 Protein purification (740955 Macherey-Nagel). Complementary DNA (cDNA) was
775 synthesized by High-Capacity cDNA Reverse Transcription Kit (4368814, Thermofisher
776 Scientific). Relative ENSGAL000100007489 and ENSGALG00010026609 cDNA copy
777 numbers were determined by real-time PCR using the Step One Real-Time PCR System
778 (Applied Biosystems) with the SYBR Green detection method and was normalized to
779 HPRT1[102]. The comparative CT (ddCT) method was used to determine the relative target
780 quantity in samples[103]. The same fraction of IMM RNA isolated from the IMMs of untrained
781 chicks was used in all experiments as a reference sample. Amplicons from randomly chosen

782 samples were sequenced at BGI BIO Solutions Co., Ltd. (Hong-Kong, China). The primer
783 sequences are provided in Table 1.

784

785 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

786

787 **Subcellular localization of lncRNAs**

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

792 **Interpretation of a correlation between a molecular quantity and**
793 **preference score**

794 A significant association between a molecular quantity (e.g., protein abundance, gene
795 expression level, etc.) and preference score may indicate (i) that the change in the quantity is
796 induced by the learning process, or that (ii) the quantity may facilitate / impede the learning
797 process, i.e., it provides some type of predisposition. In order to distinguish between these
798 hypotheses, we use a design comprising two groups of chicks: a trained group, where both
799 the molecular quantity under examination and the preference score are measured, and an
800 untrained group, where only the molecular quantity can be measured. The molecular quantity
801 is regressed over the preference score in the trained group with a linear model.

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

805 The y-intercept at preference score 50 is an estimate of the molecular quantity when no
806 learning occurs in trained chicks. It thus measures the effect of procedures such as handling,
807 locomotor activity, sensory stimulation etc. that occur during training but which are separable

808 from learning (i.e. side-effects). If the intercept at preference score 50 is not significantly
809 different from the mean value in untrained chicks, there is no evidence of side-effects.

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

814 If the correlation is due to a predisposition (see (ii) above), training has no effect on the
815 molecular quantity and the variance of the trained and untrained chicks should be statistically
816 homogeneous. The correlation accounts for some of the variance in trained chicks and the
817 variance about the regression line is correspondingly reduced: a significant *reduction* in
818 residual variance about the regression line, relative to the variance in untrained chicks, is
819 therefore evidence of a *predisposition*[3].

820 Fluorescent in-situ hybridization

821 The experiment was conducted in four batches. Each batch contained a good learner, a poor
822 learner and an untrained chick from the same hatch of eggs, that is 12 samples in total.
823 Hatching, rearing and training were performed as described for the other experiments. After
824 killing, brain tissue was removed and fixed in 4% paraformaldehyde solution for 3 days. The
825 brains were then transferred to a freshly prepared methacarn solution consisting of 60%
826 (vol/vol) absolute methanol, 30% chloroform, and 10% glacial acetic acid[104]. Dehydration
827 was performed through graded ethanol, followed by xylene, then paraffin wax for
828 embedding. Samples were sectioned at 7 µm onto superfrost glass slides. For each brain, two
829 coronal sections were selected corresponding to each of the coordinates A7.0, A7.6 and A8.0
830 of the chick brain atlas by Kuenzel and Masson[105]. Measurements of cellular fluorescence
831 were made on each section within a rectangular sampling frame measuring 0.3 mm x 0.9 mm
832 in the IMM and a square sampling frame 0.3 mm x 0.3 mm in the adjacent nidopallium
833 (Supplementary Figure S33). Fluorescence was measured at each of the three anteroposterior
834 levels for LOC121109133-O1, SLC and PPIB, and nuclei were stained with DAPI. One of
835 the sections from each chick was stained for GAD instead of PPIB.

836 In any one batch, the sections stained for GAD were at the same anteroposterior level in all
837 three experimental groups. There were three GAD sections from the posterior IMM in batch
838 1, three sections from the intermediate IMM in each of batches 2 and 3 and three sections
839 from the anterior IMM in batch 4.

840 RNAscope was conducted to manufacturer specifications using probes Gg-GAD2-C1
841 (reference number 1300861-C1, target region 448-1421, 20 pairs), Gg-PPIB-C1 (reference
842 number 453371, target region 67-876, 16 pairs), Gg-SLC17A6-C2 (reference number
843 1300871-C1, target region 591-1864, 20 pairs), Gg- LOC121109133-O1-C3 (reference
844 number 1300881-C3, target region 91-656, 11 pairs), and Opal fluorophores 520, 570, and
845 650. Samples were scanned on a Zeiss Axioscan 7 at x20 and images were imported into
846 QuPath[48]. Cell detection in QuPath was used, with the following parameters. Detection
847 channel DAPI, requested pixel size 0.5 µm, background radius 8um, use opening by
848 reconstruction yes, median filter radius 0um, sigma 1.5 µm, minimum area 10um², maximum
849 area 400um², threshold 50, cell expansion 5 µm, include cell nucleus yes, smooth boundaries
850 yes, make measurements yes, and visually validated. Subcellular detection in QuPath was
851 then used, with the following parameters, detection threshold SLC 2000, PPIB 400, LNC
852 4000, smooth before detection yes, spilt by intensity yes, split by shape yes, expected spot
853 size 1um², min spot size 0.1 µm², max spot size 4 µm², include clusters yes (still being
854 optimized for all sections), and visually validated.

855 Fluorescent in-situ hybridization data analysis

856 Quantitative measurements on fluorescent intensities provided by QuPath were analyzed
857 using the R Statistical Software[106]. Two main analyses were performed:

858 - *Contrasting GLUBK89 expression values across experimental groups*.
859 For this analysis, only sections stained for SLC17A6, PPIB and GLUBK89 were used
860 (see Supplementary Figure 30 for a schema of the experimental design). Glutamatergic
861 nuclei were identified by thresholding the SLC17A6 marker (> 1700), while a normalized
862 expression value for GLUBK89 in each nucleus was computed by dividing for its
863 fluorescent intensity for the fluorescent intensity associated to the housekeeping gene
864 PPIB. These normalized expression values at the nucleus level were then averaged across
865 sample, tissue (IMM and nidopallium), and side (left, right side of the brain). The
866 resulting average values were modeled, independently in the IMM and nidopallium,
867 according to the following linear mixed model[107]:
868 norm_expression_values ~ group * side, random = ~ 1|batch/chick
869 Which means that we model the normalized expression values depending on experimental
870 groups (good learners, poor learners, untrained), side (left or right), and their interaction,
871 as well as taking into account the variability provided by the nested random effect chick
872 (sample) and batch (see Supplementary Figure S30 for batch definitions). The statistical

873 significance of contrasts between experimental group levels within this model were
874 investigated with the Tukey method[108,109].
875 - *Contrasting GLUBK89 expression between GABAergic and glutamatergic cells*.
876 Sections stained for SLC17A6, GAD2 and GLUBK89 were considered for this analysis.
877 Glutamatergic and GABAergic nuclei were identified by thresholding the SLC17A6 (>
878 1700) and GAD2 (> 630) markers, respectively. The expression of GLUBK89 was
879 contrasted between the two cell types using a t-test.

880 ## Data and code availability

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

884 ## Funding

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

889 ## Author contributions

890 V.L., R.S., Z.Kh., and D.C.G. conceived the project and designed the experiments. L.C.,
891 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.,
892 B.J.M., V.L. performed bioinformatic analysis and analyzed the data. V.L., R.S., Z.Kh.,
893 B.J.M., D.G.C., L.A.I., and J.T. interpreted the results and wrote the manuscript with input
894 from all authors.

895 ## Competing interests

896 The authors declare no competing interests.

897

References

1. 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).

2. 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).

3. 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).

4. 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).

5. 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).

6. 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).

7. 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).

8. 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).

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

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

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

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

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

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

942 15. Harvey, R. J., McCabe, B. J., Solomonia, R. O., Horn, G. & Darlison, M.
943 G. Expression of the GABAA receptor γ4-subunit gene: Anatomical
944 distribution of the corresponding mRNA in the domestic chick forebrain
945 and the effect of imprinting training. *European Journal of Neuroscience*
946 **10**, (1998).

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

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

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

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

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

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

962 22. Chen, X. *et al.* Transcriptomic mapping uncovers Purkinje neuron
963 plasticity driving learning. *Nature* **605**, (2022).

964 23. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's
965 disease. *Nature* **570**, (2019).

966 24. Morabito, S., Miyoshi, E., Michael, N., Shahin, S. & Swarup, V. Single-
967 nuclei chromatin accessibility and transcriptomics unravels altered human
968 oligodendrocyte heterogeneity in Alzheimer's disease. *Alzheimer's &*
969 *Dementia* **16**, (2020).

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

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

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

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

29. Tosches, M. A. *et al.* Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* **360**, 881–888 (2018).

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

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

32. 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).

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

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

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

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

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

1006 38. Morabito, S., Reese, F., Rahimzadeh, N., Miyoshi, E. & Swarup, V.
1007 hdWGCNA identifies co-expression networks in high-dimensional
1008 transcriptomics data. *Cell Reports Methods* **3**, (2023).

1009 39. Wang, Q., Roy, B. & Dwivedi, Y. Co-expression network modeling
1010 identifies key long non-coding RNA and mRNA modules in altering
1011 molecular phenotype to develop stress-induced depression in rats. *Transl*
1012 *Psychiatry* **9**, (2019).

1013 40. Mukherjee, S., Banerjee, B., Karasik, D. & Frenkel-Morgenstern, M.
1014 Mrna-lncrna co-expression network analysis reveals the role of lncrnas in
1015 immune dysfunction during severe sars-cov-2 infection. *Viruses* **13**,
1016 (2021).

1017 41. Sengar, A. S. *et al.* Control of Long-Term Synaptic Potentiation and
1018 Learning by Alternative Splicing of the NMDA Receptor Subunit GluN1.
1019 *Cell Rep* **29**, (2019).

1020 42. Alberini, C. M. Transcription factors in long-term memory and synaptic
1021 plasticity. *Physiological Reviews* vol. 89 Preprint at
1022 https://doi.org/10.1152/physrev.00017.2008 (2009).

1023 43. Shen, K. & Cowan, C. W. Guidance molecules in synapse formation and
1024 plasticity. *Cold Spring Harbor perspectives in biology* vol. 2 Preprint at
1025 https://doi.org/10.1101/cshperspect.a001842 (2010).

1026 44. Wong, L. W., Tann, J. Y., Ibanez, C. F. & Sajikumar, S. The p75
1027 neurotrophin receptor is an essential mediator of impairments in
1028 hippocampal-dependent associative plasticity and memory induced by
1029 sleep deprivation. *Journal of Neuroscience* **39**, (2019).

1030 45. Yu, M. *et al.* Overexpression of Kcnmb2 in Dorsal CA1 of Offspring
1031 Mice Rescues Hippocampal Dysfunction Caused by a Methyl Donor-Rich
1032 Paternal Diet. *Front Cell Neurosci* **12**, (2018).

1033 46. Solomonia, R. O., McCabe, B. J., Jackson, A. P. & Horn, G. Clathrin
1034 proteins and recognition memory. *Neuroscience* **80**, 59–67 (1997).

1035 47. Sweeney, B. A. *et al.* RNAcentral 2021: Secondary structure integration,
1036 improved sequence search and new member databases. *Nucleic Acids Res*
1037 **49**, (2021).

1038 48. Bankhead, P. *et al.* QuPath: Open source software for digital pathology
1039 image analysis. *Sci Rep* **7**, (2017).

1040 49. Rose, S. P. R. God's organism? The chick as a model system for memory
1041 studies. *Learning and Memory* vol. 7 1–17 Preprint at
1042 https://doi.org/10.1101/lm.7.1.1 (2000).

1043 50. Matsushima, T., Izumi, T. & Vallortigara, G. The domestic chick as an
1044 animal model of autism spectrum disorder: building adaptive social
1045 perceptions through prenatally formed predispositions. *Frontiers in*
1046 *Neuroscience* vol. 18 Preprint at
1047 https://doi.org/10.3389/fnins.2024.1279947 (2024).

1048 51. Horn, G. *Memory, Imprinting, and the Brain*. (Oxford University Press,
1049 1985).

1050 52. Atoji, Y. & Karim, M. R. Homology of the mesopallium in the adult
1051 chicken identified by gene expression of the neocortical marker
1052 cholecystokinin. *Neurosci Lett* **562**, 85–89 (2014).

1053 53. Suzuki, I. K., Kawasaki, T., Gojobori, T. & Hirata, T. The Temporal
1054 Sequence of the Mammalian Neocortical Neurogenetic Program Drives
1055 Mediolateral Pattern in the Chick Pallium. *Dev Cell* **22**, (2012).

1056 54. Hecker, N. *et al.* Enhancer-driven cell type comparison reveals similarities
1057 between the mammalian and bird pallium. *Science (1979)* **387**, (2025).

1058 55. Grinman, E., Espadas, I. & Puthanveettil, S. V. Emerging roles for long
1059 noncoding RNAs in learning, memory and associated disorders. *Neurobiol*
1060 *Learn Mem* **163**, (2019).

1061 56. Irwin, A. B., Bahabry, R. & Lubin, F. D. A putative role for lncRNAs in
1062 epigenetic regulation of memory: Memory Regulation via lncRNAs.
1063 *Neurochemistry International* vol. 150 Preprint at
1064 https://doi.org/10.1016/j.neuint.2021.105184 (2021).

1065 57. Kemenes, G., Benjamin, P. R. & Kemenes, I. The role of non-coding
1066 RNAs in the formation of long-term associative memory after single-trial
1067 learning in Lymnaea. *Front Behav Neurosci* **16**, 1005867 (2022).

1068 58. Zhang, J. *et al.* Glutamate-activated BK channel complexes formed with
1069 NMDA receptors. *Proc Natl Acad Sci U S A* **115**, (2018).

1070 59. Ancatén-González, C. *et al.* BK channels mediate a presynaptic form of
1071 mGluR-LTD in the neonatal hippocampus. *Proc Natl Acad Sci U S A* **122**,
1072 (2025).

1073 60. Zhu, Y. *et al.* The long noncoding RNA glycoLINC assembles a lower
1074 glycolytic metabolon to promote glycolysis. *Mol Cell* **82**, (2022).

1075 61. Daniels, N. J. *et al.* Functional analyses of human LUC7-like proteins
1076 involved in splicing regulation and myeloid neoplasms. *Cell Rep* **35**,
1077 (2021).

1078 62. Kenny, C. J. *et al.* LUC7 proteins define two major classes of 5′ splice
1079 sites in animals and plants. *Nature Communications 2025 16:1* **16**, 1–15
1080 (2025).

1081 63. Shen, L. *et al.* Loss-of-function mutation in PRMT9 causes abnormal
1082 synapse development by dysregulation of RNA alternative splicing.
1083 *Nature Communications 2024 15:1* **15**, 1–20 (2024).

1084 64. Jolly, S., Journiac, N., Vernet-Der Garabedian, B. & Mariani, J.
1085 RORalpha, a key to the development and functioning of the brain.
1086 *Cerebellum* **11**, 451–452 (2012).

1087 65. Sarachana, T. & Hu, V. W. Genome-wide identification of transcriptional
1088 targets of RORA reveals direct regulation of multiple genes associated
1089 with autism spectrum disorder. *Mol Autism* **4**, 1–19 (2013).

1090 66. Sarachana, T. & Hu, V. W. Genome-wide identification of transcriptional
1091 targets of RORA reveals direct regulation of multiple genes associated
1092 with autism spectrum disorder. *Mol Autism* **4**, 1–19 (2013).

1093 67. Akila, S. *et al.* Molecular Signatures of Resilience to Alzheimer's Disease
1094 in Neocortical Layer 4 Neurons. *bioRxiv* 2024.11.03.621787 (2024)
1095 doi:10.1101/2024.11.03.621787.

1096 68. Valle-Bautista, R. *et al.* From songbird to humans: The multifaceted roles
1097 of FOXP2 in speech and motor learning. *Neurosci Biobehav Rev* **167**,
1098 (2024).

1099 69. Li, S., Weidenfeld, J. & Morrisey, E. E. Transcriptional and DNA binding
1100 activity of the Foxp1/2/4 family is modulated by heterotypic and
1101 homotypic protein interactions. *Mol Cell Biol* **24**, 809–822 (2004).

1102 70. Tiunova, A. A., Komissarova, N. V & Anokhin, K. V. Mapping the
1103 Neural Substrates of Recent and Remote Visual Imprinting Memory in the
1104 Chick Brain. *Front Physiol* **10**, (2019).

1105 71. Solomonia, R. O., Kiguradze, T., McCabe, B. J. & Horn, G. Neural cell
1106 adhesion molecules, CaM kinase II and long-term memory in the chick.
1107 *Neuroreport* **11**, 3139–3143 (2000).

1108 72. Behroozi, M. *et al.* Functional MRI of imprinting memory in awake
1109 newborn domestic chicks. *Communications Biology 2024 7:1* **7**, 1–13
1110 (2024).

1111 73. Brose, K. *et al.* Slit proteins bind Robo receptors and have an
1112 evolutionarily conserved role in repulsive axon guidance. *Cell* **96**, 795–
1113 806 (1999).

1114 74. Whitford, K. L. *et al.* Regulation of cortical dendrite development by Slit-
1115 Robo interactions. *Neuron* **33**, 47–61 (2002).

1116 75. Chen, X. *et al.* Computational models for lncRNA function prediction and
1117 functional similarity calculation. *Brief Funct Genomics* **18**, 58–82 (2019).

1118 76. Smith, F. V. & Bird, M. W. The relative attraction for the domestic chick
1119 of combinations of stimuli in different sensory modalities. *Anim Behav* **11**,
1120 300–305 (1963).

1121 77. Horn, G. & Johnson, M. H. Memory systems in the chick: Dissociations
1122 and neuronal analysis. *Neuropsychologia* **27**, 1–22 (1989).

1123 78. Davies, D. C., Horn, G. & McCabe, B. J. Noradrenaline and learning:
1124 Effects of the noradrenergic neurotoxin DSP4 on imprinting in the
1125 domestic chick. *Behavioral Neuroscience* **99**, 652–660 (1985).

1126 79. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling
1127 of single cells. *Nat Commun* **8**, (2017).

1128 80. Alvarez, M. *et al.* Enhancing droplet-based single-nucleus RNA-seq
1129 resolution using the semi-supervised machine learning classifier DIEM.
1130 *Sci Rep* **10**, (2020).

1131 81. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet
1132 Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest
1133 Neighbors. *Cell Syst* **8**, (2019).

1134 82. Fleming, S. J. *et al.* Unsupervised removal of systematic background
1135 noise from droplet-based single-cell experiments using CellBender. *Nat*
1136 *Methods* **20**, (2023).

1137 83. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**,
1138 (2019).

1139 84. Waltman, L. & Van Eck, N. J. A smart local moving algorithm for large-
1140 scale modularity-based community detection. *European Physical Journal*
1141 *B* **86**, (2013).

1142 85. Miller, S. A. *et al.* LSD1 and aberrant DNA methylation mediate
1143 persistence of enteroendocrine progenitors that support BRAF-mutant
1144 colorectal cancer. *Cancer Res* **81**, (2021).

1145 86. Squair, J. W. *et al.* Confronting false discoveries in single-cell differential
1146 expression. *Nat Commun* **12**, (2021).

1147 87. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change
1148 and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 1–21
1149 (2014).

1150 88. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based
1151 approach for interpreting genome-wide expression profiles. *Proc Natl*
1152 *Acad Sci U S A* **102**, 15545–50 (2005).

1153 89. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for
1154 interpreting omics data. *Innovation (Cambridge (Mass.))* **2**, (2021).

1155 90. The Gene Ontology Consortium. Gene Ontology Consortium: going
1156 forward. *Nucleic Acids Res* **43**, D1049–D1056 (2015).

1157 91. Buza, T. J., Mccarthy, F. M., Wang, N., Bridges, S. M. & Burgess, S. C.
1158 Gene Ontology annotation quality analysis in model eukaryotes. *Nucleic*
1159 *Acids Res* **36**, (2008).

1160 92. Kui, H. *et al.* Gene expression profiles of specific chicken skeletal
1161 muscles. *Sci Data* **9**, (2022).

1162 93. Nie, H. *et al.* Gene expression in chicken reveals correlation with
1163 structural genomic features and conserved patterns of transcription in the
1164 terrestrial vertebrates. *PLoS One* **5**, (2010).

1165 94. Gu, Z. & Hübschmann, D. simplifyEnrichment: A Bioconductor Package
1166 for Clustering and Visualizing Functional Enrichment Results. *Genomics*
1167 *Proteomics Bioinformatics* **21**, (2023).

1168 95. Baran, Y. *et al.* MetaCell: Analysis of single-cell RNA-seq data using K-
1169 nn graph partitions. *Genome Biol* **20**, (2019).

1170 96. Bilous, M. *et al.* Metacells untangle large and complex single-cell
1171 transcriptome networks. *BMC Bioinformatics* **23**, (2022).

1172 97. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a
1173 hierarchical cluster tree: The Dynamic Tree Cut package for R.
1174 *Bioinformatics* **24**, (2008).

1175 98. Dittmer, A. & Dittmer, J. β-Actin is not a reliable loading control in
1176 Western blot analysis. *Electrophoresis* **27**, (2006).

1177 99. Li, R. & Shen, Y. An old method facing a new challenge: Re-visiting
1178 housekeeping proteins as internal reference control for neuroscience
1179 research. *Life Sciences* vol. 92 Preprint at
1180 https://doi.org/10.1016/j.lfs.2013.02.014 (2013).

1181 100. Ghosh, R., Gilda, J. E. & Gomes, A. V. The necessity of and strategies for
1182 improving confidence in the accuracy of western blots. *Expert Review of*
1183 *Proteomics* vol. 11 Preprint at
1184 https://doi.org/10.1586/14789450.2014.939635 (2014).

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

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

103. 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).

104. 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).

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

106. 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).

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

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

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

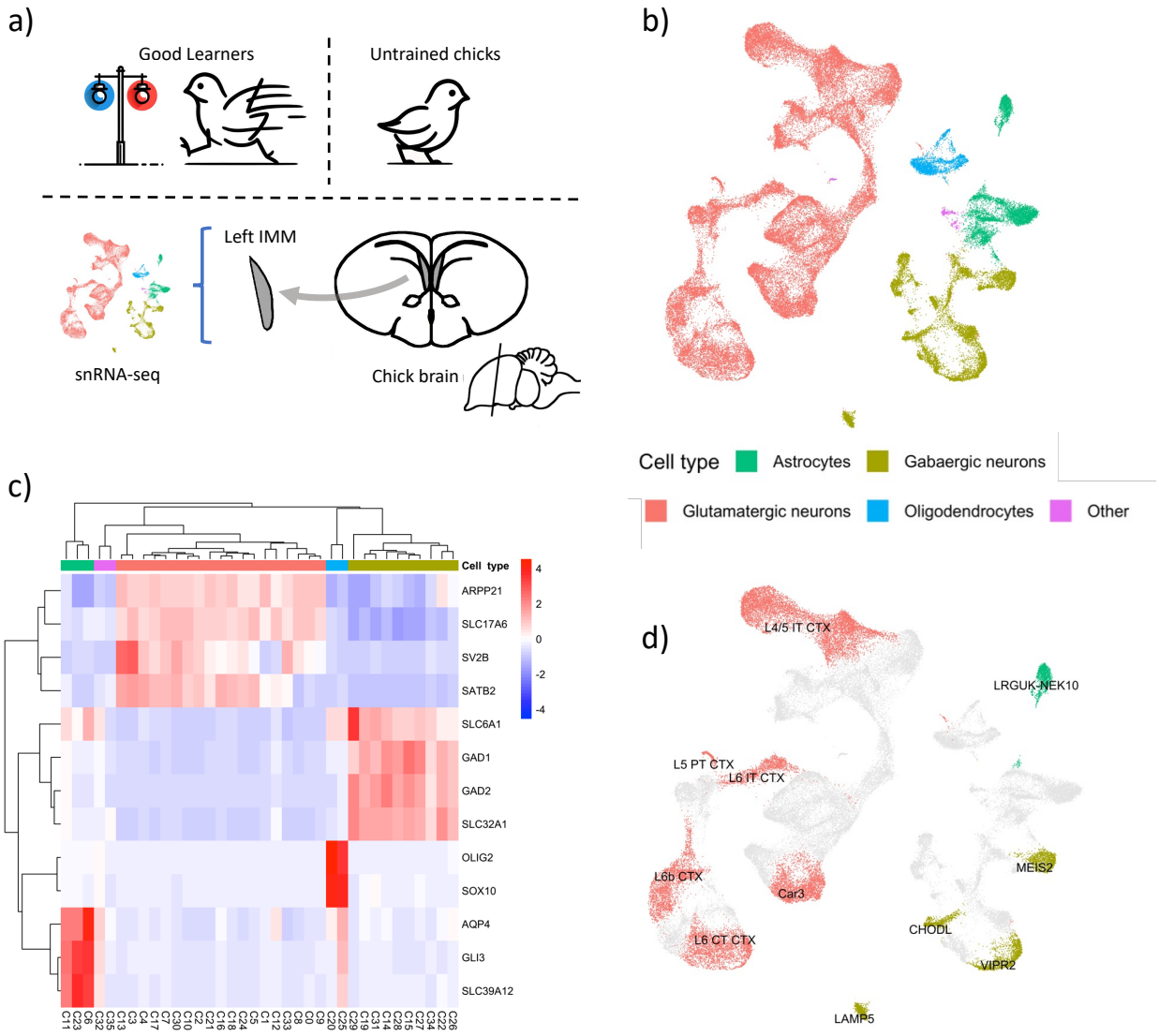

Figure 1: (a) Graphical presentation of the study design. 22-28h old chicks are either subjected to a preference test for assessing their imprinting or left untrained. Trained chicks are selected only if consistently directing themselves towards the training stimulus (good learners; see methods for details). 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, oligodendrocytes, as well as GABAergic and glutamatergic neurons. (c) Heatmap showing the expression levels of gene markers for major cell types. (d) Localization of notable cell subtypes, see main text for details.

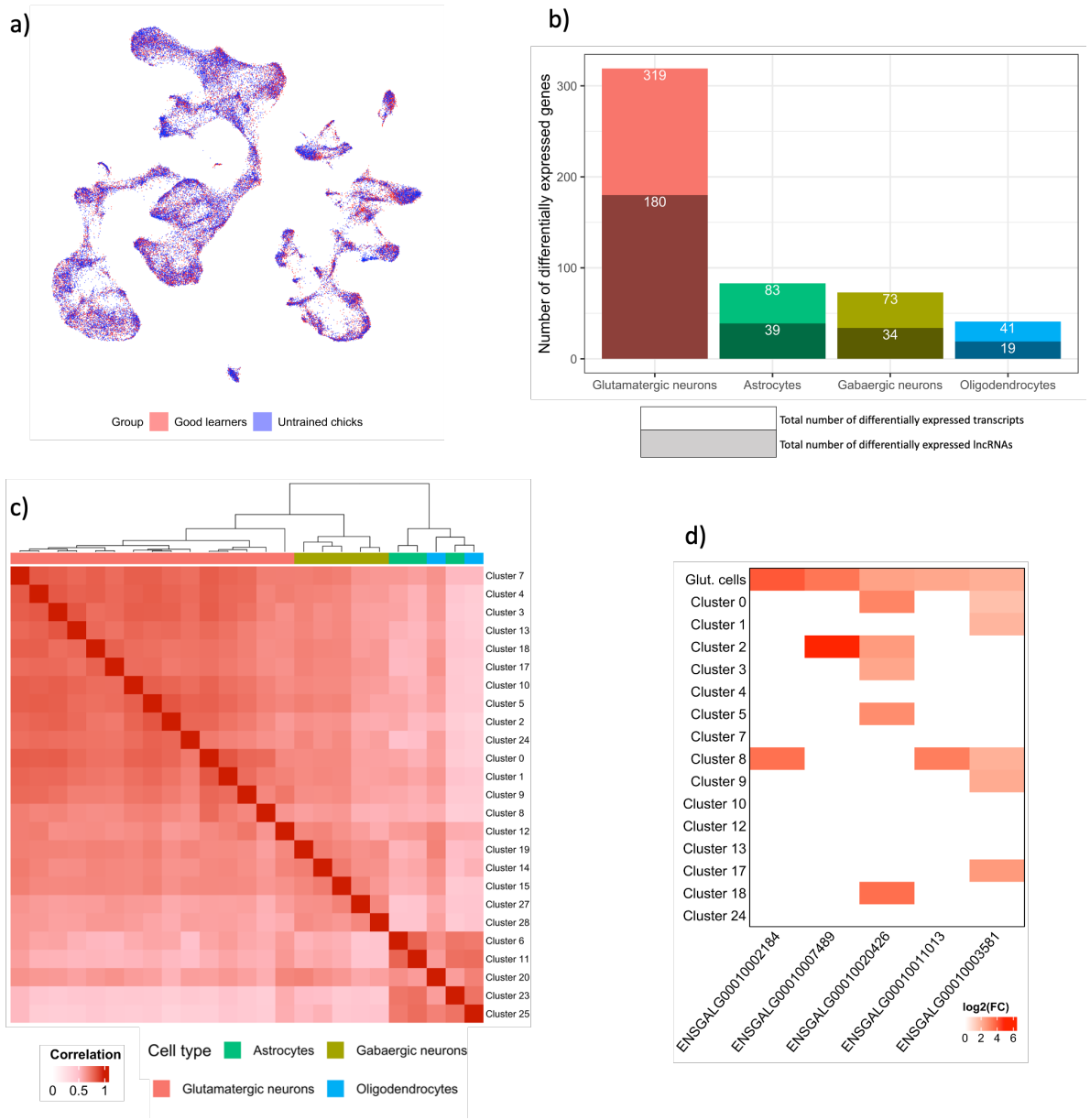

Figure 2: (a) UMAP visualizing IMM cells according to the group of their sample of origin, red good learners and blue untrained chicks. (b) Number of statistically significant transcriptomic changes (log FC > 1 in absolute value, FDR <= 0.05). Darker shaded bars indicate the number of non-coding RNA found differentially expressed. (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 all contrast resulting to a FDR <= 0.05, as well as with at least 50 cells forming each pseudo-bulk sample (see Methods for details).

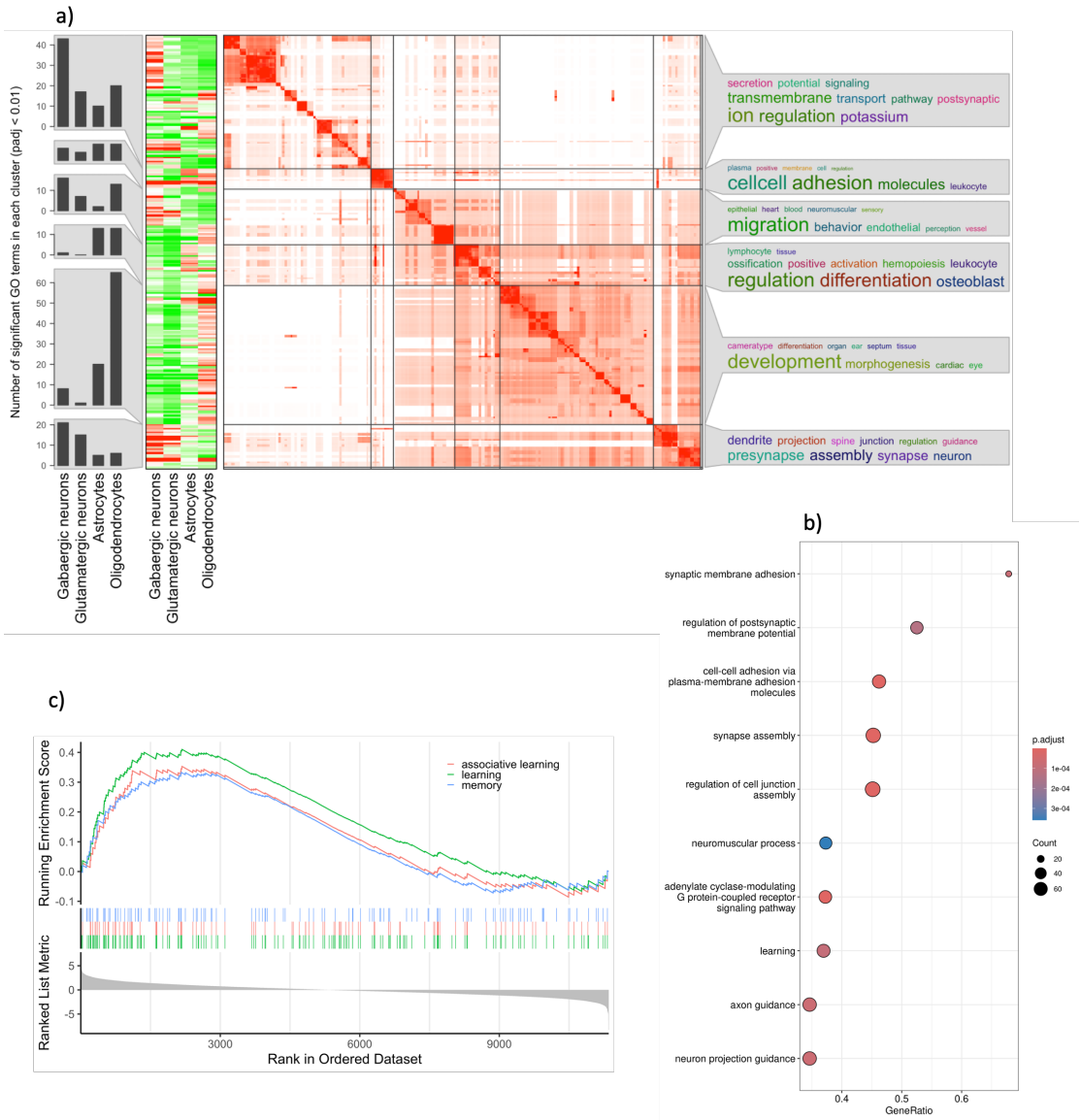

Figure 3: (a) Gene Ontology biological processes (BPs) significantly upregulated (FDR < 0.01) in good learner chicks, across all cell types, clustered and summarized. Central heatmap reports the BP-to-BP term similarity (see Methods), with BP clusters formed through the Louvain algorithm. Left side heatmap reports the significance of each BP in each cell type (red indicating significance, green close-to-1 adj. p-value). 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 type and BP cluster. (b) Dot plot presenting the most upregulated BP in glutamatergic neurons; color indicate significance, size the number of upregulated genes within the BP. (c) Enrichment of relevant BPs in glutamatergic neurons: GO:0007612 learning, GO:0007613 memory, GO:0008306 associative learning.

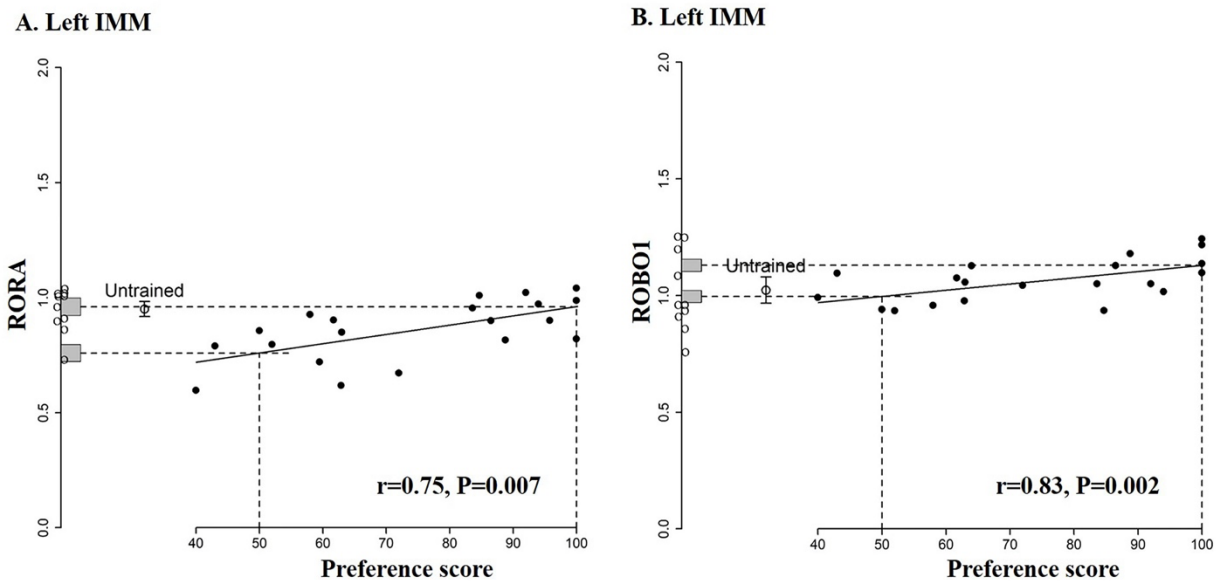

Figure 4. Standardized relative amount of RORA (A) and ROBO1 (B) in the left IMM plotted against preference score. Filled circles, trained chicks. Open circle with error bars, mean level in untrained controls ± SEM; individual values for untrained chicks are shown as open circles along the y-axis. Vertical 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 proteins correlations are significant

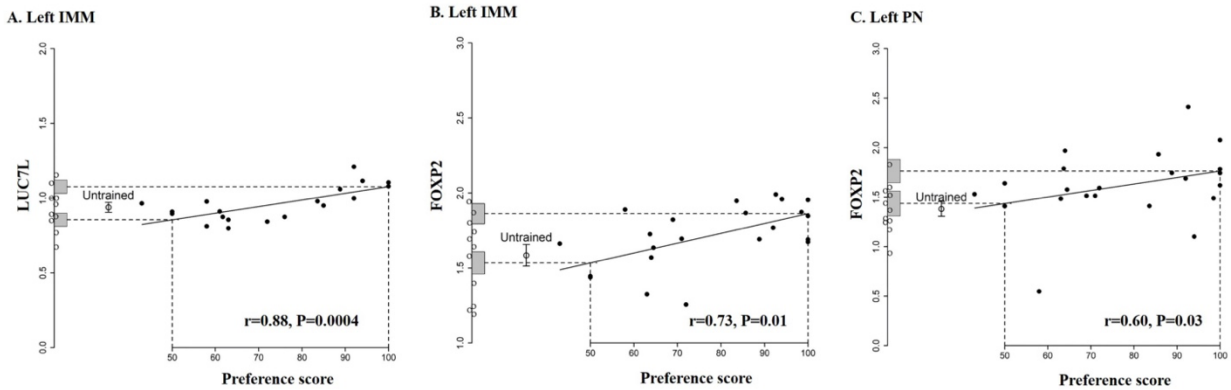

Figure 5. Standardized relative amount of LUC7L (A-left IMM) and FOXP2 (B-left IMM, C-left PPN) plotted against preference score. Filled circles, trained chicks. Open circle with error bars, mean level in untrained controls ± SEM; individual values for untrained chicks are shown as open circles along 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.

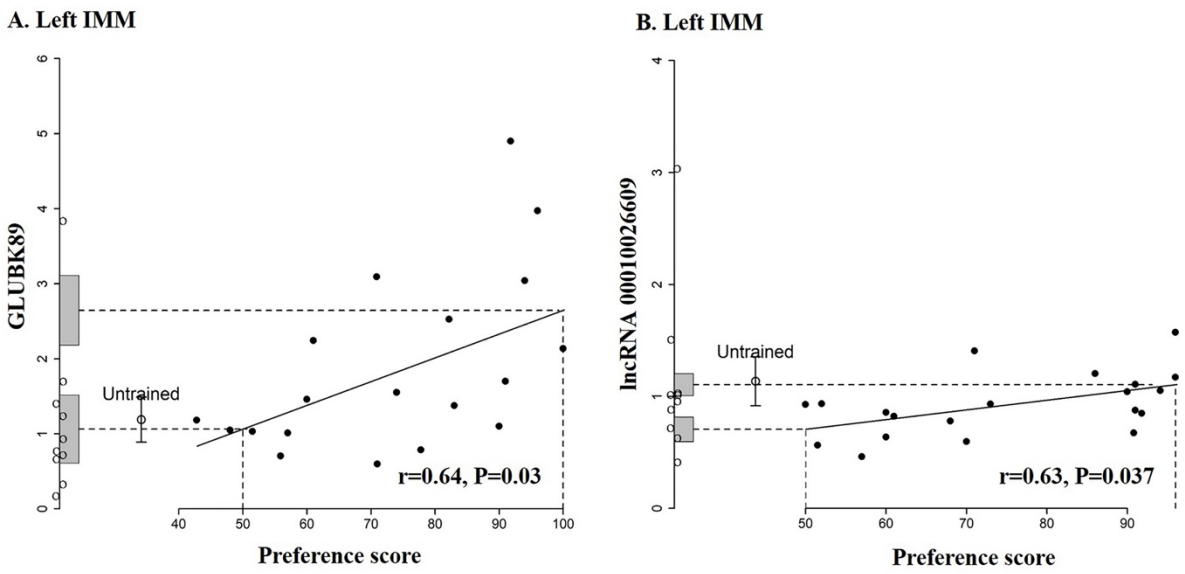

Figure 6. Standardized relative amount of GLUBK89 (A) and lncRNA 6609 (B) in the left IMM plotted against preference score. Filled circles, trained chicks. Open circle with error bars, mean level in untrained controls ± SEM; individual values for untrained chicks are shown as open circles along the y-axis. 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).

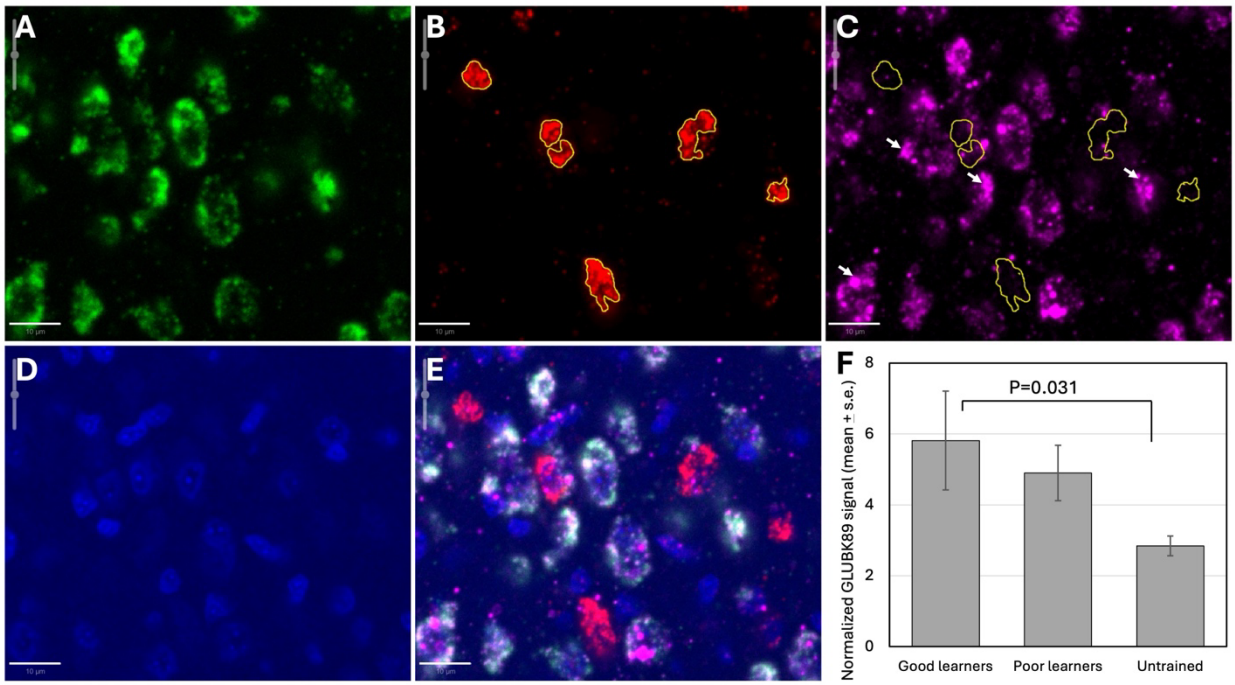

Figure 7. Representative confocal FISH imaging of the left IMM region in the brain of a trained chick. (A) Glutamatergic neurons expressing SLC17A6 are labeled with a green fluorescent marker. (B) GABAergic neurons expressing GAD2 are labeled with a red fluorescent marker, with their boundaries outlined 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 mm. (F) GLUBK89 expression, normalized by PPI, in the left IMM across groups.