nature portfolio

Peer Review File

Uncovering the Molecular Logic of Cortical Wiring between Neuronal subtypes Across Development Through Ligand–Receptor Inference

Corresponding Author: Dr Antoine de Chevigny

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The manuscript by Mathieu et al. sets out to address the fundamental question of how synaptic specificity is determined during mouse brain development. To do so, the authors compile single cell RNAseq datasets from developing and adult mouse cortex, conduct bioinformatic analyses including pseudotime, “pseudo-layer”, and ligand-receptor predictionsto explain cell-cell interactions during corticogenesis.

Overall this is a very interesting analysis, validated by the prediction of known patterns of cell cell interactions, augmented by perturbational experiments involving short hairpin RNA knock-down of candidate cell contact molecules in vivo mouse brain. I believe that this analysis and its results will be of great value and interest to the community, and that it merits publication.

Below, we list specific issues that should be addressed.

Compiling datasets

The authors collected scRNAseq and snRNAseq data from SS cortex at P0-P30.The pipeline they describe for integrating data from multiple sources seems reasonable and conventional, but nonetheless it would be helpful to validate if any insights differ between nuclear versus whole cell sequencing.

The authors should discuss to what extent their analysis, which is focused on SS cortex, could be generalizable to other cortical regions.

Relatedly, whether progenitor domain of origin of interneurons plays any significant role in the types of interactions predicted by their algorithm should be further interrogated and discussed.

Pseudo-spatiotemporal analysis

The authors performed pseudotime and pseudolayer analysis to build smooth models of gene expression across developmental time and space. The pseudotime analysis using slingshot is conventional, but the pseudolayer method is less clear. The description of the pseudolayer method in the main text mentions cross-referencing their data with patch-seq and MERFISH datasets, but there is no further explanation (not even in the methods) of how these datasets are used to estimate pseudolayer. It seems that their approach to pseudolayer estimation relying on spatial transcriptional gradients could be further discussed. The authors should consider validating their method with a more established method for spatial estimation from scRNAseq data like tangram.

The “waves” estimated from pseudotime and pseudolayer models, with shared dynamics of gene activity across time/space, demonstrate that the pseudotime and pseudolayer results are meaningful. However, I can’t find a description of how these waves are computed in the methods. Conversely, there’s no description of fitting the GAM in the main text, but there is in the methods. Small detail: it would be clearer to refer to the pseudolayer dynamics as something like “upper and lower

gradients" rather than "first and last waves."

The depiction of the developing cortex colored by cell-type in Figure 2E is useful, but too crowded and should be simplified.

LR-interaction analysis

The authors assembled the largest existing LR database and used a modified version of scSeqComm to infer magnitudes of cell-cell interactions using more conventional LR correlations between cell types (intercellular score) and less conventional pathway activation where such pathways are known (intracellular scores). With this approach, they aimed to validate their model with known LR pairs and by discovering novel LR interactions which they validated with a knockout experiment. These results give credence to their data processing and analysis pipelines.

However, a major concern with the author's validation approach is how they modify the intercellular score calculation by adding information about developmental timing and connectivity patterns from patch-seq dataset (lines 231-239). It seems that they use this score to predict connectivity between cell types which were classified using the patch-seq datasets. That is, it seems like information about the validation data entered the model during training. The authors should clarify that this was not the case. Further, they should show the implications of their modifications to the scSeqComm technique and explain why their modification to the intercellular score is necessary and valid.

The authors make a core assumption (lines 208-211) that the transcription of GlutN ligands and GABAN receptors should be correlated between synaptically connected GlutNs and GABANs. Though the converse of this statement seems true, the statement itself does not seem true necessarily. Given two connected neurons, it's possible that their transcription of an LR pair is uncorrelated across time, because the ligand and/or receptor may have roles outside of the time period relevant for the connectivity of these two neurons. The authors should discuss the potential limitations of the method.

The authors observe that very few LR combinations are unique to a single cell-type pair, while many combinations are shared between pairs. From this observations, they claim that a "combinatorial code" of common LR pairs drives synapse specificity. This hypothesis is enticing, but it could be much more strongly supported with statistical tests. Specifically, the authors should train a model to predict cell type pair from LR interaction scores, and compare the performance of this model when common or rare LR pairs are included as features. If the model using shared LR pairs outperforms the model using rare pairs, this finding would support their hypothesis.

When quantifying LR interactions during synaptogenesis and in developmental diseases, the authors use numbers of interacting LR pairs as a metric of cell-cell interaction (Figures 3 and 4). However, this metric assumes that every LR pair has the same degree of effect on cell function. It's unclear why they don't use the available intercellular and intracellular scores instead. Also, in Figures 3 and 4, the white-red-black colorbar indicating number of LR pairs is not very colorblind friendly. If an intermediate color is needed, green or orange (as long as it differs from the colors indicating GlutN and GABAN) could be used. The "normalized soma depth" colorbar should be changed to a different scheme than number of LR pairs, if it is even needed (seems redundant).

Lastly, the result that CDH13-CDH13 interactions largely take place between Sst/Pvalb and L5/6 (conveyed lines 364-367) isn't clearly conveyed by the corresponding Figure 6A-C. It's not clear what the red boxes in Figure 6A are highlighting, and it seems like there are other GABAN types (like VIP) which have CDH13 interaction with L5/6 GlutNs.

The fact that the analysis considers only interactions between neurons as opposed to glial cells should be reflected in the title.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
Summary Paragraph

This manuscript presents a single-cell transcriptomics atlas spanning the development of the mouse somatosensory cortex, integrating both newly generated and publicly available single-cell and single-nucleus RNA-seq datasets from embryonic to adult stages. This efficient approach leverages existing large-scale datasets, making optimal use of available resources. The authors assembled and curated a large ligand-receptor (LR) pair database, which they applied to their atlas to systematically infer putative LR interactions between transcriptomically-defined neuronal groups, with a focus on connections between inhibitory and excitatory neurons during circuit assembly. Some predictions are experimentally validated, including that neogenin-1 is the main receptor for CBLN4 during early synapse formation, and that cadherins CDH13 and PCDH8 have different, layer-specific roles in regulating inhibitory synapses onto deep and superficial neurons. A rich, interactive web app is provided (though currently not practically usable due to severe performance issues), potentially enabling broad access to the spatiotemporal transcriptomic data and LR predictions. A key limitation of transcriptomic data is its lack of subcellular localization information for ligands and receptors, which is essential for proper excitatory and inhibitory neuronal signaling.

This work provides a valuable reference for researchers studying cortical development and neuronal circuit wiring. The computational and bioinformatic analyses are well executed, and targeted experiments provide direct validation for selected predictions. I have a few major and mostly minor points that need to be addressed. If these are resolved, I support publication.

Major Comments

The interactive web resource, meant to be a key output of the study, is currently not functioning very well. It reacts slow and several parts don't load at all. Full release of this tool is important for maximizing the impact of the work.

While the integration of multiple single-cell datasets is a strength, the manuscript lacks critical analysis of batch effects, dataset biases, or possible artifacts arising from combining data produced under differing experimental protocols. The authors should address how integration choices may impact downstream LR inference. The manuscript would benefit from providing intermediate steps and results for data integration and label transfer, enabling readers to better assess the accuracy and reliability of the integration process and cell-type annotation. The authors could explore the integrated dataset more thoroughly by presenting developmental trajectories, UMAPs labeled by developmental stage (as shown in Di Bella et al.), and cell type abundances across stages.

The study primarily emphasizes interactions between inhibitory and excitatory neurons but does not thoroughly address possible roles of non-neuronal cells (e.g., astrocytes, oligodendrocytes, microglia) in circuit wiring, even though some of these cell types are included in the single-cell data. The authors should clarify whether such interactions were systematically explored or filtered out, and, if not, discuss this as a limitation and an avenue for future work.

The manuscript would benefit from a clearer description of how thresholds for significance and specificity in LR interactions were determined, and how robust these results are to parameter tuning. Furthermore, differences in cell numbers across clusters could affect the quantification of LR interactions. I recommend performing a downsampling analysis to assess how varying cell numbers per cell type influence LR inference and the stability of predicted interactions.

Minor Comments

Since the manuscript posits that specific ligand-receptor combinations encode the molecular logic for establishing cell type-specific cortical wiring, directly testing whether these LR expression patterns can distinguish between subtypes would provide strong support for this hypothesis. I suggest the authors explore to what extent neuronal cell types can be distinguished or classified based on their ligand and receptor expression patterns. For example, using clustering or machine learning approaches on LR expression profiles, the authors could assess whether cell-type identities can be accurately recovered from LR combinations and from which developmental stage on.

Line 103–105: Cell type classification is a crucial step that impacts all downstream analyses. The authors should provide more detailed information on the reliability of their label transfer, especially for immature neurons. For the ANN used for cell type classification, please report performance metrics such as ROC-AUC on the training dataset. Clarify how well the two different prediction methods agree with each other, and how many cells were excluded due to low prediction scores.

Lines 133–146: To characterize the identified gene expression waves, the authors could perform unsupervised GO enrichment analysis for each wave. It would be informative to analyze how many genes are shared across different cell types or classes within each wave, as this could reveal the extent to which a global maturation program is active versus cell type-specific expression dynamics.

Lines 153–160: The definition and biological interpretation of the “spatial axis” in this section require further clarification. The approach appears similar to the maturation axis analysis described earlier, raising the question whether the pseudo-layer score captures spatial information, or if it largely reflects developmental progression. Please clarify the methodology for constructing the pseudo-layer score and provide evidence that it identifies gene sets distinct from the maturation axis. If there is substantial overlap or limited added value — especially given that glutamatergic cell types are already defined by their laminar position — it may be worth reconsidering the inclusion of this analysis.

The LR-Coord diagrams are confusing in their current form and lack a concise description.

Please provide a clearer explanation of Figure 3a

Please clarify in the main text how morphological reconstructions were used to infer the spatial locations of cell types. Additionally, for interneurons of the same type or subtype that occupy different spatial positions (e.g., upper versus deep layers), please report and discuss which genes show differential expression depending on spatial location.

(Remarks on code availability)

The code and documentation are generally good. However, the Shiny app web tool has major usability and performance issues, which currently limit its value as a community resource.

Reviewer #3

(Remarks to the Author)
In this manuscript, Mathieu and colleagues used single-cell RNA sequencing (scRNA-seq) to characterise gene expression dynamics among developing neurons during the first month of postnatal life in the mouse somatosensory cortex. Using this comprehensive atlas, which they integrated with previously published datasets from the embryonic and adult cortex, they identified developmental trajectories for the main types of glutamatergic and GABAergic neurons. They then assembled a comprehensive map of putative receptor-ligand interactions and explored possible expression patterns underlying connectivity between glutamatergic and GABAergic neurons.

This is a very useful resource for identifying putative molecular interactions between distinct neuronal populations in the mouse somatosensory cortex. As such, this study will be of great utility to the community. I only have a few suggestions that help clarify some aspects of the manuscript.

Main issues

One important caveat that limits the impact of the findings reported by the authors is the assumption that specific receptor-ligand interactions are associated with known functions for a particular pair of molecules. For example, the authors conclude that "…processes like cell death or differentiation may rely more on intrinsic mechanisms" (lines 224 and 225). A similar assertion is made in the paragraph starting on line 282. This is counterintuitive because, for example, interneuron cell death is known to depend on interactions between excitatory and inhibitory cells. I suggest the authors take a more agnostic approach when linking specific ligand-receptor pairs to function. There is a danger in overplaying GO terminology, as it only relies on what we currently know. I don't have a problem with the author's focus on synaptic connectivity, but they shouldn't rule out that cell-cell interactions are important for other developmental processes during this period.

Using an adult connectivity matrix restricts the potential universe of ligand-receptor interactions in a practical manner. However, it assumes that the connectivity we see in the adult cortex is the same as during early postnatal development. Numerous examples indicate that this is not always the case, as transient connectivity is abundant during these stages. The manuscript should acknowledge this limitation, and relevant references should be added in this context.

The authors suggest that cell-cell interactions may play a more prominent role in regulating the connectivity of excitatory cells with SST and PV interneurons than with other cortical interneurons, such as VIP cells (lines 270-273). Are there alternative explanations? Could the number of cells or the stages sampled (which may not directly overlap with the wiring of VIP interneurons, for instance) limit the exploration of those interactions? In any case, conclusions like this (and other examples in the Results section), which are speculative, should probably be moved to the Discussion.

The authors should perform functional experiments (similar to those performed for CDH13) to confirm the role of NEO1 in the connectivity between Martinotti cells and excitatory neurons. In contrast to the other example (ErbB4-Nrg3), no functional evidence supports a role for NEO1-Cbln4 interactions in this context.

Minor

Line 257. I assume "CR-CR interactions" means interactions among Cajal-Retzius cells. Please revise the use of abbreviations throughout.

Line 270. Please consider Modol et al., Neuron 2024, in this context.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I really appreciate the authors' incorporation of both major and minor suggestions, and overall agree in the few cases where the authors rebut some minor suggestions.

scRNA-seq vs snRNA-seq
The comparison of scRNA-seq and snRNA-seq is very useful and the authors' motivation to include both and the manner of handling both is reasonable. Moreover, I appreciate the concession on page 12 that snRNA-seq failing to capture short transcripts can bias results.

Region-specific cell-types

The clarification and caveat on extrapolating findings to the rest of the cortex is well-stated. I appreciate the new branch of analysis on interneuron regional origin which the authors interpreted nicely.

Pseudo-spatiotemporal analysis
The finer detail on pseudolayer analysis is good, especially the careful descriptions of how MERFISH and patch-seq were used. The term “wave” instead of “gradient” is fine, but please make sure it’s used consistently. Figure 2E also looks very clean and interpretable now.

A deeper gene enrichment analysis would be a nice complement to the spatiotemporal clusters found in panel F, beyond just mentioning the processes of lamination and synaptogenesis in the text. It would be nice to label the spatiotemporal pattern clusters in panel F by their enriched functions.

A new concern is about the claim of immature signatures of upper vs lower layer neurons (page 8, the last paragraph of the “Spatial transcriptional gradients” section). If you claim that upper layer neurons show immaturity markers at E18.5-P8 and deep layer neurons do not, corresponding to their relative birthdates, you need to demonstrate that past P8, these maturity differences disappear. It would also help to demonstrate that deep layer neurons show the same/similar immaturity markers at earlier developmental stages if feasible.

New LR scoring metrics
Descriptions on page 12 of developmental score and connectivity score, and how they are used to filter (rather than manipulate) the results of scSeqComm, are much clearer. However, like Reviewer #2, I would appreciate a very explicit statement of how thresholds on the intercellular score were made. I understand that the developmental and connectivity scores helped to define this threshold, but the specific method for converting from a set of continuous scores to a binary LR pair (significant/non-significant) should be explicitly stated. I do appreciate the authors checking that using the intercellular score rather than a binary LR pair provides very similar results.

Limitations of transcriptomics
The limitations of transcriptomics for LR interaction studies on page 26 are a nice addition and provide sufficient skepticism to the claim that paired ligands and receptors will have correlated expressions over time.

LR specificity to cell type pairs
I really like the direction you took my suggestion of classifying cell type pairs based on ligand-receptor, and how you stratified by LR class (shared, unique, etc.) and developmental stage in Fig. S17. I can see how this is a very difficult classification problem, and the interpretation you derive provides very strong support to your claims about LR class contribution to cell type.

Source-target plots in Figures 3 and 4
The colormaps in Figures 3 and 4 are very colorblind accessible now. However, the light to dark scale makes small signals hard to see in some cases. In those cases (top right panel of Fig 3D, left panel of Fig 4C), a break can be added to the colorbar to solve this problem.

On that note, I recommend that the colorbar in Figure 4 refer to the number of LR interactions as in Figure 3, unless there is a good reason otherwise. An absolute number of LR pairs is clearer than “proportion of total ontology.” If you prefer using proportion of total ontology, it would be best to have the total number of LRs next to the colorbar.

That being said, in Figures 4C and 4D, I dislike how the color scale is normalized to the total LR ontology across all developmental stages rather than for each developmental stage. Using the same color scale for all developmental stages in Figures 4C and 4D exaggerates how the number of LR pairs increases with maturation, masking other trends. What’s more interesting than an increase in the number of LR pairs over time is how the configurations of LR interactions across cell types change over time. This can be accomplished simply by setting the max color value for each column (developmental stage) independently. That would remove the trend of an increasing number of LR interactions over time and leave us with the residual timepoint-specific patterns for emphasis.

Another concern: can the increase in the number of significant LR pairs over developmental time (Figure 4A) be explained by some confounding feature of the datasets? For instance, could more cells in older datasets explain why there are more significant LR pairs? It would be easy to alleviate this concern by repeating your analysis multiple times on subsetted data to balance the number of cells at each developmental stage and see whether the trend in number of significant LR pairs holds.

Lastly, I’m not quite convinced about the claim on page 15 about how DL GABANs show higher LR engagement than UL GABANs early on, as Figure 4B makes this trend seem very slight. Statistically testing the relationship between depth and rate of LR engagement would provide much better support to this claim.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The authors have addressed all of my previous concerns in a satisfactory manner. The revisions improve the clarity of the manuscript and strengthen the conclusions. The additional analyses are appropriate and support the main claims. I do not have further comments.

I recommend the manuscript for publication in its current form.

(Remarks on code availability)
I examined the provided code. The code is clearly organized. The code should be usable by others in the community with standard computational experience. I have no concerns regarding reproducibility.

Reviewer #3

(Remarks to the Author)
This is a very nice study and an important resource. The authors have done a good job addressing my comments, and I have no further suggestions.

(Remarks on code availability)

Reviewer #1 (Remarks to the Author):

The manuscript by Mathieu et al. sets out to address the fundamental question of how synaptic specificity is determined during mouse brain development. To do so, the authors compile single cell RNAseq datasets from developing and adult mouse cortex, conduct bioinformatic analyses including pseudotime, “pseudo-layer”, and ligand-receptor predictions explain cell-cell interactions during corticogenesis. Overall, this is a very interesting analysis, validated by the prediction of known patterns of cell cell interactions, augmented by perturbational experiments involving short hairpin RNA knock-down of candidate cell contact molecules in vivo mouse brain. I believe that this analysis and its results will be of great value and interest to the community, and that it merits publication.

We thank the reviewer for his/her detailed evaluation and positive appreciation of our work.

The authors collected scRNAseq and snRNAseq data from SS cortex at P0-P30.The pipeline they describe for integrating data from multiple sources seems reasonable and conventional, but nonetheless it would be helpful to validate if any insights differ between nuclear versus whole cell sequencing.

We thank the reviewer for this comment. We chose snRNA-seq primarily for later developmental ages, which are underrepresented in publicly available datasets. At these stages, scRNA-seq is known to result in unbalanced cell-type distributions due to high levels of cell death. To directly assess whether nuclear versus whole-cell sequencing impacts our conclusions, we performed a systematic comparison at P8, a stage where both scRNA-seq and snRNA-seq were available. Specifically, we compared clustering outcomes, cell-type proportions, and gene expression profiles. The UMAP projections (Fig. S15A) revealed consistent clustering across both modalities, with all major excitatory and inhibitory subtypes well represented. The relative abundance of cell types was also preserved (Fig. S15B), with only modest differences in proportions. Gene expression was strongly correlated between modalities (Pearson r = 0.72, Fig. S15C). Finally, we compared ligand–receptor (LR) interaction inference across modalities under identical filtering and scoring criteria, ensuring equal sampling size across conditions. Both scRNA-seq and snRNA-seq recovered highly overlapping LR pairs (Fig. S15D. Taken together, these analyses demonstrate that scRNA-seq and snRNA-seq yield highly comparable results in our framework, with snRNA-seq offering the key advantage of profiling late developmental stages where whole-cell dissociation is not feasible. These new control are now described in Results p12.

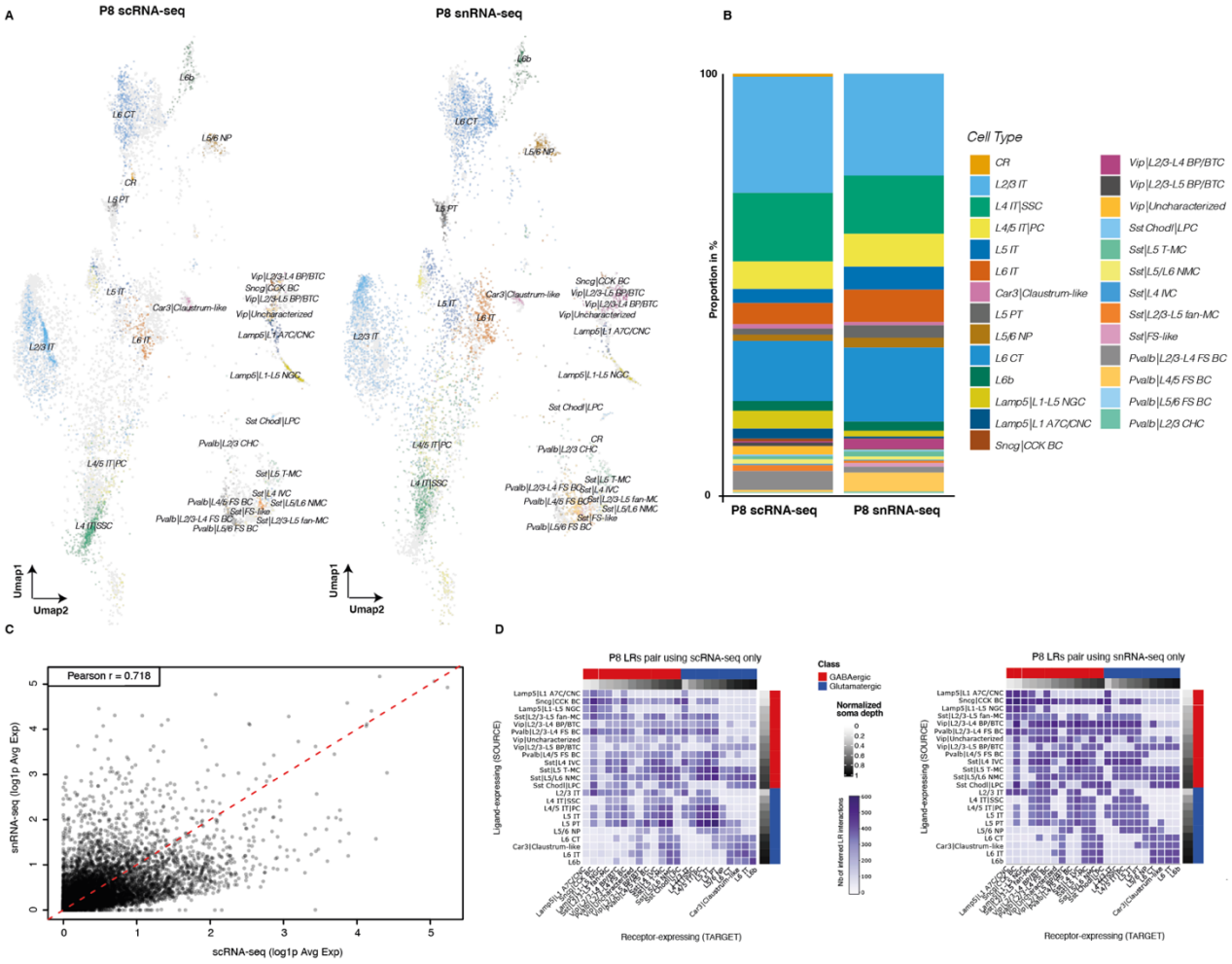


Figure S15. scRNA-seq and snRNA-seq datasets comparaison. (A) UMAP projections of single-cell (left) and single-nucleus (right) RNA-seq datasets from postnatal day 8 (P8) mouse cortex. Cells are colored and annotated at cell type level. **(B)** Stacked bar plots showing the proportional representation of cell types detected by scRNA-seq (left) and snRNA-seq (right). While the overall diversity of populations is preserved, relative subtitle abundances differ between methods. **(C)** Correlation of average gene expression between scRNA-seq and snRNA-seq across all cell types. Each dot represents a gene, with the red dashed line indicating the perfect correlation. Pearson correlation coefficient is r = 0.718, demonstrating strong concordance between the two modalities. **(D)** Ligand–receptor (LR) interaction analysis at P8. Heatmaps show predicted LR pairs identified using scRNA-seq only (left) or snRNA-seq only (right). Rows represent ligand-expressing clusters and columns represent receptor-expressing clusters, with signal strength indicated by color intensity. GABAergic and glutamatergic classes are highlighted, and normalized soma depth is provided on the side bars. No major differences are observed between scRNA-seq and snRNA-seq predictions.

The authors should discuss to what extent their analysis, which is focused on SS cortex, could be generalizable to other cortical regions.

We thank the reviewer for raising this important point. As reported in Yao et al. (2021), most GABAergic clusters defined in the Allen Institute taxonomy are shared across isocortical regions (see Fig. 4c and 4h in Yao et al., Cell, 2021), and the relative proportions of cells within these clusters are largely consistent across areas. For glutamatergic cell types, there is a modest, gradual transcriptomic variation across regions, and some degree of regional specificity is observed, particularly at the cluster level and in isocortical areas located at the rostral and caudal extremes. In contrast, adjacent cortical regions tend to share the majority of clusters.
Importantly, our analysis is anchored at the supertype level; a higher-order classification that is more stable than individual clusters and broadly conserved across cortical areas. As described in Yao et al., supertypes are consistently observed across the isocortex, with only a few known exceptions, such as the L5 PT and Car3 supertypes. Since the cell types in our study were assigned at the supertype level, we believe that our findings are likely to generalize neighboring cortical regions.
That said, we emphasize that our analyses were specifically conducted in the somatosensory cortex. Any extrapolation to other cortical areas should therefore be made with appropriate caution. We now clarify this point in the revised manuscript p 26.

Relatedly, whether progenitor domain of origin of interneurons plays any significant role in the types of interactions predicted by their algorithm should be further interrogated and discussed.

We thank the reviewer for this suggestion, which provided important new results that further strengthen our conclusions. Specifically, we tested the hypothesis that GABAergic subtypes derived from the MGE, CGE, or POA exhibit either similar or distinct predicted ligand–receptor (LR) interactions with their glutamatergic counterparts in the pallium. To address this, we first regrouped all cell types by their ganglionic eminence of origin (MGE, CGE, or POA) and all glutamatergic cells by pallial VZ origin. We then quantified, at each developmental stage, the fraction of LR pairs that were either shared across at least two GEs or specific to one GE, by comparing VZ pallium–MGE, VZ pallium–CGE, and VZ pallium–POA interactions.

As summarized in Fig. S23A, our analysis revealed a dynamic temporal pattern of GE-specific versus shared LR modules. The fraction of shared LR interactions increased from embryonic stages to a peak at P8, followed by a progressive decline toward adulthood, reflecting a transition from early GE-specific programs to common communication modules, and then a return to specificity later in life. Fig. S23B further shows the distribution of individual LR pairs across development, with clear segregation of GE-specific versus shared modules. To determine whether this segregation reflected random partitioning or functionally distinct gene modules, we performed enrichment analyses of shared versus GE-specific genes at each age. This revealed distinct ontologies unique to GE-specific modules that were not observed in the shared sets, highlighting biological processes tied to the developmental origin of interneurons (Fig. S23C). Importantly, while early GE-specific functional modules (E18–P5) were enriched for programs linked to partner recognition and cortical integration, these converged into shared programs between P8 and P30, dominated by generic processes such as synapse organization and neuronal signaling. In adulthood, GE-specific modules re-emerged, suggesting renewed specialization and potentially subtype-specific synaptic communication.

Together, these results demonstrate that LR modules are both fate- and stage-dependent. Early GE-specific LR programs may facilitate partner recognition during migration and initial circuit entry, mid-developmental shared programs likely support global synapse formation, and adult specificity may reflect fine-tuned, lineage-dependent

communication. This dynamic pattern of LR modules is a unique feature of GE-derived interneurons and underscores the developmental logic of their integration into cortical networks. These new findings are now described in Results, section p13, and further discussed in Discussion, section 25-26.

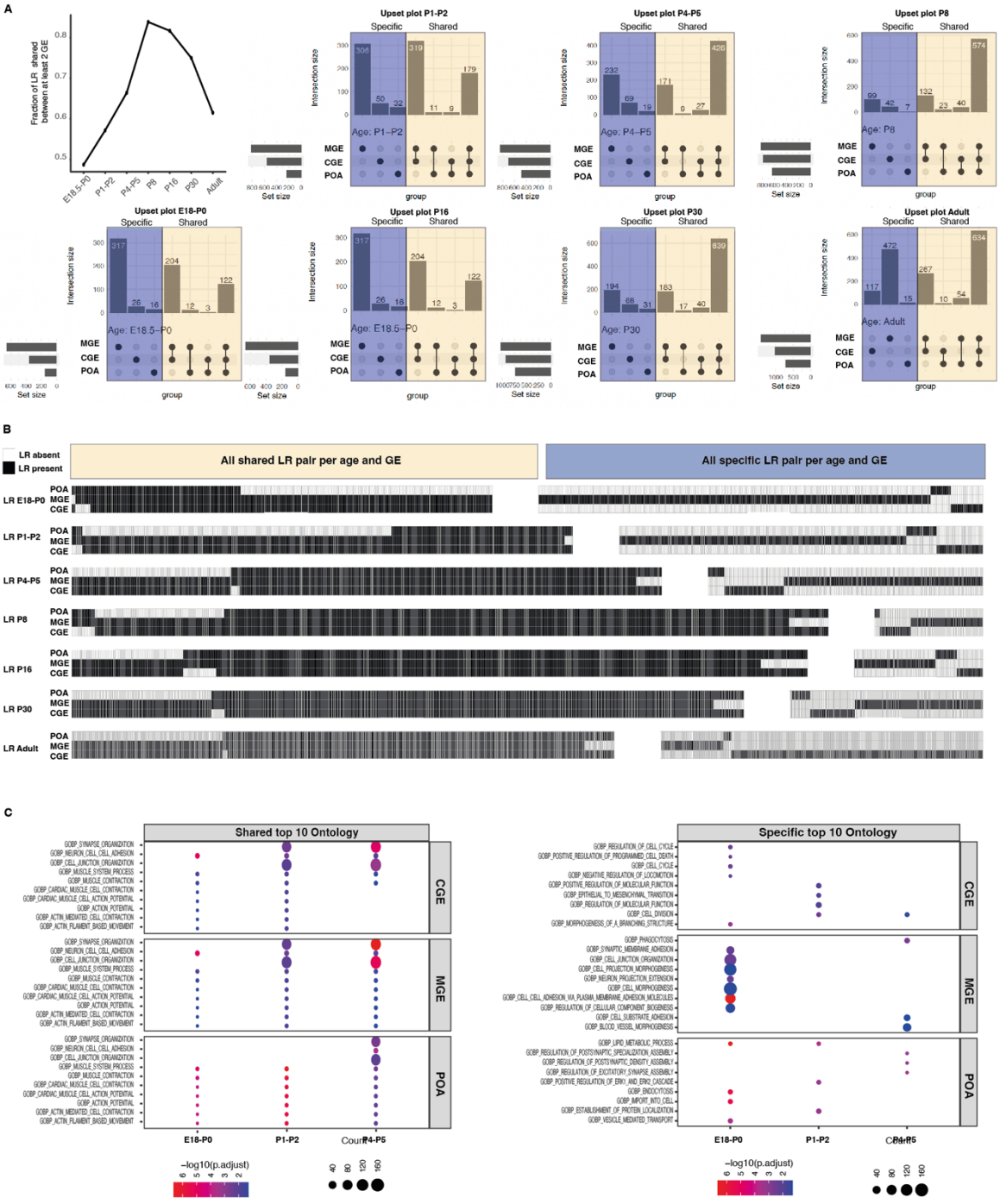


Figure S23. Ganglionic Eminence show specific and shared ligand–receptor modules across cortical development. (A) Temporal dynamics of sharing versus specificity of predicted ligand–receptor (LR) interactions, (left) line plot showing the fraction of LR pairs that are shared by at least 2 ganglionic eminence (GE) lineages (MGE, CGE, POA) at each developmental stage (E18.5–Adult) when Target cell type is Glutamatergic pallial cells. Right, for each stage, UpSet plots partition significant LR pairs into GE-specific (left panels) and shared (right panels). Bars on the left of each UpSet plot indicate the size of the LR sets for Source: MGE, CGE, and POA and Target (Glutamatergic pallial cells); intersection bars show the number of LR pairs common to the indicated GE combinations. **(B)** Presence/absence "barcode" maps of LR pairs across development. Left block: all shared LR pairs (columns) with rows for POA, MGE, and CGE at each stage; black = present, white = absent. Right block: all GE-specific LR pairs, displayed in the same format. This panel highlights the switch from early

GE-specific modules to a more shared program in mid-development, followed by a re-emergence of GE specificity in adulthood. **(C)** Gene Ontology enrichment for LR modules. Left, shared LR modules; right, GE-specific LR modules. For each GE and stage, the top 10 GO Biological Process terms are shown. Dot color indicates –log10(adjusted p value) and dot size indicates the number of LR genes contributing to the term. Enrichments emphasize synapse- and adhesion-related processes among shared modules and GE-biased functional programs among specific modules**.**

Pseudo-spatiotemporal analysis

The authors performed pseudotime and pseudolayer analysis to build smooth models of gene expression across developmental time and space. The pseudotime analysis using slingshot is conventional, but the pseudolayer method is less clear. The description of the pseudolayer method in the main text mentions cross-referencing their data with patch-seq and MERFISH datasets, but there is no further explanation (not even in the methods) of how these datasets are used to estimate pseudolayer. It seems that their approach to pseudolayer estimation relying on spatial transcriptional gradients could be further discussed. The authors should consider validating their method with a more established method for spatial estimation from scRNAseq data like tangram.

We thank the reviewer for this helpful comment, which highlighted the need for clarification. We recognize that the main text was not sufficiently explicit and that some methodological details were not optimally placed. To clarify: cross-referencing with Patch-seq and MERFISH was not used to construct the pseudolayer itself, which is based purely on transcriptional profiles (hence the term "pseudo"). Nor was it required to establish the relative positions of cell types within families, since these could already be inferred from our nomenclature, which incorporates predominant laminar location based on the correspondence between Patch-seq and the reference taxonomy in Yao et al. Instead, Patch-seq and MERFISH cross-referencing was indispensable only for assigning exact (rather than relative) normalized laminar positions, information required for calculating the connectivity score in our ligand–receptor inference analysis.

To improve clarity, we now explicitly describe in the Results (p7-8) how these laminar positions were assigned and have relocated the corresponding Methods subsection (p 33-34) so that it directly follows the pseudolayer analysis, ensuring better logical flow. We also thank the reviewer for suggesting Tangram. As the mapping between MERFISH data and the transcriptomic taxonomy we used has already been published, we relied on this existing resource to interpret spatial positions (see Fig. S2D) and therefore did not find it necessary to perform an additional Tangram-based projection. We will clarify this rationale in the revised manuscript.

The "waves" estimated from pseudotime and pseudolayer models, with shared dynamics of gene activity across time/space, demonstrate that the pseudotime and pseudolayer results are meaningful. However, I can't find a description of how these waves are computed in the methods. Conversely, there's no description of fitting the GAM in the main text, but there is in the methods. Small detail: it would be clearer to refer to the pseudolayer dynamics as something like "upper and lower gradients" rather than "first and last waves."

We thank the reviewer for this helpful comment. We would like to clarify that the procedures used to compute the genetic waves were already described at the end of the "Pseudo-maturation analysis" and "Pseudo-layer analysis" sections of the Methods . To enhance clarity and accessibility, we have now added explicit subheadings; "Temporal gene wave computation" (p. 32) and "Spatial gene wave computation" (p. 33); to clearly highlight this content. We believe that "wave" better captures the biological dynamics than "gradient," but we are happy to adjust the terminology further if the reviewers prefer.

The depiction of the developing cortex colored by cell-type in Figure 2E is useful, but too crowded and should be simplified.

Thank you for this helpful suggestion. We agree that the original version of Figure 2E was overly crowded. We have now simplified the depiction of the developing cortex by reducing the density of labels and streamlining the color scheme to improve clarity, while still preserving the key cell-type information.

LR-interaction analysis

The authors assembled the largest existing LR database and used a modified version of scSeqComm to infer magnitudes of cell-cell interactions using more conventional LR correlations between cell types (intercellular score) and less conventional pathway activation where such pathways are known (intracellular scores). With this

approach, they aimed to validate their model with known LR pairs and by discovering novel LR interactions which they validated with a knockout experiment. These results give credence to their data processing and analysis pipelines.

We thank the reviewer for their positive assessment and appreciation of our work. We are encouraged that the reviewer recognizes both the value of assembling the largest LR database to date and the robustness of our analytical framework.

However, a major concern with the author's validation approach is how they modify the intercellular score calculation by adding information about developmental timing and connectivity patterns from patch-seq dataset (lines 231-239). It seems that they use this score to predict connectivity between cell types which were classified using the patch-seq datasets. That is, it seems like information about the validation data entered the model during training. The authors should clarify that this was not the case. Further, they should show the implications of their modifications to the scSeqComm technique and explain why their modification to the intercellular score is necessary and valid.

We thank the reviewer for raising this important point and for the opportunity to clarify. First, we emphasize that neither the intercellular nor the intracellular scores generated by scSeqComm were modified in any way, and the core algorithm itself was not altered. The intercellular score was calculated exactly as described in the original publication and was returned for all possible LR pairs across all cell-type combinations.

The additional connectivity score and developmental score were used only afterwards as biologically informed filters to prioritize among the large number of potential LR interactions predicted by scSeqComm. In line with the recommendations of the scSeqComm authors, applying thresholds to intercellular scores is necessary to define significant interactions. Rather than using an arbitrary cutoff, we integrated orthogonal information on connectivity potential (from Patch-seq–based experiments) and developmental timing. This prior knowledge was not used to generate or "train" the intercellular scores themselves, but only to increase stringency and highlight interactions most likely to occur in vivo at the relevant developmental stages.

We believe this refinement is both necessary and valid: by constraining the predictions to biologically plausible interactions, we reduce false positives and increase interpretability, while still retaining the full scSeqComm output for transparency. Importantly, we experimentally validated several of these filtered interactions (Figs. 5–6), supporting the utility of this approach. At the same time, we wish to stress that our analysis is not intended as an exhaustive catalogue of all valid LR interactions. Rather, it represents a framework and resource to help identify the most promising candidate interactions, which will ultimately require further functional validation.

We have now clarified these points more explicitly in the revised manuscript (p. 12).

The authors make a core assumption (lines 208-211) that the transcription of GlutN ligands and GABAN receptors should be correlated between synaptically connected GlutNs and GABANs. Though the converse of this statement seems true, the statement itself does not seem true necessarily. Given two connected neurons, it's possible that their transcription of an LR pair is uncorrelated across time, because the ligand and/or receptor may have roles outside of the time period relevant for the connectivity of these two neurons. The authors should discuss the potential limitations of the method.

We thank the reviewer for this thoughtful comment. We agree that for a given ligand–receptor (LR) pair, transcript levels for the ligand and receptor do not necessarily correlate over time. For example, a ligand transcript may be present at P8 but absent at P16, while the receptor transcript shows the opposite pattern. Yet, if the ligand protein is stable, both proteins could coexist at P16 and mediate communication. Our current approach, which focuses on correlations at individual time points, would miss such interactions. We now acknowledge this limitation in the manuscript (p. 26). At the same time, we view this stringency as an advantage: when an LR interaction is transcriptionally predicted, it is more likely to reflect a real event. While transcript co-expression at the same time point does not guarantee protein interaction, lack of co-expression across consecutive stages provides even weaker evidence for interaction at later stages.
Conversely, as the reviewer also points out, concomitant transcription of a ligand in one cell type and its receptor in another does not in itself imply functional communication. We now discuss this as well (p. 26). Transcript abundance is only a proxy: it may not reflect protein levels, and even if both proteins are present, functional interaction depends on trafficking and localization to appropriate subcellular compartment; information that single-cell RNA-seq cannot provide. For instance, receptors may remain intracellular, or ligands may fail to be secreted, preventing interaction.

The authors observe that very few LR combinations are unique to a single cell-type pair, while many combinations are shared between pairs. From this observations, they claim that a "combinatorial code" of common LR pairs drives synapse specificity. This hypothesis is enticing, but it could be much more strongly supported with statistical tests. Specifically, the authors should train a model to predict cell type pair from LR interaction scores, and compare the performance of this model when common or rare LR pairs are included as features. If the model using shared LR pairs outperforms the model using rare pairs, this finding would support their hypothesis.

We thank the reviewer for this suggestion. We would like to clarify that our interpretation does not imply that shared LR pairs alone drive synapse specificity. In fact, our analysis shows that specificity arises from the unique combinatorial patterns of both shared and context-dependent LR pairs used by different cell-type pairs across development. Truly unique LR pairs or rarely share are limited, but the way rare and widely shared pairs are assembled together differs markedly between cell-type pairs, and we interpret this as a combinatorial code. Instead, our results indicate that specificity is encoded in the joint usage of LR subsets across cell-type pairs, rather than in shared or rare pairs considered in isolation.

After this clarification we want appreciate the reviewer's idea to test the "combinatorial code" hypothesis with a predictive model which is really elegant. To note directly training a classifier on the native scSeqComm output is not feasible: scSeqComm returns one summary value per cell-type pair for each LR pair (i.e., no within–cell-type replicates), so any supervised model would overfit by construction; each (LR, cell-type pair) appears once. To address the spirit of the suggestion without violating this constraint, we built an alternative machine-learning set-up that creates single-cell replicates for every cell-type pair. Concretely, for each age we expanded every SOURCE→TARGET pair into all single-cell combinations (source cell × target cell). For a chosen LR feature set, each example is a vector of ligand×receptor expression products (log1p and z-scaled), and the label is the SOURCE→TARGET cell-type pair. We evaluated four LR feature sets: (1) all significant LR pairs; (2) "shared" pairs (highly reused); (3) "rare" pairs (≤10 cell-type pairs); (4) "unique" pairs (used once); plus a size-matched false LR list (random gene–gene pairs). To avoid information leakage, the train/test split was done by cells (no source or target cell used in training appears in testing). We trained a regularized MLP classifier (with and without weight pruning) and report accuracy per age (example in Fig. SX; dashed = unpruned, solid = pruned). Models trained on unique, shared-only, or fake LR lists perform at (or very near) chance across ages. In contrast, models using rare-only LR pairs consistently exceed chance and contribute meaningful predictive signal. Crucially, the best performance is obtained when combining LR sets (all significant = shared + rare + occasional unique), supporting our central claim: specificity is not carried by shared pairs alone nor by unique pairs alone, but by their combination, with "rare" LR pairs providing discriminative power that sharpens cell-type-pair identity. This pattern holds across developmental stages (see accuracy-by-age panels: shared/unique/fake ≈ chance; rare > chance; all = best). We have add the supplement figure (Fig. S17) and clarify in the Results (12-13), Discussion (p25) and Methods (p 40-41).

One more point of context: this is an extremely hard prediction task (>600 classes), so overall accuracy is expected to be low simply because the chance baseline is tiny. Despite that, many individual cell-type pairs achieve high ROC–AUC values, showing strong separability well above chance even when the aggregate accuracy is modest. This pattern reinforces that the model captures real discriminative signal; consistent with a combinatorial LR code; rather than random fluctuations.

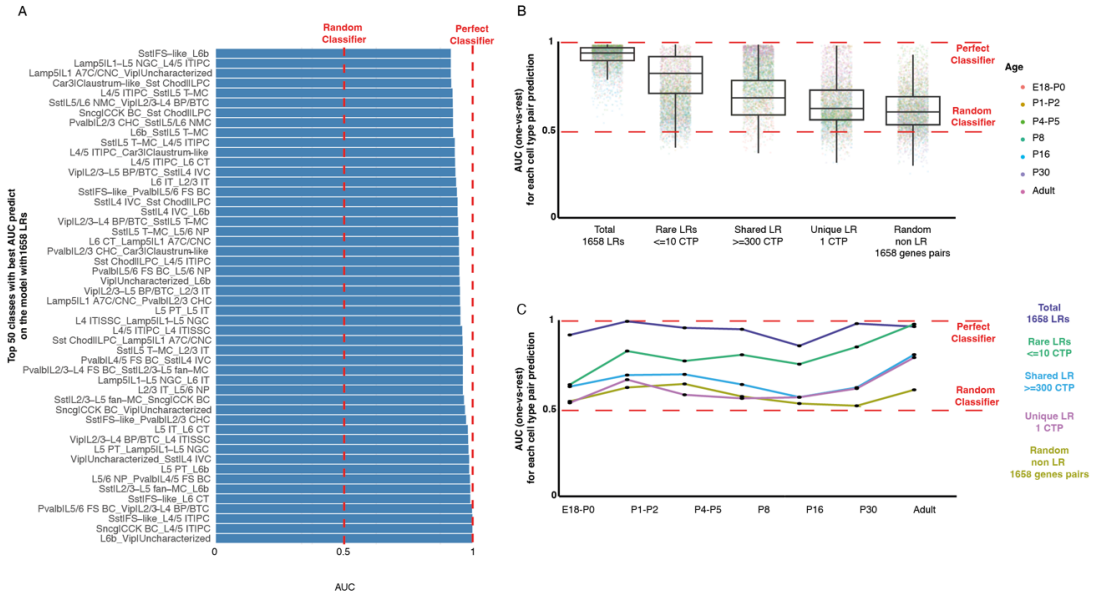


Figure S17. Predicting source–target cell-type pairs with machine learning reveals specificity encoded by combinatorial ligand–receptor patterns.
(A) Classification performance for the top 50 cell-type pairs (CTPs) predicted using all significant LR pairs. Bars show ROC–AUC values (one-vs-rest), with most CTPs reaching well above chance (red dashed line, random classifier) and approaching high separability (perfect classifier, AUC = 1).
(B) Comparison of predictive performance across different LR feature sets. Models trained on all significant pairs perform best, while models restricted to shared or unique LR pairs perform at or near chance. Models trained only on rare LR pairs (≤10 CTPs) consistently exceed chance and contribute meaningful signal. Randomized non-LR gene–gene pairs do not provide predictive power. (**C)** Developmental trajectory of classifier performance across ages (E18–P0 to Adult). Rare-only LR pairs show reproducible predictive signal above chance across stages, while shared or unique pairs alone remain near chance. Models trained on the full LR set (shared + rare + occasional unique) achieve the highest performance, consistent with specificity arising from the joint usage of different LR subsets.

When quantifying LR interactions during synaptogenesis and in developmental diseases, the authors use numbers of interacting LR pairs as a metric of cell-cell interaction (Figures 3 and 4). However, this metric assumes that every LR pair has the same degree of effect on cell function. It's unclear why they don't use the available intercellular and intracellular scores instead. Also, in Figures 3 and 4, the white-red-black colorbar indicating number of LR pairs is not very colorblind friendly. If an intermediate color is needed, green or orange (as long as it differs from the colors indicating GlutN and GABAN) could be used. The "normalized soma depth" colorbar should be changed to a different scheme than number of LR pairs, if it is even needed (seems redundant).

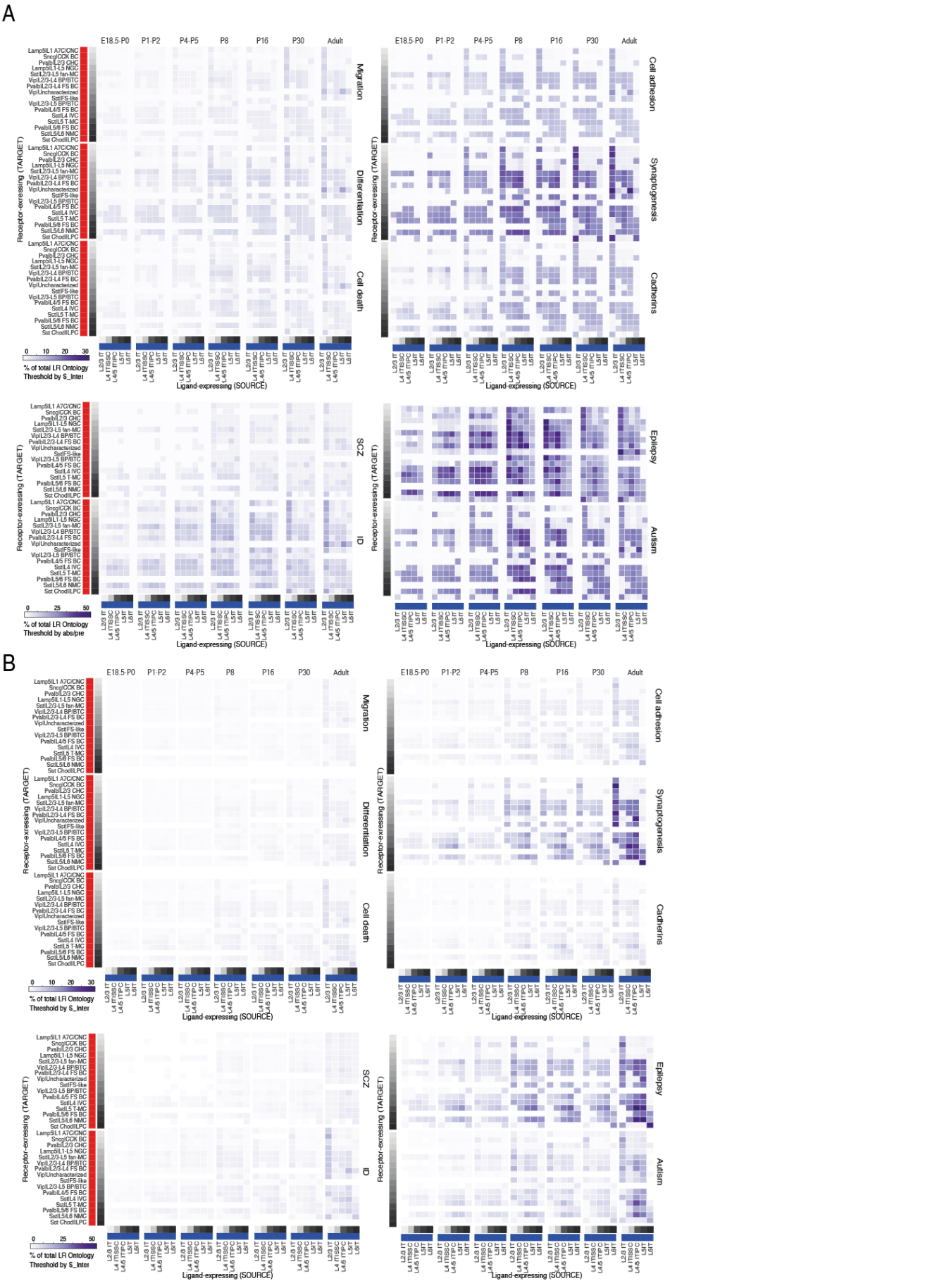
We thank the reviewer for raising this, because it is central to interpretation. After re-running the analyses with the intercellular score (S_inter) instead of presence/absence, we observed the same trend of qualitative patterns emphasized in the manuscript's Results. Together, these S_inter-based results support the same interpretation as the count metric: at the cell-type–pair level, the binary detection of a significant LR already captures most of the informative signal, with score magnitude adding little power sensitivity to these aggregate trends.

Using S_inter, we recover the same patterns: migration, differentiation, and cell-death LRs are sparse across development, whereas synaptogenesis (and to a lesser extent adhesion) are robust from E18.5 through adulthood, with cadherins rising sharply from P8. Disease-linked profiles are also preserved: autism/epilepsy LRs show high utilization (peaking at P8 and sustained), while SCZ/ID LRs remain consistently low.

In practice, this concordance indicates that, at the cell-type–pair level, the binary detection of a significant LR interaction already captures most of the informative signal. Biologically, this is consistent with a threshold-like behavior in which once an LR pair is expressed above a minimal significant level, additional increases in ligand or receptor transcript abundance contribute relatively little to these aggregate patterns. We therefore keep the presence/absence representation in the main figures for clarity and interpretability, and we now show the S_inter versions in the Supplement and state explicitly that both yield the same qualitative conclusions.

We also revised the figures and all related Suplemental to address accessibility. The white–red–black scale has been replaced throughout by a single-hue purple palette that is color-blind friendly and visually distinct from the colors used to denote GlutNs and GABANs and the one for normalized soma depth.

Figure for rebuttal: Comparaison of LR pairs based on neurodevelopmental ontologies and diseases using S_inter or Pre/Abs input. (A top) Percentage of LR pairs of 5 main neurodevelopmental ontologies and of the cadherin LR family that are predicted as utilized between IT and GABAN *cell-type*s as source and target cells, respectively, from E18 to adulthood. (A Bottom) Percentage of LR pairs with both ligand and receptor associated with neurodevelopmental diseases that are predicted as utilized between IT and GABAN *cell-type*s as source and target cells, respectively, from E18 to adulthood. Using either Pre/Abs LR pairs as input matrix (A) or S_Inter treshold for LR pairs as input matrix.


A
B
E18.5-P0
P1-P2
P4-P5
P8
P16
P30
Adult
Migration
Differentiation
Cell death
Cell adhesion
Synaptogenesis
Cadherins
SCZ
ID
Epilepsy
Autism
Receptor-exressing (TARGET)
Ligand-expressing (SOURCE)
% of total LR Ontology
Threshold by S_Inter
Threshold by abs/pre

Lastly, the result that CDH13-CDH13 interactions largely take place between Sst/Pvalb and L5/6 (conveyed lines 364-367) isn't clearly conveyed by the corresponding Figure 6A-C. It's not clear what the red boxes in Figure 6A are highlighting, and it seems like there are other GABAN types (like VIP) which have CDH13 interaction with L5/6 GlutNs.

We thank the reviewer for this observation. Our main point regarding CDH13-CDH13 interactions was to highlight the specificity of interactions between the PVALB subclass, specifically Pvalb|FS BC, and deep layer (DL) glutamatergic neurons, as opposed to upper layer (UL) glutamatergic neurons. We agree with the reviewer that CDH13-CDH13 interactions are not exclusive to the PVALB subclass. We concur that it would be valuable to further explore CDH13-CDH13 interactions across other GABAergic subclasses in future analyses.

The fact that the analysis considers only interactions between neurons as opposed to glial cells should be reflected in the title.

We thank the reviewer for this comment. We now change the title to "Uncovering the Molecular Logic of Cortical Wiring Between Neuronal Subtypes Across Development via Ligand–Receptor Inference".

Reviewer #2 (Remarks to the Author):

Summary Paragraph

This manuscript presents a single-cell transcriptomics atlas spanning the development of the mouse somatosensory cortex, integrating both newly generated and publicly available single-cell and single-nucleus RNA-seq datasets from embryonic to adult stages. This efficient approach leverages existing large-scale datasets, making optimal use of available resources. The authors assembled and curated a large ligand-receptor (LR) pair database, which they applied to their atlas to systematically infer putative LR interactions between transcriptomically-defined neuronal groups, with a focus on connections between inhibitory and excitatory neurons during circuit assembly. Some predictions are experimentally validated, including that neogenin-1 is the main receptor for CBLN4 during early synapse formation, and that cadherins CDH13 and PCDH8 have different, layer-specific roles in regulating inhibitory synapses onto deep and superficial neurons. A rich, interactive web app is provided (though currently not practically usable due to severe performance issues), potentially enabling broad access to the spatiotemporal transcriptomic data and LR predictions. A key limitation of transcriptomic data is its lack of subcellular localization information for ligands and receptors, which is essential for proper excitatory and inhibitory neuronal signaling.

This work provides a valuable reference for researchers studying cortical development and neuronal circuit wiring. The computational and bioinformatic analyses are well executed, and targeted experiments provide direct validation for selected predictions. I have a few major and mostly minor points that need to be addressed. If these are resolved, I support publication.

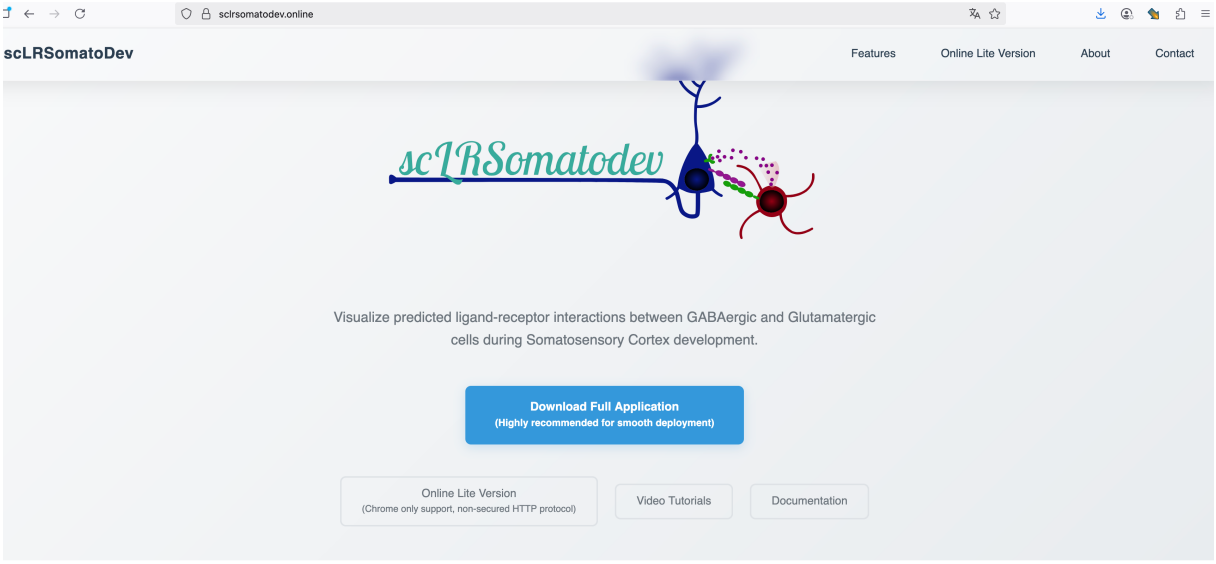
We thank the reviewer for his/her detailed evaluation and positive appreciation of our work.

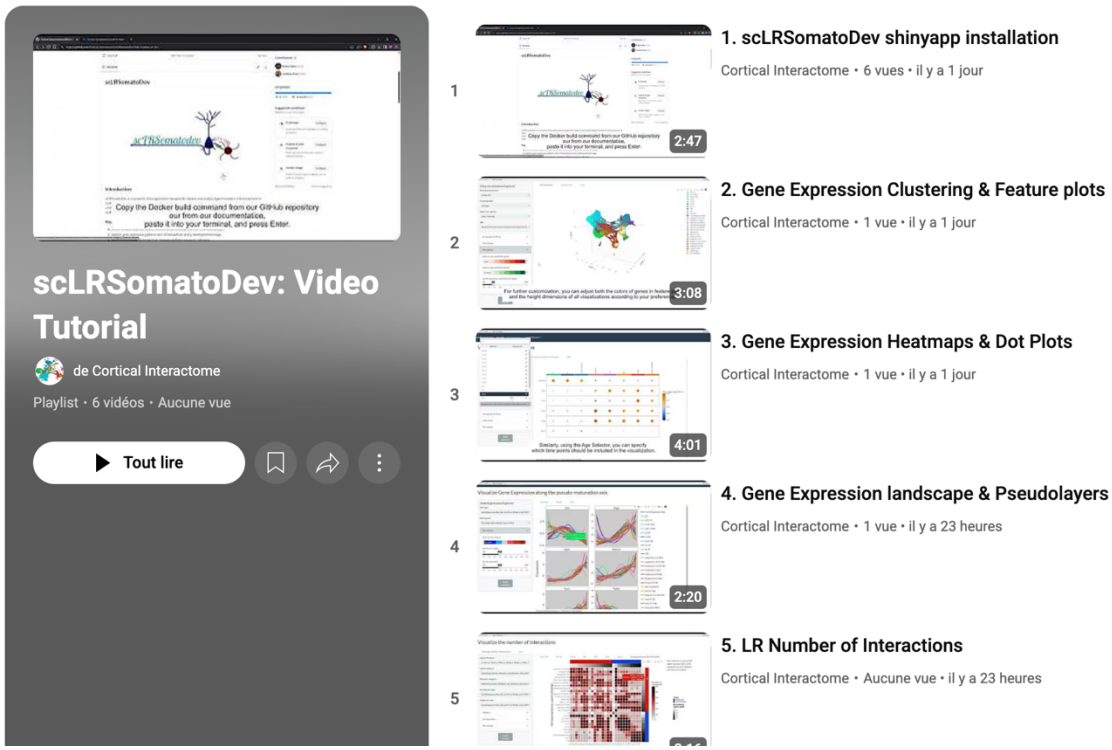
Major Comments

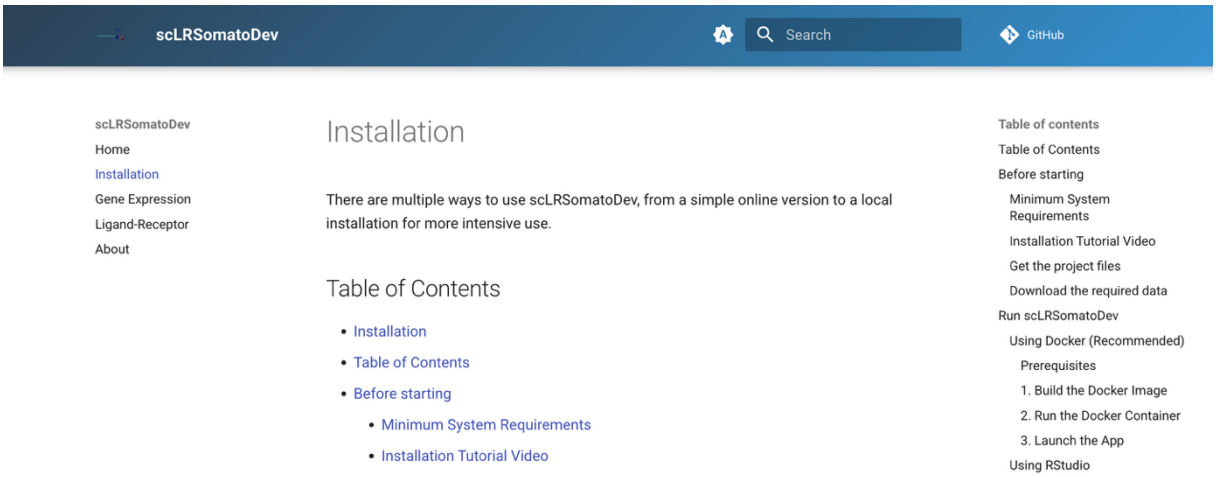
The interactive web resource, meant to be a key output of the study, is currently not functioning very well. It reacts slow and several parts don't load at all. Full release of this tool is important for maximizing the impact of the work.

We thank the reviewer for their feedback and acknowledge that the hosted Shiny app can experience slow response times and occasional loading issues. We would like to clarify that these performance problems are primarily related to the hosting server environment rather than the app itself, as the current server has limited resources.

For optimal performance, the app should be installed and run locally using our provided Docker file. In this setup, it operates smoothly, with acceptable speed for rendering plots and navigating the interface. Since the app mainly accesses pre-loaded data without intensive computation, its RAM usage remains still significant (20Go), and the data storage size does not significantly impact runtime performance. Our team has also worked to optimize the hosting server to improve user experience. The source code will be publicly available on GitHub, accompanied by comprehensive installation instructions, we have now made multiple video tutorials, and a dedicated documentation website (see print screen bellow). The application is designed for local installation with minimal system requirements and has not been optimized for scalable deployment on web services. After publication, it will be easier for us to obtain support from the University for such deployment.
For users who require only basic functionality and may not want to follow the local Docker setup; we have now developed a Lite version of the app and release this one online only support by "chrome browser without HTTPs support only HTTP". This version focuses on essential features and is optimized to run effectively within the limitations of our current hosting infrastructure (to be confirmed in practice). The source code will be publicly available on GitHub, accompanied by comprehensive installation instructions, multiple video tutorials, and a dedicated documentation website.
We believe that by addressing these bottlenecks, users will be able to fully realize the app's potential.





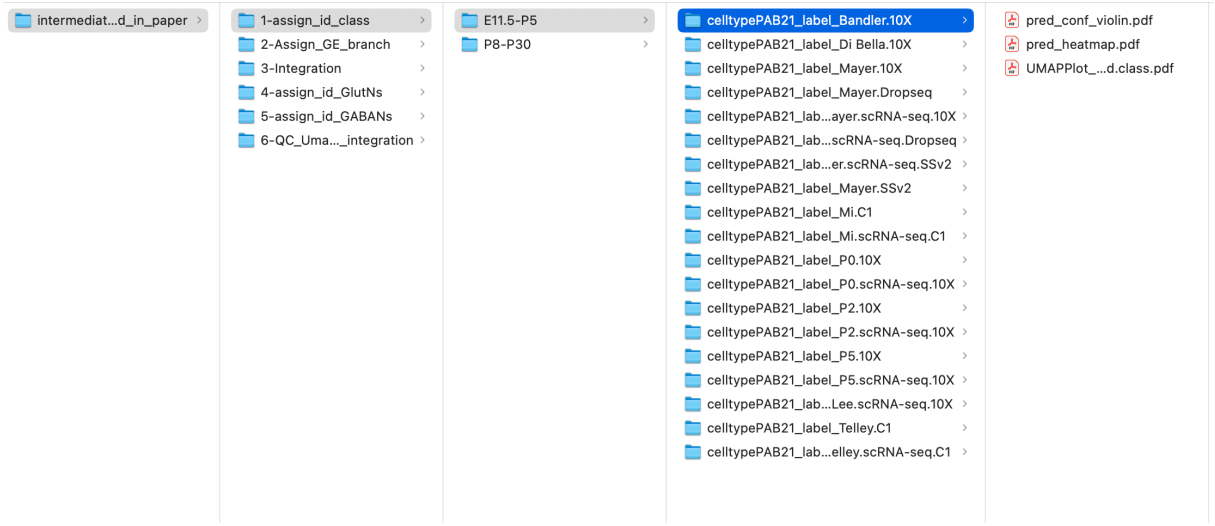


While the integration of multiple single-cell datasets is a strength, the manuscript lacks critical analysis of batch effects, dataset biases, or possible artifacts arising from combining data produced under differing experimental protocols. The authors should address how integration choices may impact downstream LR inference. The manuscript would benefit from providing intermediate steps and results for data integration and label transfer, enabling readers to better assess the accuracy and reliability of the integration process and cell-type annotation. The authors could explore the integrated dataset more thoroughly by presenting developmental trajectories, UMAPs labeled by developmental stage (as shown in Di Bella et al.) , and cell type abundances across stages.

Thank you for raising this, it's an important point when aggregating heterogeneous single-cell datasets. First, we agree that integration choices can bias downstream analyses if not handled carefully. In our workflow, joint embeddings were used only for visualization and label transfer, whereas LR inference was run on library-size–normalized expression within each dataset (log-scale CPM) using scSeqComm, precisely to avoid injecting cross-dataset batch corrections into the quantitative scores. In other words, batch integration informed who is what (cell-type/age labels), but not how strong an LR pair is which was computed from the underlying expression matrices. This separation reduces the risk that over-correction or alignment artifacts propagate into LR estimates. Second, we benchmarked multiple integration/mapping strategies and chose the one that best preserved known biological structure while minimizing over-correction. Class assignment however is the critical step for our LR analysis , to make the processus transparent, we have now provided intermediate steps and QC. Because the number of datasets and plots is large, we have deposited the full set of intermediate objects and figures (embeddings, metadata tables, label-transfer reports, and per-step parameters) on Zenodo (url bellow) which will be release publicly after publication. Key snapshots of the folder organization is included bellow for your convenience. We had clarified this in the revised manuscript (p5)

Zenodo url:

https://zenodo.org/records/11634657?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjE1NmEyZGZlLWU4ZTYtNDY4Mi1iNWFlLWE1ZGE5OGI3YjAxZCIsImRhdGEiOnt9LCJyYW5kb20iOiJiNTcxZGRiZTUzZTBlMzFmNTQwZDkzZDA2ZDYyMmE0ZiJ9.JtfV4UuHViK70BTaH_YVuhF5TK0cNS25mJXEBEP6vOJtJmT7CbT0k1KIL32-pluOaOvxoOGLLGPd-irv4EMHAw



The study primarily emphasizes interactions between inhibitory and excitatory neurons but does not thoroughly address possible roles of non-neuronal cells (e.g., astrocytes, oligodendrocytes, microglia) in circuit wiring, even though some of these cell types are included in the single-cell data. The authors should clarify whether such interactions were systematically explored or filtered out, and, if not, discuss this as a limitation and an avenue for future work.

We thank the reviewer for this important comment. As stated in the main text and detailed in the Methods section ("Assigning Broad Class Identity") p30, our study specifically focuses on post-mitotic glutamatergic (GlutN) and GABAergic (GABAN) neurons. Although non-neuronal were present in the single-cell RNA-seq data, we chose to exclude them from the current analysis. We acknowledge that non-neuronal cells likely play important roles in circuit wiring, and this is discussed as a limitation and future direction in the Discussion section of the manuscript (p27).

The manuscript would benefit from a clearer description of how thresholds for significance and specificity in LR interactions were determined, and how robust these results are to parameter tuning. Furthermore, differences in cell numbers across clusters could affect the quantification of LR interactions. I recommend performing a downsampling analysis to assess how varying cell numbers per cell type influence LR inference and the stability of predicted interactions.

We thank the reviewer for the valuable suggestion to perform a downsampling analysis. However, we would like to emphasize that the intercellular score calculation in scSeqComm already accounts for both the variability in ligand/receptor expression and the number of cells per cluster. Specifically, the method adjusts the reliability of the average ligand (or receptor) expression level by weighting it according to cluster size and expression variance. As a result, ligand or receptor signals observed in small or noisy clusters receive lower intercellular scores compared to those from larger, more homogeneous clusters. Following the reviewer's recommendation, we performed a downsampling analysis to directly test the robustness of LR inference to differences in cluster cell numbers. We took the P8 dataset for example, for this test, we downsampled the dataset to the same number of cells per cluster or the full P8 dataset (keeping the same number of cell type pairs between conditions >50 cells per cell type) and re-ran the scSeqComm analysis. The results were highly reproducible, supporting the robustness of our method to differences in cluster size (see Figure below for your convenience). Regarding thresholds used for defining significance and specificity of LR interactions. We want emphasize that neither the intercellular nor the intracellular scores generated by scSeqComm were modified in any way, and the core algorithm itself was not altered. The intercellular score was calculated exactly as described in the original publication and was returned for all possible LR pairs across all cell-type combinations. The additional connectivity score and developmental score were used afterwards as biologically informed filters to prioritize among the large number of potential LR interactions predicted by scSeqComm. In line with the recommendations of the scSeqComm authors, applying thresholds to intercellular scores is necessary to define significant interactions. Rather than using an arbitrary

cutoff, we integrated orthogonal information on connectivity potential (from Patch-seq–based experiments) and developmental timing. This prior knowledge was used to increase stringency and highlight interactions most likely to occur in vivo at the relevant developmental stages. We believe this refinement is both necessary and valid: by constraining the predictions to biologically plausible interactions, we reduce false positives and increase interpretability, while still retaining the full scSeqComm output for transparency. Importantly, we experimentally validated several of these filtered interactions (Figs. 5–6), supporting the utility of this approach. At the same time, we wish to stress that our analysis is not intended as an exhaustive catalogue of all valid LR interactions. Rather, it represents a framework and resource to help identify the most promising candidate interactions, which will ultimately require further functional validation.

However we agree that the criteria and thresholds used for defining significance and specificity of LR interactions could be described more clearly. We have now clarified these points more explicitly in the revised manuscript (p. 12).

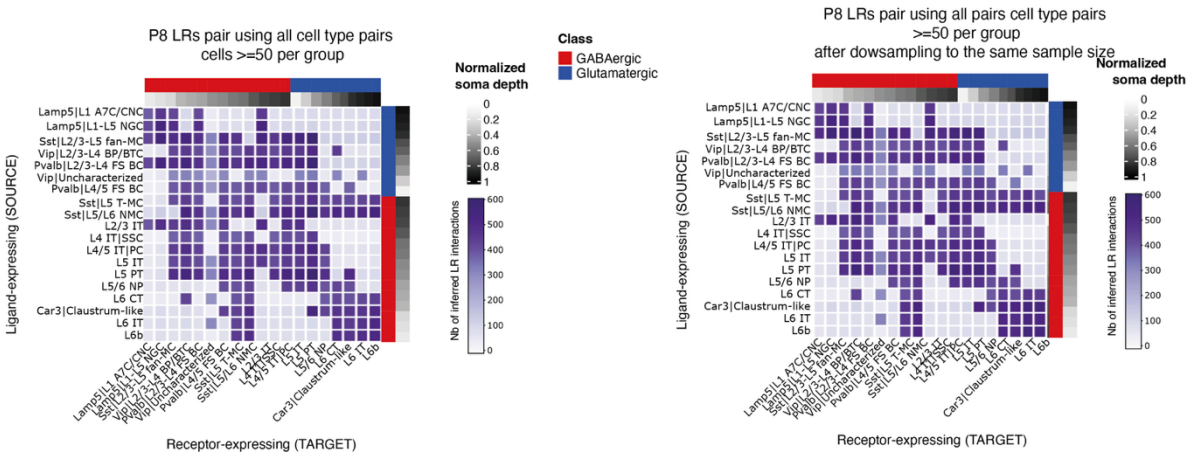


Figure for rebuttal. Effect of cluster size Ligand–receptor (LR) interaction analysis at P8. Heatmaps show predicted LR pairs identified using full dataset (left) or downsample to the same sample size per cluster (right). The same number of cell type pairs between conditions >50 cells per cell type were preserved between conditions. Rows represent ligand-expressing clusters and columns represent receptor-expressing clusters, with signal strength indicated by color intensity. GABAergic and glutamatergic classes are highlighted, and normalized soma depth is provided on the side bars. No major differences are observed between the full or downsample datasets.

Minor Comments

Since the manuscript posits that specific ligand-receptor combinations encode the molecular logic for establishing cell type-specific cortical wiring, directly testing whether these LR expression patterns can distinguish between subtypes would provide strong support for this hypothesis. I suggest the authors explore to what extent neuronal cell types can be distinguished or classified based on their ligand and receptor expression patterns. For example, using clustering or machine learning approaches on LR expression profiles, the authors could assess whether cell-type identities can be accurately recovered from LR combinations and from which developmental stage on.

We thank the reviewer for this important suggestion, which echoes a similar point raised by another reviewer. We fully agree that directly testing whether LR expression patterns can distinguish neuronal subtypes provides a strong evaluation of the "combinatorial code" hypothesis. A practical constraint is that the native scSeqComm output yields a single summary value per cell-type pair per LR (i.e., no within–cell-type replicates), so training a classifier on that matrix would be statistically ill-posed and prone to overfitting. To address the spirit of the suggestion without this limitation, we built an alternative ML framework that creates single-cell replicates for each SOURCE→TARGET cell-type pair. For each developmental stage, we expanded every SOURCE→TARGET pair into all single-cell combinations (source cell × target cell). Each example is a vector of ligand×receptor expression products (log1p, z-scaled), and the label is the SOURCE→TARGET cell-type pair. We evaluated five feature sets: (i) all significant LR pairs; (ii) shared-only (highly reused) pairs; (iii) rare-only pairs (≤10 cell-type pairs); (iv) unique-only pairs (used once); and (v) a size-matched random "fake" LR list (non-LR gene pairs). To

prevent information leakage, the train/test split was done by cells (no source or target cell seen during training appears in testing). We trained a regularized MLP (with/without first-layer weight pruning) and report performance per age (new Fig. S17). Our models trained on unique-only, shared-only, or fake lists perform at; or very near; chance across ages. In contrast, rare-only LR pairs consistently exceed chance and provide meaningful predictive signal. Most importantly, the best performance is achieved when combining LR sets (all significant = shared + rare + occasional unique), supporting our central claim: specificity resides in the combination of LR pairs, not in shared or unique subsets considered in isolation, with rare, context-restricted LRs providing much of the discriminative power. This result is robust across developmental stages. For context, this is an exceptionally hard task (>600 classes), so overall accuracy is necessarily modest; nevertheless, many individual cell-type pairs exhibit high ROC–AUC, demonstrating clear separability well above chance and indicating that the model captures genuine signal rather than noise. We have add the supplement figure (Fig. S17) and clarify in the Results (12-13), Discussion (p25) and Methods (p 40-41) .

One more point of context: this is an extremely hard prediction task (>600 classes), so overall accuracy is expected to be low simply because the chance baseline is tiny. Despite that, many individual cell-type pairs achieve high ROC–AUC values, showing strong separability well above chance even when the aggregate accuracy is modest. This pattern reinforces that the model captures real discriminative signal; consistent with a combinatorial LR code; rather than random fluctuations.

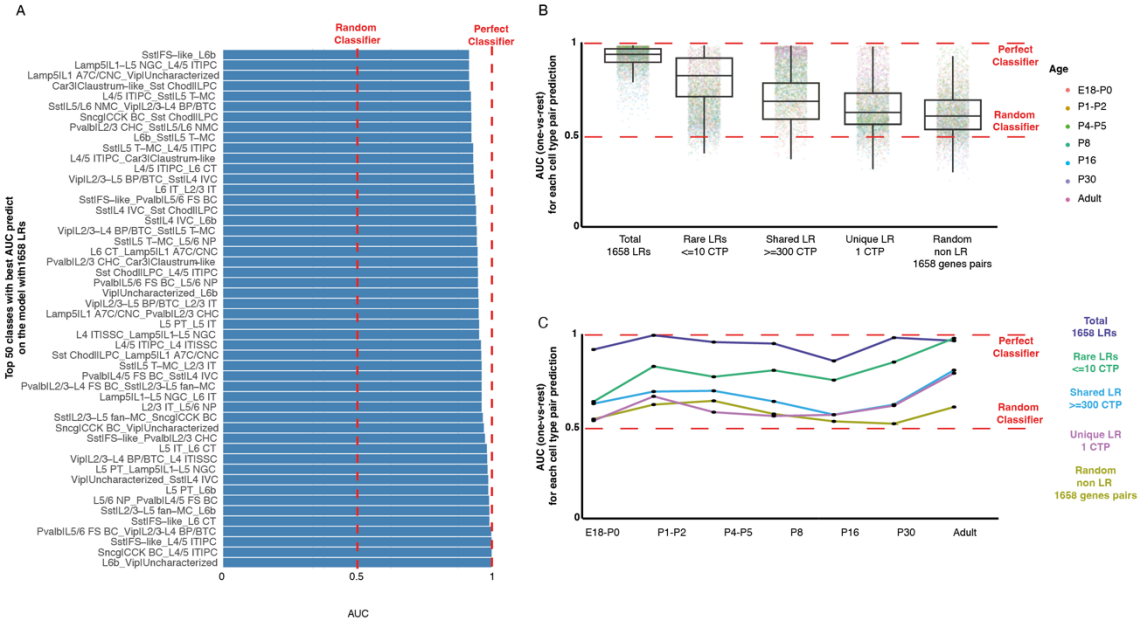


Figure S17. Predicting source–target cell-type pairs with machine learning reveals specificity encoded by combinatorial ligand–receptor patterns.

(A) Classification performance for the top 50 cell-type pairs (CTPs) predicted using all significant LR pairs. Bars show ROC–AUC values (one-vs-rest), with most CTPs reaching well above chance (red dashed line, random classifier) and approaching high separability (perfect classifier, AUC = 1).

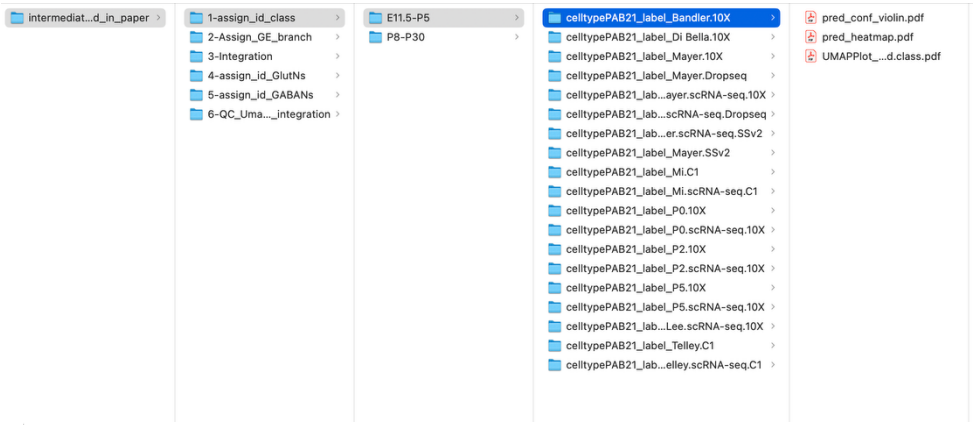
(B) Comparison of predictive performance across different LR feature sets. Models trained on all significant pairs perform best, while models restricted to shared or unique LR pairs perform at or near chance. Models trained only on rare LR pairs (≤10 CTPs) consistently exceed chance and contribute meaningful signal. Randomized non-LR gene–gene pairs do not provide predictive power. **(C)** Developmental trajectory of classifier performance across ages (E18–P0 to Adult). Rare-only LR pairs show reproducible predictive signal above chance across stages, while shared or unique pairs alone remain near chance. Models trained on the full LR set (shared + rare + occasional unique) achieve the highest performance, consistent with specificity arising from the joint usage of different LR subsets.

Line 103–105: Cell type classification is a crucial step that impacts all downstream analyses. The authors should provide more detailed information on the reliability of their label transfer, especially for immature neurons. For the ANN used for cell type classification, please report performance metrics such as ROC-AUC on the training dataset. Clarify how well the two different prediction methods agree with each other, and how many cells were excluded due to low prediction scores.

We thank the reviewer for this important comment. We would like to clarify that the artificial neural network (ANN) classifier was used exclusively to assign broad class identities (e.g., glutamatergic vs. GABAergic neurons) and DI Bella cell type for early developmental stages (E11.5 to P5), as shown in Figure S1. For later stages (P8 to P30), and for all downstream hierarchical annotation levels, we employed the nearest-centroid classifier from the scrattch.hiact package of the Allen Institute. Regarding performance metrics: we acknowledge that ROC-AUC scores for the ANN classifier are not currently included in the manuscript. We are now including example of them in the revised version Fig S1.E.
To assess prediction confidence for the nearest-centroid-based annotations, we used a bootstrapped label transfer procedure and report prediction probabilities. These results, including violin plots of prediction scores, are available now on Zenodo for all assignments made with scrattch.hicat, and we had clarified this in the revised manuscript (p5). Key snapshots of the folder organization is included bellow for your convenience.

Zenodo url:

https://zenodo.org/records/11634657?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjE1NmEyZGZlLWU4ZTYtNDY4Mi1iNWFlLWE1ZGE5OGI3YjAxZCIsImRhdGEiOnt9LCJyYW5kb20iOiJiNTcxZGRiZTUzZTBlMzFmNTQwZDkzZDA2ZDYyMmE0ZiJ9.JtfV4UuHViK70BTaH_YVuhF5TK0cNS25mJXEBEP6vOJtJmT7CbT0k1KIL32-pluOaOvxoOGLLGPd-irv4EMHAw



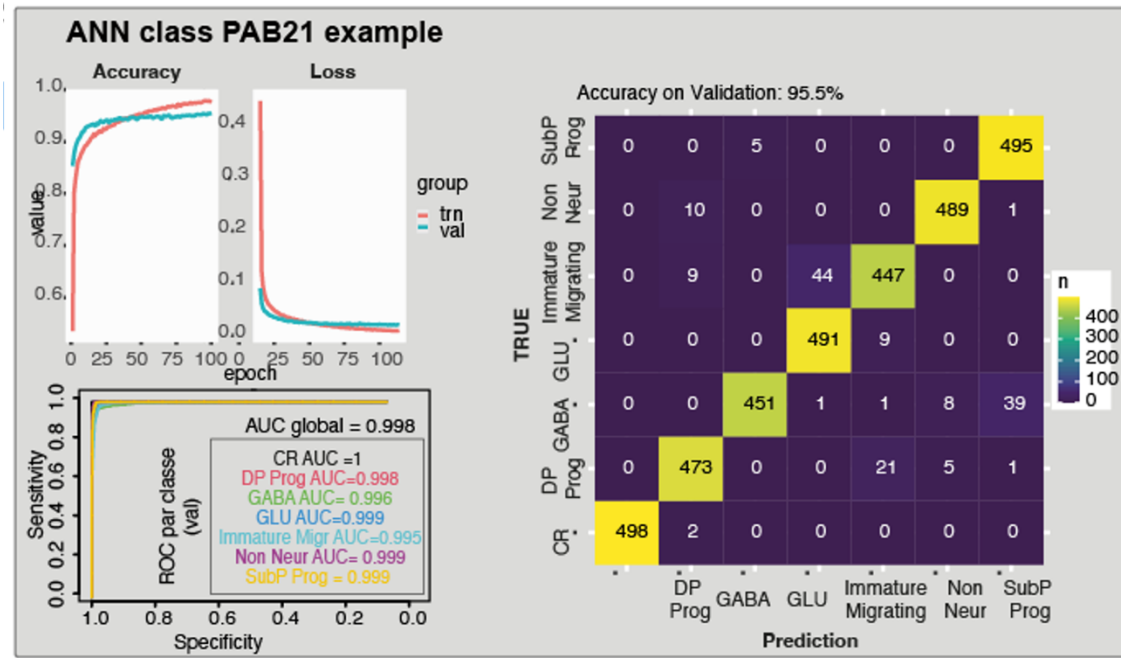


Figure S1. scRNA-seq pipeline and analysis workflow. e) Example performance of the ANN classifier. Training and validation curves indicate high accuracy (~95%) with steadily decreasing loss. ROC analysis confirms excellent discrimination across all classes (global AUC = 0.998). The confusion matrix on the right illustrates that most predictions fall on the diagonal, with only minor misclassifications, supporting the overall validation accuracy of 95.5%.

Lines 133–146: To characterize the identified gene expression waves, the authors could perform unsupervised GO enrichment analysis for each wave. It would be informative to analyze how many genes are shared across different cell types or classes within each wave, as this could reveal the extent to which a global maturation program is active versus cell type-specific expression dynamics.

We thank the reviewer for this constructive suggestion. As described in the Methods section and shown in Figures S5 and S6, GO enrichment analyses were already performed for each gene expression wave, both along the pseudo-maturation and pseudo-layer axes we show only key ontological process.

To address the second part of the comment, we agree that evaluating the extent of gene sharing across cell types is important. This suggestion echoes a point raised by Reviewer 1, which we addressed by testing the hypothesis using ganglionic eminence–derived interneurons as the source and pallial-origin glutamatergic cells as the target.

Specifically, we tested the hypothesis that GABAergic subtypes derived from the MGE, CGE, or POA exhibit either similar or distinct predicted ligand–receptor (LR) interactions with their glutamatergic counterparts in the pallium. To address this, we first regrouped all cell types by their ganglionic eminence of origin (MGE, CGE, or POA) and all glutamatergic cells by pallial VZ origin. We then quantified, at each developmental stage, the fraction of LR pairs that were either shared across at least two GEs or specific to one GE, by comparing VZ pallium–MGE, VZ pallium–CGE, and VZ pallium–POA interactions.

As summarized in Fig. S23A, our analysis revealed a dynamic temporal pattern of GE-specific versus shared LR modules. The fraction of shared LR interactions increased from embryonic stages to a peak at P8, followed by a progressive decline toward adulthood, reflecting a transition from early GE-specific programs to common communication modules, and then a return to specificity later in life. Fig. S23B further shows the distribution of individual LR pairs across development, with clear segregation of GE-specific versus shared modules. To determine whether this segregation reflected random partitioning or functionally distinct gene modules, we performed enrichment analyses of shared versus GE-specific genes at each age. This revealed distinct ontologies unique to GE-specific modules that were not observed in the shared sets, highlighting biological processes tied to the developmental origin of interneurons (Fig. S23C). Importantly, while early GE-specific functional modules (E18–P5) were enriched for programs linked to partner recognition and cortical integration, these converged into shared programs between P8 and P30, dominated by generic processes such as synapse organization and neuronal signaling. In adulthood, GE-specific modules re-emerged, suggesting renewed specialization and potentially subtype-specific synaptic communication.

Together, these results demonstrate that LR modules are both fate- and stage-dependent. Early GE-specific LR programs may facilitate partner recognition during migration and initial circuit entry, mid-developmental shared programs likely support global synapse formation, and adult specificity may reflect fine-tuned, lineage-dependent communication. This dynamic pattern of LR modules is a unique feature of GE-derived interneurons and underscores the developmental logic of their integration into cortical networks. These new findings are now described in Results, section p13, and further discussed in Discussion, section p25-26.

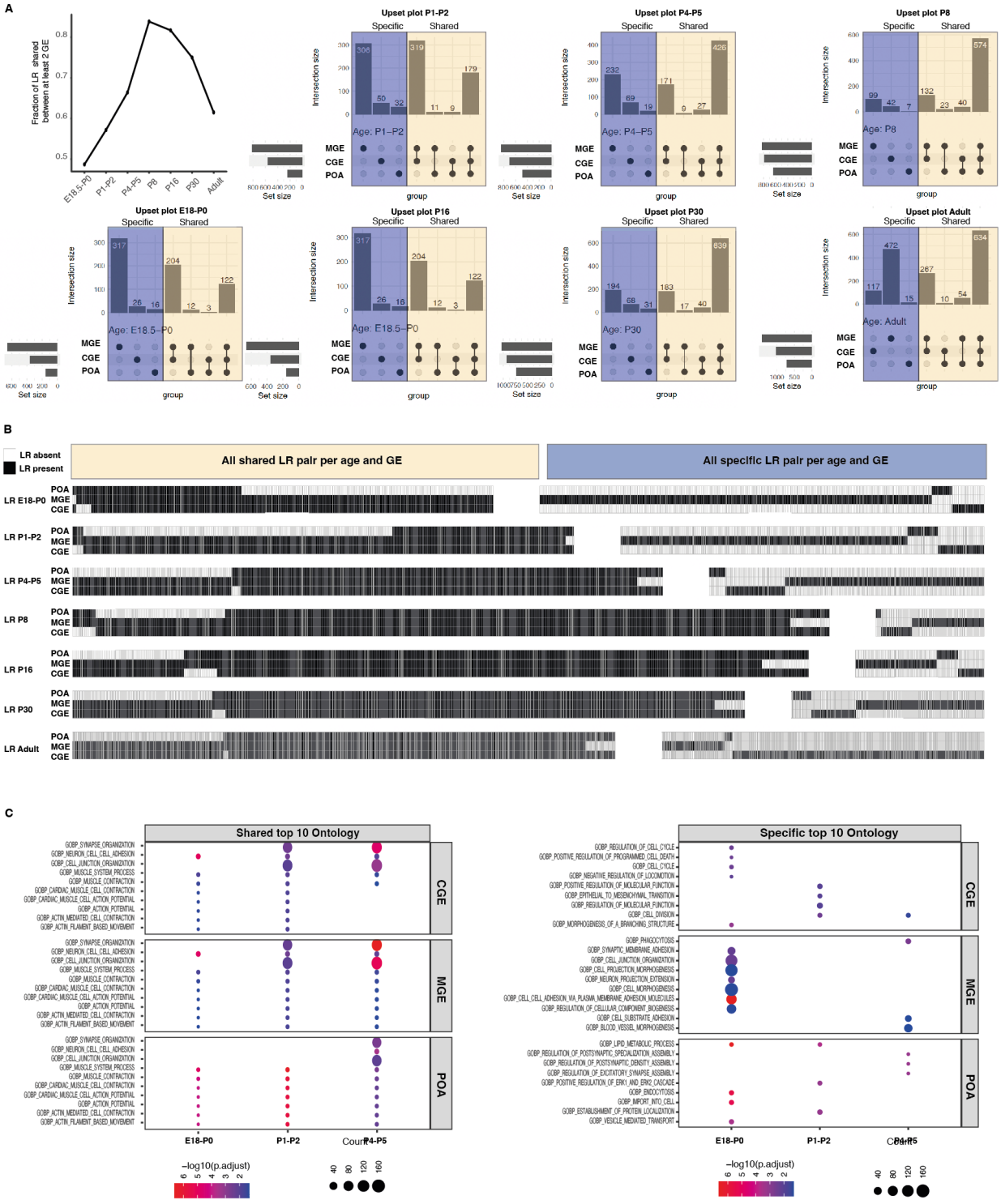


Figure S23. Ganglionic Eminence show specific and shared ligand–receptor modules across cortical development. (A) Temporal dynamics of sharing versus specificity of predicted ligand–receptor (LR) interactions, (left) line plot showing the fraction of LR pairs that are shared by at least 2 ganglionic eminence (GE) lineages (MGE, CGE, POA) at each developmental stage (E18.5–Adult) when Target cell type is Glutamatergic pallial cells. Right, for each stage, UpSet plots partition significant LR pairs into GE-specific (left panels) and shared (right panels). Bars on the left of each UpSet plot indicate the size of the LR sets for Source: MGE, CGE, and POA and Target (Glutamatergic pallial cells); intersection bars show the number of LR pairs common to the indicated GE combinations. **(B)** Presence/absence "barcode" maps of LR pairs across development. Left block: all shared LR pairs (columns) with rows for POA, MGE, and CGE at each stage; black = present, white = absent. Right block: all GE-specific LR pairs, displayed in the same format. This panel highlights the switch from early

GE-specific modules to a more shared program in mid-development, followed by a re-emergence of GE specificity in adulthood. **(C)** Gene Ontology enrichment for LR modules. Left, shared LR modules; right, GE-specific LR modules. For each GE and stage, the top 10 GO Biological Process terms are shown. Dot color indicates –log10(adjusted p value) and dot size indicates the number of LR genes contributing to the term. Enrichments emphasize synapse- and adhesion-related processes among shared modules and GE-biased functional programs among specific modules**.**

Lines 153–160: The definition and biological interpretation of the "spatial axis" in this section require further clarification. The approach appears similar to the maturation axis analysis described earlier, raising the question whether the pseudo-layer score captures spatial information, or if it largely reflects developmental progression. Please clarify the methodology for constructing the pseudo-layer score and provide evidence that it identifies gene sets distinct from the maturation axis. If there is substantial overlap or limited added value ; especially given that glutamatergic cell types are already defined by their laminar position ; it may be worth reconsidering the inclusion of this analysis.

The methodology for constructing the pseudo-layer score is described in detail in the Methods section ("Pseudo-layer score analysis"). This analysis is essential for defining the transcriptional landscapes presented later in the manuscript and cannot be removed. In the case of glutamatergic neurons, the pseudo-layer axis does partially correlate with the maturation axis at early developmental stages, which is expected given the inside-out pattern of cortical layer formation. However, this spatial axis remains relevant even at later stages, as it reflects persistent gene expression differences along the cortical depth, independent of temporal progression.
Importantly, we also observed a similar pseudo-layer gradient in GABAergic neurons. This suggests that spatially patterned gene expression is not exclusive to glutamatergic neurons but also contributes to the organization of GABAergic subtypes within the cortex.

The LR-Coord diagrams are confusing in their current form and lack a concise description.

We thank the reviewer for this helpful comment. We agree that the LR-Coord diagrams were confusing, and upon reflection we realized that they did not contribute substantially to the overall message of the manuscript. We have therefore decided to remove them and appreciate the reviewer for pointing this out.

Please provide a clearer explanation of Figure 3a

We apologize for the lack of clarity and agree that Figure 3A required a clearer description. This figure illustrates how we constructed our ligand–receptor database (LRDB). Specifically, we compiled 109 sources to generate a comprehensive reference containing 8,789 ligand–receptor pairs, involving 1,421 unique ligand genes and 1,233 unique receptor genes. However, this global LRDB is not brain-specific. To refine the database for our study and increase the resolution of the downstream ontological analyses, we defined a brain-focused LRDB by subsetting only the genes expressed in our single-cell dataset. This filtering resulted in 5,341 ligand–receptor pairs, comprising 1,022 ligand genes and 848 receptor genes. We have now revised the main text and legend to better explain this figure and to clarify the rationale behind creating both the comprehensive LRDB and the brain-specific version used in our analyses (p15).

Please clarify in the main text how morphological reconstructions were used to infer the spatial locations of cell types. Additionally, for interneurons of the same type or subtype that occupy different spatial positions (e.g., upper versus deep layers), please report and discuss which genes show differential expression depending on spatial location.

We thank the reviewer for this comment and are happy to clarify. Spatial localization of cell types in our study was determined using MERFISH and patch-seq datasets (see Methods, "Determining the distribution of cell types within a cortical column"). We now add explicitly stated this in the revised manuscript (p 7-8). Morphological reconstructions were not used to assign spatial positions. Instead, they were incorporated exclusively to build a 3D model of a cortical column and to compute the connectivity score, which estimates the probability of potential synapses between reconstructed cell types.

Reviewer #2 (Remarks on code availability):

The code and documentation are generally good. However, the Shiny app web tool has major usability and performance issues, which currently limit its value as a community resource.

Thank you for pointing this out. As mentioned above, we have worked on this critical point and substantially improved the usability and performance of the Shiny app web tool, making it a more reliable and valuable resource for the community.

Reviewer #3 (Remarks to the Author):

In this manuscript, Mathieu and colleagues used single-cell RNA sequencing (scRNA-seq) to characterise gene expression dynamics among developing neurons during the first month of postnatal life in the mouse somatosensory cortex. Using this comprehensive atlas, which they integrated with previously published datasets from the embryonic and adult cortex, they identified developmental trajectories for the main types of glutamatergic and GABAergic neurons. They then assembled a comprehensive map of putative receptor-ligand interactions and explored possible expression patterns underlying connectivity between glutamatergic and GABAergic neurons.

This is a very useful resource for identifying putative molecular interactions between distinct neuronal populations in the mouse somatosensory cortex. As such, this study will be of great utility to the community. I only have a few suggestions that help clarify some aspects of the manuscript.

We thank the reviewer for his/her detailed evaluation and positive appreciation of our work.

Main issues

One important caveat that limits the impact of the findings reported by the authors is the assumption that specific receptor-ligand interactions are associated with known functions for a particular pair of molecules. For example, the authors conclude that "…processes like cell death or differentiation may rely more on intrinsic mechanisms" (lines 224 and 225). A similar assertion is made in the paragraph starting on line 282. This is counterintuitive because, for example, interneuron cell death is known to depend on interactions between excitatory and inhibitory cells. I suggest the authors take a more agnostic approach when linking specific ligand-receptor pairs to function. There is a danger in overplaying GO terminology, as it only relies on what we currently know. I don't have a problem with the author's focus on synaptic connectivity, but they shouldn't rule out that cell-cell interactions are important for other developmental processes during this period.

We thank the reviewer for this insightful comment, and we fully agree that our initial wording may have conveyed an overly strong interpretation of the GO analyses. As suggested, we have revised the text to adopt a more cautious and agnostic tone when linking specific ligand–receptor pairs to developmental processes such as cell death or differentiation. Our intention was not to exclude the possibility of intercellular signaling in these processes, but rather to highlight that our GO-based inference is limited by the current scope of available annotations. We now emphasize that GO terminology reflects existing knowledge and may underrepresent certain biological roles, including those supported by experimental studies (e.g., interneuron cell death depending on excitatory–inhibitory interactions). We have modified the relevant passages (p 11, 15, 27) to reflect this more measured interpretation.

Using an adult connectivity matrix restricts the potential universe of ligand-receptor interactions in a practical manner. However, it assumes that the connectivity we see in the adult cortex is the same as during early postnatal development. Numerous examples indicate that this is not always the case, as transient connectivity is abundant during these stages. The manuscript should acknowledge this limitation, and relevant references should be added in this context.

We thank the reviewer for raising this important point. We fully agree that using an adult-derived connectivity matrix constrains the scope of inferred ligand–receptor interactions to those supported by mature

circuit architecture and therefore does not fully capture the transient and dynamic connectivity patterns that are known to occur during early postnatal development. This limitation is indeed acknowledged in our approach. Our choice to rely on adult connectivity was motivated by the need to increase biological plausibility and reduce inflated false positives in the inferred interactions. To partly address this limitation, we incorporated a *developmental score* that modulates the influence of the adult-derived connectivity matrix based on known developmental trajectories and cell-type–specific maturation profiles (see Table S4 and *Methods*). We fully agree that having access to developmental-stage-specific connectivity data would provide a more accurate framework; however, such spatially resolved datasets are currently not available with the sequencing methods used in this study. We have now revised the manuscript to more explicitly state this limitation and to cite relevant literature documenting transient and stage-specific connectivity during early cortical development (p. 27).

The authors suggest that cell-cell interactions may play a more prominent role in regulating the connectivity of excitatory cells with SST and PV interneurons than with other cortical interneurons, such as VIP cells (lines 270-273). Are there alternative explanations? Could the number of cells or the stages sampled (which may not directly overlap with the wiring of VIP interneurons, for instance) limit the exploration of those interactions? In any case, conclusions like this (and other examples in the Results section), which are speculative, should probably be moved to the Discussion.

We agree with the reviewer's comment that some of the conclusions, such as those about increased interactions with SST and PV interneurons (lines 270–273), are speculative and would be more appropriate in the Discussion. However, we can justify this observation beyond sampling bias. The number of predicted LR interactions may reflect the morphological and subtype diversity within each subclass, with SST interneurons being the most diverse (e.g., fan cells, T-shaped Martinotti cells, NMCs, IVCs), followed by PV (FS basket cells and chandelier cells). In contrast, VIP cells are more homogenous (mostly bipolar/BTC), and Lamp5 consists of only two main types.
Additionally, SST and PV interneurons are more broadly distributed across cortical depth, whereas VIP is largely restricted to L2/3 and Lamp5 to L1. This spatial distribution likely contributes to increased opportunities for interactions with glutamatergic neurons, making our findings biologically plausible even if somewhat speculative. Still, I agree we should tone down the claims in the Results section (p15) and move this type of interpretation to the Discussion (p26). In addition, we have also added a statement at the end of the Discussion highlighting the limitations of our approach, including sampling- and stage-related caveats.

The authors should perform functional experiments (similar to those performed for CDH13) to confirm the role of NEO1 in the connectivity between Martinotti cells and excitatory neurons. In contrast to the other example (ErbB4-Nrg3), no functional evidence supports a role for NEO1-Cbln4 interactions in this context.

We thank the reviewer for highlighting the importance of functionally validating the role of NEO1–Cbln4 interactions in Martinotti–glutamatergic connectivity. We agree that such experiments would be of great interest. As suggested, we have already initiated collaborations to test this interaction in vivo using conditional Neo1 mice combined with AAV-based manipulations of SST neurons (which require an MTA with the US and a complex breeding strategy). In parallel, we have performed in utero electroporation at E15.5 with either shRNA Neo1 or shRNA Ctl, together with a GFP plasmid. At P0, we injected AAV-BisSTe4_dTomato (Addgene) intraventricularly to label Martinotti cells in red. The goal is to perform co-staining for Gad65 and Gephyrin in order to visualize, in layer 1, Martinotti axon terminals (red) expressing Gad65 in contact with Glutamatergic neuron dendrites (green) containing Gephyrin, thus identifying Martinotti–Glutamatergic synapses. These experiments, however, require substantial additional time to complete and, in our view, fall beyond the scope of the present study, which is primarily focused on establishing a broadly applicable LRDB framework and demonstrating its predictive value through multiple already validated examples (e.g., Nrg3–ErbB4 and Cdh13/Pcdh8 cases, Figs. 5–6).

We therefore consider the NEO1–Cbln4 interaction an exciting direction for future work, and we thank the reviewer for this suggestion. Importantly, our current atlas already highlights this candidate with strong computational and transcriptional support, underscoring the utility of the resource in generating new, testable hypotheses.

Minor

Line 257. I assume “CR-CR interactions” means interactions among Cajal-Retzius cells. Please revise the use of abbreviations throughout.

We thank the reviewer for pointing this out. Yes, “CR” refers to Cajal-Retzius cells. We apologize for not defining this abbreviation earlier in the manuscript and will ensure that all abbreviations, including “CR,” are clearly defined upon first use in the revised version (p8).

Line 270. Please consider Modol et al., Neuron 2024, in this context.

We thank the reviewer for the very interesting suggestion. We had incorporate it p15.

REVIEWERS' COMMENTS
Reviewer #1 (Remarks to the Author):

I really appreciate the authors' incorporation of both major and minor suggestions, and overall agree in the few cases where the authors rebut some minor suggestions.

We thank the reviewer for their detailed evaluation and positive appreciation of our work. His/her constructive feedback has helped us further clarify key points and improve the overall quality and rigor of the manuscript.

scRNA-seq vs snRNA-seq
The comparison of scRNA-seq and snRNA-seq is very useful and the authors' motivation to include both and the manner of handling both is reasonable. Moreover, I appreciate the concession on page 12 that snRNA-seq failing to capture short transcripts can bias results.

Thank you very much we are glad that the revisions have addressed your suggestion.

Region-specific cell-types
The clarification and caveat on extrapolating findings to the rest of the cortex is well-stated. I appreciate the new branch of analysis on interneuron regional origin which the authors interpreted nicely.

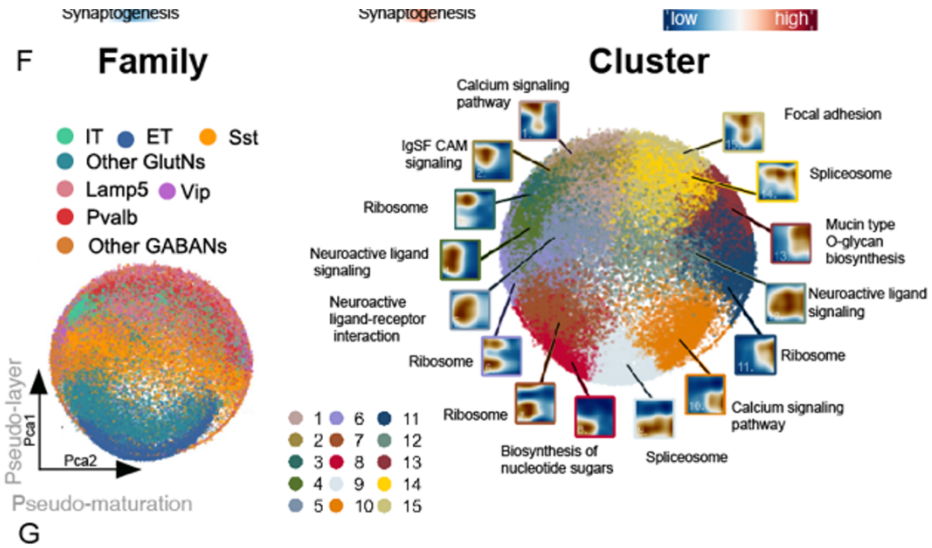
Thank you very much we are glad that the revisions have addressed your suggestion.

Pseudo-spatiotemporal analysis
The finer detail on pseudolayer analysis is good, especially the careful descriptions of how MERFISH and patch-seq were used. The term "wave" instead of "gradient" is fine, but please make sure it's used consistently. Figure 2E also looks very clean and interpretable now.

We thank the reviewer for the positive feedback regarding the pseudolayer analysis and the clarification of MERFISH and Patch-seq methodologies. Following the reviewer's suggestion, we have carefully revised the manuscript to use the term "gradient" consistently when referring to the pseudo-spatiotemporal pattern, replacing the previous use of "wave."

A deeper gene enrichment analysis would be a nice complement to the spatiotemporal clusters found in panel F, beyond just mentioning the processes of lamination and synaptogenesis in the text. It would be nice to label the spatiotemporal pattern clusters in panel F by their enriched functions.

We thank the reviewer for this suggestion. We have now performed a deeper KEGG enrichment analysis and labeled each spatiotemporal cluster in panel F with its corresponding top enriched functional category (Fig. 2G; Table S8). This provides clearer biological interpretation of the developmental programs represented by each cluster.



A new concern is about the claim of immature signatures of upper vs lower layer neurons (page 8, the last paragraph of the "Spatial transcriptional gradients" section). If you claim that upper layer neurons show immaturity markers at E18.5-P8 and deep layer neurons do not, corresponding to their relative birthdates, you need to demonstrate that past P8, these maturity differences disappear. It would also help to demonstrate that deep layer neurons show the same/similar immaturity markers at earlier developmental stages if feasible.

We thank the reviewer for this important comment. We agree that our previous wording overstated the interpretation regarding relative immaturity of upper- versus deep-layer glutamatergic neurons, as we did not provide analyses at later postnatal stages nor earlier embryonic stages to substantiate this developmental trajectory. To avoid overinterpretation, we have revised the text to remain descriptive and remove any claims about maturation differences beyond the observed time window. The revised paragraph now simply reports the transcriptional patterns present at E18.5–P5 without inferring later convergence or earlier signatures. We believe this correction strengthens the accuracy and clarity of the manuscript.

New LR scoring metrics
Descriptions on page 12 of developmental score and connectivity score, and how they are used to filter (rather than manipulate) the results of scSeqComm, are much clearer. However, like Reviewer #2, I would appreciate a very explicit statement of how thresholds on the intercellular score were made. I understand that the developmental and connectivity scores helped to define this threshold, but the specific method for converting from a set of continuous scores to a binary LR pair (significant/non-significant) should be explicitly stated. I do appreciate the authors checking that using the intercellular score rather than a binary LR pair provides very similar results.

We thank the reviewer for pointing out the need for a clearer explanation of how continuous intercellular scores are converted into binary ligand–receptor decisions. We apologize for the lack of clarity in the original version. In the revised manuscript, we have now added a dedicated section on page 29 of the Methods that explicitly describes how the cell-type score, together with the intercellular score, is used to define a pair- and cell-type–specific threshold, and how this threshold is applied to classify each LR pair as significant or non-significant. This new section makes it clear that continuous scSeqComm scores are not modified but rather filtered using these thresholds.

Limitations of transcriptomics
The limitations of transcriptomics for LR interaction studies on page 26 are a nice addition and provide sufficient skepticism to the claim that paired ligands and receptors will have correlated expressions over time.

Thank you very much we are glad that the revisions have addressed your suggestion.

LR specificity to cell type pairs
I really like the direction you took my suggestion of classifying cell type pairs based on ligand-receptor, and how you stratified by LR class (shared, unique, etc.) and developmental stage in Fig. S17. I can see how this is a very difficult classification problem, and the interpretation you derive provides very strong support to your claims about LR class contribution to cell type.

Thank you very much we are glad that the revisions have addressed your suggestion.

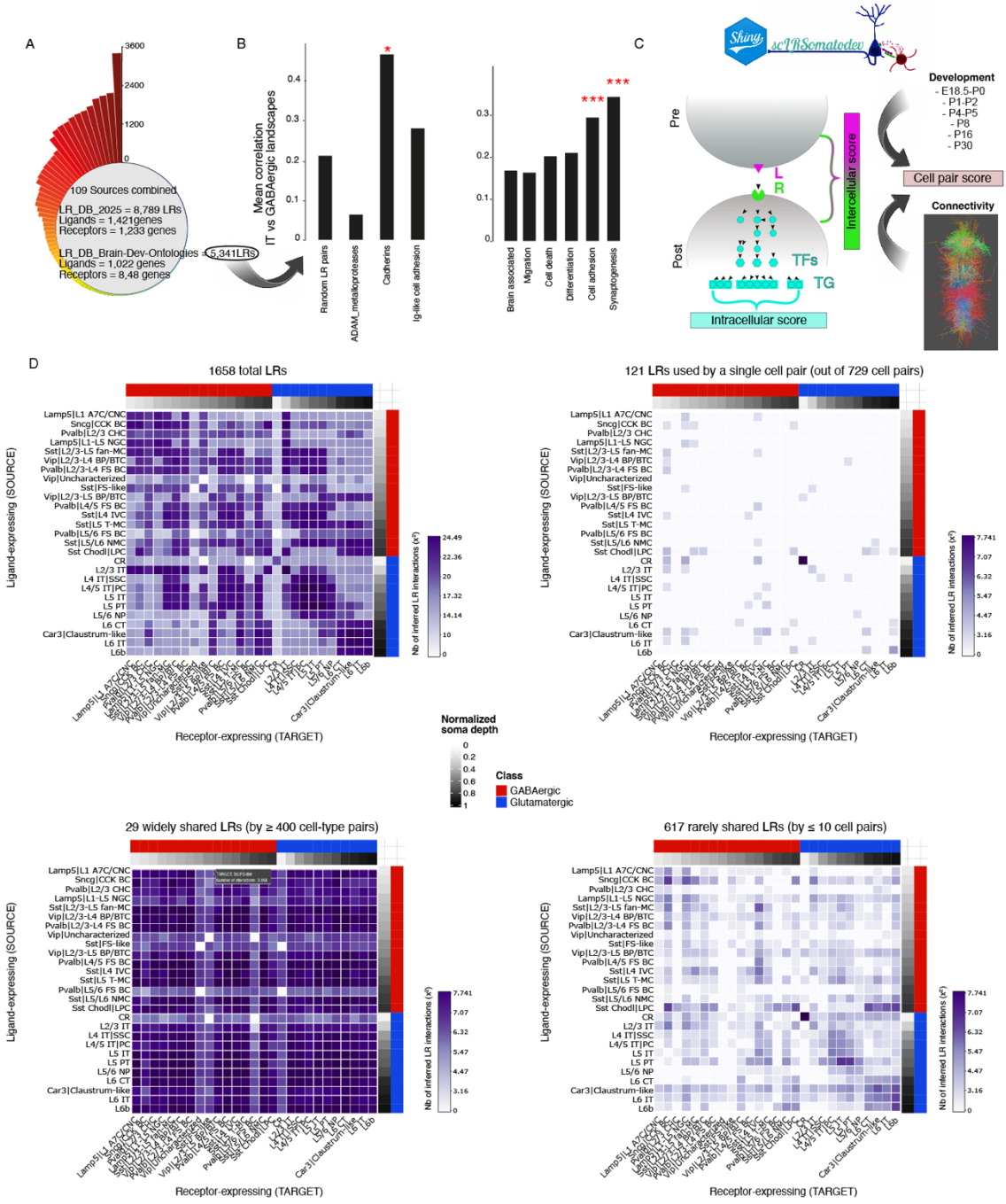
Source-target plots in Figures 3 and 4

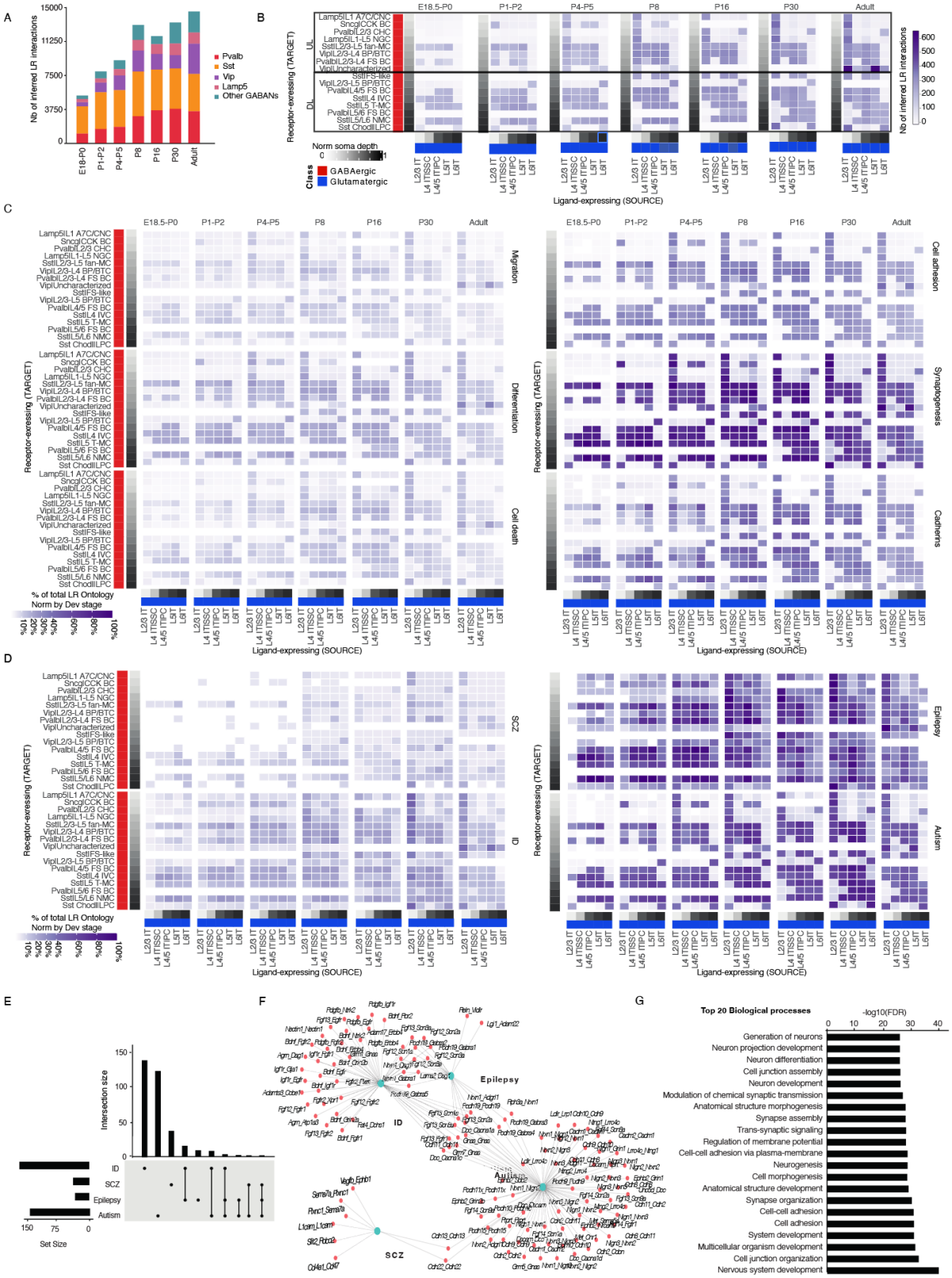
The colormaps in Figures 3 and 4 are very colorblind accessible now. However, the light to dark scale makes small signals hard to see in some cases. In those cases (top right panel of Fig 3D, left panel of Fig 4C), a break can be added to the colorbar to solve this problem.

We thank the reviewer for this helpful suggestion. For Figures 3 and 4A–B, which are generated using interactive heatmaps (plotly-based), the plotting engine unfortunately does not support adding breaks in the color scale. To address the reviewer's point and improve visibility of low-intensity values, we have modified the colormap range by applying a square-root transformation to the LR counts. This transformation enhances contrast for lower values while preserving the overall distribution, and it provides a visual effect comparable to introducing a colorbar break.

For Figure 4C, which is generated with ggplot2, breaks in the color scale can be implemented directly. Following the reviewer's advice, we have added explicit colorbar breaks to improve the visualization of low-percentage values.

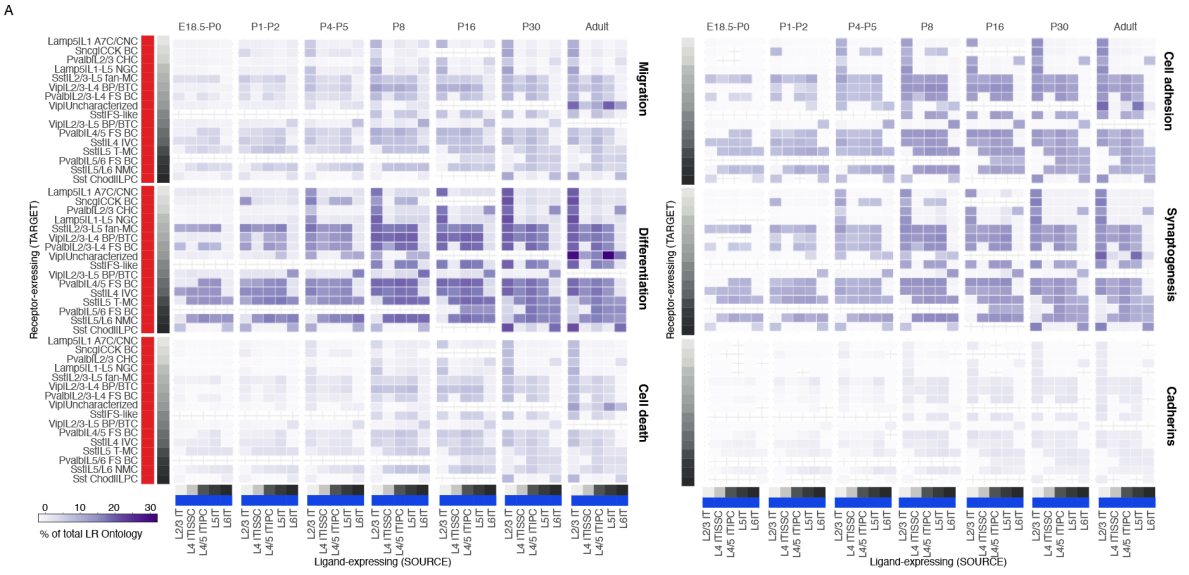
We are grateful for this suggestion, which has improved the clarity and readability of these figures (see below).

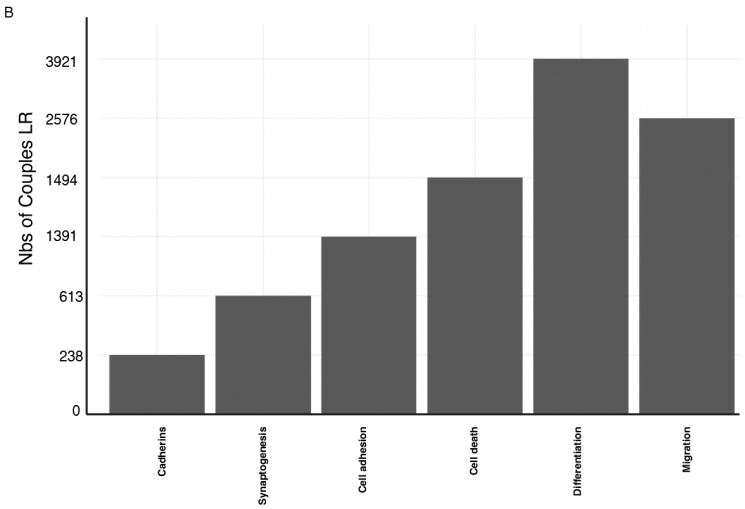



A
Nb of interrecl LR interactions
Pvalb
Sst
Vip
Lamp5
Other GABANs
E18-P0
P1-P2
P4-P5
P8
P16
P30
Adult
B
Receptor-exressing (TARGET)
Ligand-expressing (SOURCE)
Norm soma depth
Class
GABAergic
Glutamatergic
C
Migration
Differentiation
Cell death
Cell adhesion
Synaptogenesis
Cadherins
% of total LR Ontology
Norm by Dev stage
D
SCZ
ID
Epilepsy
Autism
E
Intersection size
Set Size
F
Epilepsy
ID
SCZ
G
Top 20 Biological processes
-log10(FDR)
Generation of neurons
Neuron projection development
Neuron differentiation
Cell junction assembly
Neuron development
Modulation of chemical synaptic transmission
Anatomical structure morphogenesis
Synapse assembly
Trans-synaptic signaling
Regulation of membrane potential
Cell-cell adhesion via plasma-membrane
Neurogenesis
Cell morphogenesis
Anatomical structure development
Synapse organization
Cell-cell adhesion
Cell adhesion
System development
Multicellular organism development
Cell junction organization
Nervous system development

On that note, I recommend that the colorbar in Figure 4 refer to the number of LR interactions as in Figure 3, unless there is a good reason otherwise. An absolute number of LR pairs is clearer than "proportion of total ontology." If you prefer using proportion of total ontology, it would be best to have the total number of LRs next to the colorbar.

We thank the reviewer for this suggestion. However, using the absolute number of significant LR pairs in the color scale would introduce a strong bias due to the unequal number of possible LR pairs across ontologies. As shown in Pannel B, each ontology contains a different total number of ligand–receptor combinations. Plotting only the raw counts Pannel A would therefore cause the heatmap to be dominated by ontologies that simply contain more LR pairs in total, rather than reflecting meaningful biological differences as you can see.
For this reason, we normalize the number of significant LR interactions by the total number of LR combinations available within each ontology category. This proportional normalization removes the artefact caused by unequal ontology sizes and allows comparison across categories on an equal footing. In other words, it highlights the relative enrichment of significant LR interactions within each ontology rather than the absolute count, which would otherwise be driven by ontology size rather than biology. For clarity, we now explicitly state this rationale in the methods (page 31).





That being said, in Figures 4C and 4D, I dislike how the color scale is normalized to the total LR ontology across all developmental stages rather than for each developmental stage. Using the same color scale for all developmental stages in Figures 4C and 4D exaggerates how the number of LR pairs increases with maturation, masking other trends. What's more interesting than an increase in the number of LR pairs over time is how the configurations of LR interactions across cell types change over time. This can be accomplished simply by setting the max color value for each column (developmental stage) independently. That would remove the trend of an increasing number of LR interactions over time and leave us with the residual timepoint-specific patterns for emphasis.

We thank the reviewer for this suggestion. Following this recommendation, we now normalize the values within each developmental stage independently. Specifically, for every developmental stage, all ligand–receptor (LR) interaction values are rescaled relative to the maximum value of that stage. The revised Figures 4C and 4D now highlight the temporal reorganization of LR interactions exactly.

Another concern: can the increase in the number of significant LR pairs over developmental time (Figure 4A) be explained by some confounding feature of the datasets? For instance, could more cells in older datasets explain why there are more significant LR pairs? It would be easy to alleviate this concern by repeating your analysis multiple times on subsetted data to balance the number of cells at each developmental stage and see whether the trend in number of significant LR pairs holds.

We thank the reviewer for this insightful suggestion. We agree that differences in cell numbers across developmental stages could, in principle, bias the number of inferred interactions. Importantly, the intercellular score in scSeqComm already accounts for both cluster size and expression variability: the reliability of the average ligand or receptor expression is weighted by the number of cells and its variance, such that small or heterogeneous clusters naturally receive lower intercellular scores. This built-in correction in scSeqComm algorithm reduces the risk that larger clusters artificially inflate the number of significant LR interactions.
Nonetheless, following the reviewer's recommendation, we performed a dedicated downsampling analysis to directly assess the robustness of scSeqComm algorithm results to differences in cluster size. Using the P8 dataset as an example, we downsampled cell numbers so that each cluster contained the same number of cells as in the smallest retained dataset (while maintaining the same set of cell-type pairs with >50 cells per type),and re-ran the full scSeqComm analysis. The results from the downsampled and full datasets were highly concordant, showing no major differences in the number, distribution, or identity of predicted LR interactions (see figure below). This confirms that the observed developmental increase in LR interactions is most likely not driven by biological differences in cell numbers. Overall, both the properties of the scSeqComm scoring framework and our downsampling analysis support that the developmental increase in LR interactions cannot be explained by differences in cell numbers across datasets.

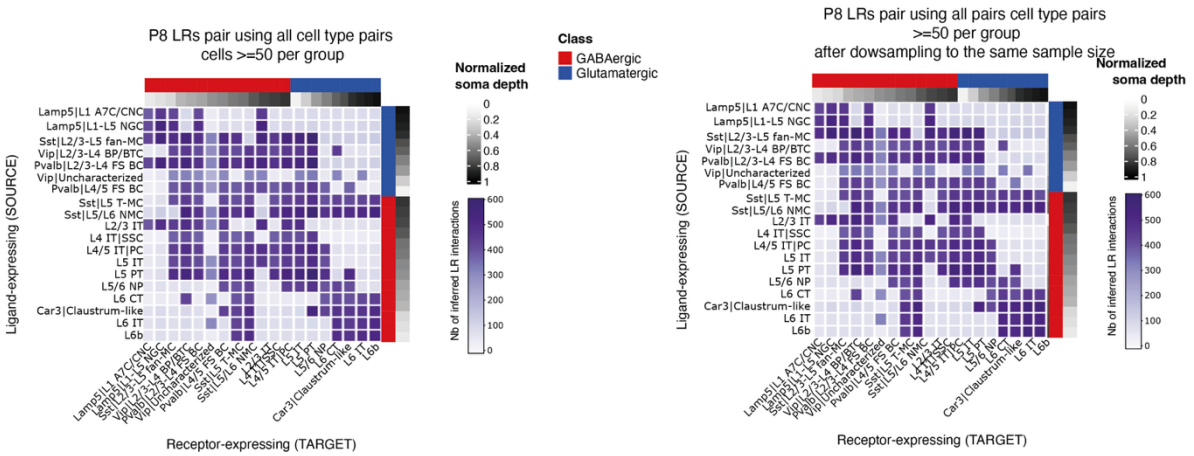


Figure for rebuttal. Effect of cluster size Ligand–receptor (LR) interaction analysis at P8. Heatmaps show predicted LR pairs identified using full dataset (left) or downsample to the same sample size per cluster (right). The same number of cell type pairs between conditions >50 cells per cell type were preserved between conditions. Rows represent ligand-expressing clusters and columns represent receptor-expressing clusters, with signal strength indicated by color intensity. GABAergic and glutamatergic classes are highlighted, and normalized soma depth is provided on the side bars. No major differences are observed between the full or downsample datasets.

Lastly, I'm not quite convinced about the claim on page 15 about how DL GABANs show higher LR engagement than UL GABANs early on, as Figure 4B makes this trend seem very slight. Statistically testing the relationship between depth and rate of LR engagement would provide much better support to this claim.

We thank the reviewer for pointing this out. We agree that the difference between UL and DL GABANs in early LR engagement is subtle. In the revised manuscript, we have toned down our interpretation of this observation. Specifically, we now describe this pattern as a possible trend rather than a strong depth-dependent difference (page 10). We also clarified that additional statistical analyses would be required to rigorously establish whether this represents a meaningful depth-related effect.

Reviewer #2 (Remarks to the Author):

The authors have addressed all of my previous concerns in a satisfactory manner. The revisions improve the clarity of the manuscript and strengthen the conclusions. The additional analyses are appropriate and support the main claims. I do not have further comments.

I recommend the manuscript for publication in its current form.

We thank the reviewer for their detailed evaluation and positive appreciation of our work. His/her constructive feedback has helped us further clarify key points and improve the overall quality and rigor of the manuscript.

Reviewer #2 (Remarks on code availability):

I examined the provided code. The code is clearly organized. The code should be usable by others in the community with standard computational experience. I have no concerns regarding reproducibility.

Reviewer #3 (Remarks to the Author):

This is a very nice study and an important resource. The authors have done a good job addressing my comments, and I have no further suggestions.

We thank the reviewer for their detailed evaluation and positive appreciation of our work. His/her constructive feedback has helped us further clarify key points and improve the overall quality and rigor of the manuscript.