THE EMBO JOURNAL

Temporal and sequential transcriptional dynamics define lineage shifts in corticogenesis

Tanzila Mukhtar, Jeremie Breda, Manal Adam, Marcelo Boareto, Pascal Grobecker, Zahra Karimaddini, Alice Grison, Katja Eschbach, Ramakrishnan Chandrasekhar, Swen Vermeul, Michal Okoniewski, Mikhail Pachkov, Corey Harwell, Suzana Atanasoski, Christian Beisel, Dagmar Iber, Erik van Nimwegen, and Verdon Taylor
DOI: 10.15252/embj.2022111132

Corresponding authors: Tanzila Mukhtar (tmturkee@gmail.com) , Verdon Taylor (verdon.taylor@unibas.ch), Erik van Nimwegen (erik.vannimwegen@unibas.ch), Dagmar Iber (dagmar.iber@bsse.ethz.ch)

Review Timeline:

Submission Date:	10th Mar 22
Editorial Decision:	28th Apr 22
Revision Received:	27th Jul 22
Editorial Decision:	30th Aug 22
Revision Received:	9th Sep 22
Accepted:	26th Sep 22

Editor: Kelly Anderson

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. Depending on transfer agreements, referee reports obtained elsewhere may or may not be included in this compilation. Referee reports are anonymous unless the Referee chooses to sign their reports.)

Dear Dr. Mukhtar,

Thank you for submitting your manuscript for consideration by the EMBO Journal. It has now been seen by three referees whose comments are shown below.

Given the referees' positive recommendations, I would like to invite you to submit a revised version of the manuscript, addressing the comments of all three reviewers. I should add that it is EMBO Journal policy to allow only a single round of revision, and acceptance of your manuscript will therefore depend on the completeness of your responses in this revised version. I have attached a guide for revisions to this email. It would be good to discuss your plan for revisions and I am available to do so in the coming weeks by email or zoom.

When preparing your letter of response to the referees' comments, please bear in mind that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process, please visit our website: https://www.embopress.org/page/journal/14602075/authorguide#transparentprocess

We generally allow three months as standard revision time. As a matter of policy, competing manuscripts published during this period will not negatively impact our assessment of the conceptual advance presented by your study. However, we request that you contact us as soon as possible upon publication of any related work, to discuss how to proceed. Should you foresee a problem in meeting this three-month deadline, please let us know in advance and we may be able to grant an extension.

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

Yours sincerely,

Kelly M Anderson, PhD
Editor
The EMBO Journal
k.anderson@embojournal.org

Further information is available in our Guide For Authors: https://www.embopress.org/page/journal/14602075/authorguide

We realize that it is difficult to revise to a specific deadline. In the interest of protecting the conceptual advance provided by the work, we recommend a revision within 3 months (27th Jul 2022). Please discuss the revision progress ahead of this time with the editor if you require more time to complete the revisions. Use the link below to submit your revision:

Link Not Available

Referee #1:

This is a resource paper analyzing the transcriptomic landscape of the developing mouse cerebral cortex along the entire neurogenic and early gliogenic periods, distinguishing separately between Radial Glia Cells, Intermediate Progenitor Cells and Newborn Neurons. The authors use two different transgenic reporter mouse lines to isolate these populations based on their expression of GFP under the specific promoter of Hes5 and Tbr2, where IPCs and NBNs are distinguished by high and low Tbr2::GFP levels. The study is very carefully performed and provides an in-depth characterization of the transcriptional landscape of these cell types at the population levels, and their heterogeneity at the single cell level. The most relevant distinction of this study over others is the identification of the transcriptome of specific cell populations without a priori knowledge or establishment of marker genes (although it is based on Hes5 and Tbr2 expression). This is important to ascribe cell identities based on transcriptome without bias on a priori knowledge of marker proteins, which commonly leads to misidentifications of cell types. Unfortunately, while this value of the study is unquestionable, the novelty that it brings over previous similar analyses (i.e. Telley et al 2019, Science) is insufficiently highlighted in the manuscript, and it may be not obvious to the reader.
One of the main caveats of this study is the extent of its utility as a resource for other scientists in the field, because all the results are based on the use of the two transgenic reporter mouse lines, which other studies will likely not use. The identification of cell groups (clusters) in single cell transcriptomic studies is strongly influenced by the collection of input cells, and thus the conclusions reached here may not be easily translatable to studies analyzing all cell classes in the embryonic cortex combined. In this sense, it would be a very significant improvement if the authors compared their results with data from non-reporter mice (i.e. publicly available data from other studies) to demonstrate that the cell populations identified here, as well as their heterogeneity and temporal evolution, are distinguishable without the need of using the Hes5 and Tbr2 reporters.
Other major points:

Fig EV1B should show also DAPI for reference of the cortical layers. Labeling with Tbr2 seems strange, as cells expressing endogenous Tbr2 protein are usually spread apically from SVZ into VZ (see Englund et al 2005, Fig 1), but Fig EV1B indicates otherwise. Hes5::GFP clearly labels cells only in VZ, but cells in the basal part of VZ are much less bright (or negative). Does this marker label a selection of RGCs? Validation in Fig EV1D should include quantifications, as well as co-expression of Pax6 and Tbr2 protein in cycling cells.

In page 6, the authors indicate that "Principal component analysis (PCA) capturing 60% of the total variance (PC1 and PC2) separated the samples based on cell type (NSC, BP and NBN) and developmental stage.". However, this separation is minimal for BPs and NBNs. Also BPs overlap completely with NSC mid-neurogenesis. This is very relevant because variance explained by PC1 is quite low, and PC1+2 is only 60% (very low).

The authors indicate that "DEG analyses revealed that the majority of the highly expressed genes in NSCs are downregulated by BPs and reduced further by NBNs. (page 7)". This is a fascinating finding not previously stressed in other studies. Does this mean that a large number of NSC genes must be turned off so that already expressed "neuronal" genes can exert their function in BPs and nascent neurons? These trends in gene expression, shown in Fig 1E, would be better identified if the heatmap was ordered in a hierarchical manner. Given this conclusion, when (and how) do the many different types of cortical neurons become distinct? The authors should at least speculate in Discussion on this question.

The authors report that expression of Shh pathway genes mark specifically the early expansion phase and their expression is low upon the onset of fate determination (Figure EV2D-G). But what is the intrepretation for Jag1 being expressed high in the expansion and gliogenesis periods? Were any genes expressed high in the two early or the two late phases? This is not currently reflected in Fig EV2. Also, given previous conclusions that most genes are progressively downregulated along the cell lineage, to know what are the proportions of DEGs following each temporal trend would be very informative.

In page 9, the authors indicate that NBNs showed less transcriptional dynamics over time. However, their PCA data shows that these are all bunched together, except for those in gliogenesis.

The authors report that genes including Dlx1, Dlx5 and Dlx2 separated NSCs (page 9). Expression data for Dlx1,2 and 5 are contrary to public ISH (Genepaint) and scRNAseq (Humous) datasets in terms of expression timing along the cell lineage in the embryonic mouse neocortex. In fact, Dlx genes are well known as markers of subpallial lineage.

Data provided in Figs 3E and 3I are interesting and very relevant to the study, but these figures are excessively complex and poorly intuitive. The manuscript would greatly benefit from a significant improvement at this level. Is there an order or sense in the location of items in space? I don't understand the colored scale "time-cell type". Does it refer to the little frames for the plots? Not clear.

"The single-cell transcriptomes of highly variable genes (HVGs) revealed a low heterogeneity within the NSCs during expansion and gliogenesis (Figure 4A)." Why was this restricted to HVGs? How were these defined? Even if these genes carry most of cellular variability, the authors should present a PCA analysis of all cells with all their genes. In Fig 4D, the authors should show UMAPs, for cluster analysis and feature plots, as it is well known that UMAPs represent transcriptional proximity much better than tSNE plots.

"the single cell transcriptomes reflected the heterogeneity of the population at the respective time point (Figure 4B)." This is not clear in Fig 4B. This panel does not show heterogeneity of NSCs at each time point, only between time points.

Related to my main comment above, do clusters for NSCs, BPs and NBNs segregate simlarly well when analyzed together? This is very important for the value of these findings as a resource, because other future/existing studies mostly will not be performed using these reporter mouse lines, so the usefullness of this study is only as good as the independency from the use of these mouse lines.

In Figure 4H there is no feature map for a gene characterizing cluster 2

The authors conclude that "The scRNA-Seq data enabled a high-resolution definition of gene signatures for each cluster (cell type) of NSCs, BPs and NBNs." However, the limited single cell heterogeneity of NBNs is very surprising (2 clusters?!), potentially due to the very small numbers of cells sampled, as those studied have a significant degree of variability, as seen in PCA and t-SNE plots. The sentence "These findings demonstrate an unprecedented heterogeneity in NSCs, BPs and NBNs over time and a dynamic shift in gene expression of these cells at the population and single-cell levels" is definitely inappropriate to describe the actual findings.

What genes and gene functions distinguish NBN clusters? Are these corresponding with known markers of upper and lower layers? Or functions of upper and lower layers neurons?
Do these results indicate that neuron subtypes are already specified at NBN stage? Or further subspecified later on? More should be interpreted in Discussion (¿?)

"This suggested that the transcriptional program that defines cortical neuron subtypes is initiated in NSCs long before their exit from cell cycle." This conclusion is glaringly different from previous studies (i.e. Telley 2019), and should be highlighted much more and discussed appropriately. It would be good to overlap the datasets from NSCs, BPs and NBNs on expression of individual genes, to compare relative expresion levels and their temporal dynamics across cell types.

In Discussion, the authors state (second paragraph) that "Here we posed the questions of how gene expression changes in stem cells, progenitors, and newly formed neurons over time and whether we can uncover distinct traits and patterns within specific cell-types defined based not on an ad hoc identification using a selection of RNA transcripts". However, this is precisely what was done here using the reporter mouse lines.

Referee #2:

In the present study, Mukhtar et al. performed RNA-seq of murine cortical NSCs, BPs, and NBNs from E10.5 to P0 and uncovered the transcriptional heterogeneity of cortical neural lineages. They identified potential intercellular interactions and signaling pathways that may play critical roles in cortical development. Moreover, they found transcriptional programs were highly dynamic over time and respond differently to signals. In general, this study is of great value for understanding biological processes underlying cortical development and neurological disorders. In its current form, however, there are several points that need to be addressed prior to publication.
Major concerns:
1. In Figure 2B, the author claimed that NSCs in the expansion phase preferentially expressed blood cell-related genes Hbb-bh1, Hba-x, and Hbb-y. what would the expression of these genes in the scRNA-seq dataset look like? As a control for sample quality, it would be helpful to show the expression of housekeeping genes and conduct an integration analysis with the publically available dataset to see if similar signatures could be observed in the published dataset.

2. It is not very clear how the regulatory networks in Figure 3E and I were constructed, please provide the supporting statistical calculation evidence in the supp. material.
conventional methods widely used by Seurat and Scanpy?

3. Figure 3E is interesting in the sense of showing the different potential of NSC but quite confusing. What is the exact value of each axis? Why was the number of points in each panel different?

4. In Figure4, the Mukhtar et al. uncovered heterogeneity within NSCs, BPs, and NBNs from scRNA-seq data. NSCs were clustered into 5 sub-groups and showed higher cellular diversity, while 2 NBN sub-groups were identified. Are there any associations between NSCs and NBN subclusters? How do NSCs contribute to the diversity of NBNs?

5. In Figure 5D, how was the scale bar for each image computed, the size of cells from each image varies.

Minor points:
1. In Figure 1 C and E, it would be helpful if the scale bar for both heatmaps could be consistent.
2. On page 12, please correct "Figure S5A" to "EV5A"
3. In the legend of Figures 5D, F, and H, a description of the scale bar should be added.
4. In Figures 6B, D, and Figure 7B-D, it would be helpful to add a figure legend beside the graphs to show what each color represents?
5. In Figure EV1 E and F, the scale bar is missing.
6. In Figure EV2 C and Figure EV3 D, please provide a description of the statistical test in the figure legend.

Referee #3:

This well written and beautifully illustrated study uses state of the art sequencing techniques and bioinformatics to address an important question in the field, how do neural progenitors in the developing cerebral cortex change over time during the phases of expansion, neurogenesis, and gliogenesis, and how does this relate to the diversity of cells in the mature cerebral cortex. The authors use GFP reporters to isolate mouse NSCs, BPs, and NBNs for bulk and single cell sequencing and generate a data-set which will be a valuable resource for the community.
The most interesting part of the study is the discovery that the molecular properties of these cells changes with time and, most important, that the transcriptome fingerprints of progenitors matches the cells they will differentiate into.
This last claim (data shown in Figure 5) is critical to the significance of this study and could be better supported, for example by comparing their scRNAseq data for progenitors with published single cell mouse transcriptomes for mature cell types corresponding to different cortical layers using CCA analysis (or similar) to demonstrate how the progenitor transcriptome fingerprints correlates to differentiated cell type transcriptome fingerprints and whether this supports their central hypothesis.

We thank the reviewers for their support and constructive suggestions. We have taken their comments to heart and addressed each and every point. We have made major additions and changes which have improved our manuscript and improved the readability. Below we address the comments point-by-point.

Referee #1:

This is a resource paper analyzing the transcriptomic landscape of the developing mouse cerebral cortex along the entire neurogenic and early gliogenic periods, distinguishing separately between Radial Glia Cells, Intermediate Progenitor Cells and Newborn Neurons. The authors use two different transgenic reporter mouse lines to isolate these populations based on their expression of GFP under the specific promoter of Hes5 and Tbr2, where IPCs and NBNs are distinguished by high and low Tbr2::GFP levels. The study is very carefully performed and provides an in-depth characterization of the transcriptional landscape of these cell types at the population levels, and their heterogeneity at the single cell level. The most relevant distinction of this study over others is the identification of the transcriptome of specific cell populations without a priori knowledge or establishment of marker genes (although it is based on Hes5 and Tbr2 expression). This is important to ascribe cell identities based on transcriptome without bias on a priori knowledge of marker proteins, which commonly leads to misidentifications of cell types. Unfortunately, while this value of the study is unquestionable, the novelty that it brings over previous similar analyses (i.e. Telley et al 2019, Science) is insufficiently highlighted in the manuscript, and it may be not obvious to the reader.

One of the main caveats of this study is the extent of its utility as a resource for other scientists in the field, because all the results are based on the use of the two transgenic reporter mouse lines, which other studies will likely not use. The identification of cell groups (clusters) in single cell transcriptomic studies is strongly influenced by the collection of input cells, and thus the conclusions reached here may not be easily translatable to studies analyzing all cell classes in the embryonic cortex combined. In this sense, it would be a very significant improvement if the authors compared their results with data from non-reporter mice (i.e. publicly available data from other studies) to demonstrate that the cell populations identified here, as well as their heterogeneity and temporal evolution, are distinguishable without the need of using the Hes5 and Tbr2 reporters.

We thank the reviewer for their supportive and insightful comments. In order to make the relevance of our data clearer, we have now added more details about our major findings in the results and discussion. We took the reviewer's advice and have now included a comparison of our single cell C1 data with the extensive Linnarsson 10X genomics dataset in the revised manuscript (La Manno et al., 2021) (revised Figure EV8). We performed two major comparisons with the Linnarsson developing mouse brain dataset: first, we compared all our NSCs, BPs and NBNs to their forebrain and dorsal forebrain cells (E9-E18), second, we compared our NSCs, with their radial glial cells from the same regions.

In addition, we repeated our single cell analyses and used KNN graph-based clustering and UMAP visualizations in addition to the previous K-means clustering and tSNE visualizations. In these new analyses, we see similar clustering of individual cell types as we showed previously, and together we identify 8 well characterized clusters of cells (revised Figure 5). We then performed CCA integration analyses on the Linnarsson cells and our C1 data. Our cells integrated as expected into the Linnarsson dataset with the expression of distinct marker genes. We identified 10 distinct cell clusters, and our cells maintained their distinct groupings even in these analyses (revised Figure EV8B). UMAP visualization reveals our cells segregate within the Linnarsson dataset post-CCA integration. These analyses also revealed that the cells we define as NSCs and NBNs in our data fall within the clusters defined by Linnarsson as radial glial and neurons, respectively. Hence, these findings validate that both methods identify the same cells but indicate that the increased depth of C1 data allows the identification of subtypes not discernable by 10X sequencing.

The Linnarsson dataset does not identify BPs as a separate population and, indeed, our BPs fall into the neuron clusters defined by Linnarsson. However, our BP cells do overlap with Tbr2 expressing cells in the Linnarsson data (revised Figure EV8F). These findings support the added value and power of the deeper SmartSeq2 C1 sequencing approach. We integrated our NSC cells with the Linnarsson cells classified as radial glial cells and found distinct subpopulations of mainly dividing early and late progenitors revealed by the expression of marker genes. We also found that the Linnarsson 10X dataset shows low expression of the “mature” genes of radial glia cells compared to our C1 data, which we interpret as being due to the shallow sequence depth of 10X.

We agree that others have presented extensive single-cell sequencing resources from the developing mouse brain. We consider the use of transgenic *Hes5::GFP* and *Tbr2::GFP* lines as a major strength of the paper because we are able to isolate clean populations of NSCs, BPs and NBNs using these lines. Our aim was to enrich for the progenitors, and newborn neurons and interrogate their underlying temporal transcriptional dynamics and cellular heterogeneity. The excellent suggestion of the reviewer to compare our C1 data with a 10X resource highlights one of the caveats of 10X genomics, the relatively low read-depth compared to SmartSeq data. In addition, this comparison demonstrated the greater sensitivity of our data for identifying heterogeneity and cell subpopulations. We feel that the new comparisons made to the Linnarsson dataset validate our data and its value as an additional resource for the community.

Other major points:

Fig EV1B should show also DAPI for reference of the cortical layers. Labeling with Tbr2 seems strange, as cells expressing endogenous Tbr2 protein are usually spread apically from SVZ into VZ (see Englund et al 2005, Fig 1), but Fig EV1B indicates otherwise. Hes5::GFP clearly labels cells only in VZ, but cells in the basal part of VZ are much less bright (or negative). Does this marker label a selection of RGCs? Validation

in Fig EV1D should include quantifications, as well as co-expression of Pax6 and Tbr2 protein in cycling cells.

We thank the reviewer for the comment, and in the revised manuscript we include images with Dapi in Figure EV1B. The reviewer is correct, Tbr2 is expressed by BPs which are formed in the VZ and migrate out of the VZ along the radial processes of NSCs. This is evident in the *Tbr2::GFP* images in Figure EV1B (arrowheads). We present new images of GFP, Pax6 and Tbr2 of E17.5 *Hes5::GFP* and *Tbr2::GFP* embryonic cortices in Figure EV1B. All Pax6 cells in cortical VZ of *Hes5::GFP* embryos are GFP positive (arrows) which confirms our previous data (Basak and Taylor, 2007; Lugert et al., 2012; Zhang et al., 2019). Pax6 is not exclusively expressed by NSCs but also by BPs in the SVZ. This is also evident from the Pax6 mRNA expression profile in Figure 1C. We also show Pax6, and Tbr2 immunostaining of *Tbr2::GFP* embryonic cortices (arrowheads in Figure EV1B).

In page 6, the authors indicate that "Principal component analysis (PCA) capturing 60% of the total variance (PC1 and PC2) separated the samples based on cell type (NSC, BP and NBN) and developmental stage.". However, this separation is minimal for BPs and NBNs. Also BPs overlap completely with NSC mid-neurogenesis. This is very relevant because variance explained by PC1 is quite low, and PC1+2 is only 60% (very low).

We thank the reviewer for the comments. Firstly, with all respect, we do not agree 60% of the total variance across the samples is very low. Rather the contrary, this indicates that the majority of the differential gene expression between NSCs, BPs and NBNs is captured by the first two PCs. However, we must also point out that the value itself is not particularly relevant. What is more important is that the first PC almost perfectly separates the three cell types (with NBNs clearly separable from the BPs and NSCs. Along the first PC, only a few NSCs overlap with the BPs during the neurogenic period. The second PC orders the cells of all three types according to developmental time (i.e., the time axis runs in the same direction for all three cell types). This is remarkable because there is *a priori* no reason for the first two PCs to correspond to cell type and time so cleanly. In the revised manuscript, we have now pointed out that the BPs and NSCs from mid-neurogenesis project to overlapping positions on the first two PCs.

The authors indicate that "DEG analyses revealed that the majority of the highly expressed genes in NSCs are downregulated by BPs and reduced further by NBNs. (page 7)". This is a fascinating finding not previously stressed in other studies. Does this mean that a large number of NSC genes must be turned off so that already expressed "neuronal" genes can exert their function in BPs and nascent neurons? These trends in gene expression, shown in Fig 1E, would be better identified if the heatmap was ordered in a hierarchical manner.

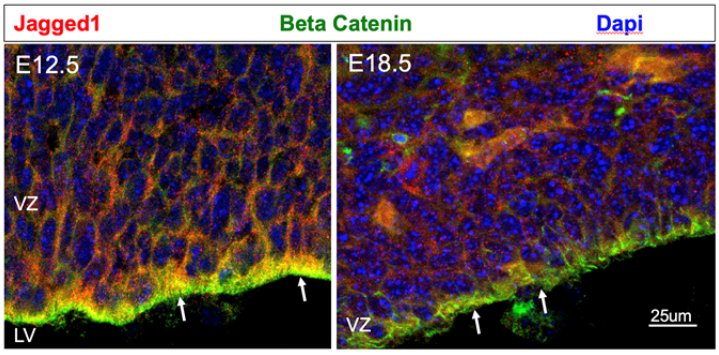
We agree these trends in gene expression are interesting. We identified many neuronal RNAs expressed by NSCs with no detectable protein expression and think there could be essential post-transcriptional regulation programs active during neurogenic differentiation. Several mechanisms of post-transcriptional regulation have previously

been reported including non-canonical regulation by Drosha (Knuckles et al., 2012), and m^6 RNA methylation (Yoon et al., 2017). It is likely that NSC genes turn off and more neurogenic genes are activated in BPs and NBNs upon differentiation. A systematic ATAC-seq or multiome analyses may be able to address these biological questions in future. In addition, validating the different transcriptional nodes predicted by ISMARA in Figure 3E and 3I will be an interesting extension of our findings. We have included comments to this end in the discussion of the revised manuscript.
In accordance with the reviewer's suggestions, we have re-arranged the heatmap in Figure 1E to remain consistent with other heatmaps in the manuscript.

…. when (and how) do the many different types of cortical neurons become distinct? The authors should at least speculate in Discussion on this question.
The reviewer raises a very interesting and important question. However, as an mRNA sequencing is a snapshot of gene expression and not a lineage trace, it is difficult to unequivocally answer this question with sequence data. However, our data do provide a potential insight into the timepoints when the different neurons become distinct.
When we assess the distribution of deep layer and upper layer cortical marker expression by NSCs, BPs and NBNs, we observe distinct temporal patterns (Figure EV9).
NSCs move from expressing deep to both deep and upper layer neuron genes and subsequently to increased expression of upper layer neuron markers. We observe a similar trend in neuronal gene expression dynamics in the BPs throughout the course of corticogenesis but with a 24-48 hour delay compared to the gene expression by NSCs. Our findings provide hints about some of the intrinsic programs active in NSCs and BPs that regulate neuronal fate determination, but further functional assays and temporal lineage tracing are required to understand the commitment to neuron subtype fates.

The authors report that expression of Shh pathway genes mark specifically the early expansion phase and their expression is low upon the onset of fate determination (Figure EV2D-G). But what is the interpretation for Jag1 being expressed high in the expansion and gliogenesis periods? Were any genes expressed high in the two early or the two late phases? This is not currently reflected in Fig EV2. Also, given previous conclusions that most genes are progressively downregulated along the cell lineage, to know what the proportions of DEGs are following each temporal trend would be very informative.
We thank the reviewer for this comment. We observe trends in gene expression correlating with specific phases of NSC expansion, neurogenesis and gliogenesis. In addition, we defined dynamic gene expressions as upregulated, downregulated, transient upregulated, and transient downregulated, for NSCs, BPs and NBNs (EV Table 2: Clustering NSCs, Clustering BPs and Clustering NBNs tabs). Their expression plots are also available on the NeuroStemX website (http://neurostemx.ethz.ch/). We have also included the proportions of these gene expression profiles on the respective tabs.

Jagged1 is a canonical Notch ligand, and it is involved in both expansion/neurogenic and gliogenic time points. Jagged1 (Jag1) is expressed at the apical end-feet of NSC, similar to what has been observed for Notch1 (Nyfeler et al., 2005). We have ongoing projects in the lab with preliminary data addressing the function of Jag1 during early and late developmental stages of the brain. We speculate that Jag1 plays different roles in regulating neurogenic and gliogenic periods of corticogenesis. We hope that the reviewer accepts at, as we do not have concrete experimental evidence for these roles of Jag1, we are reluctant to speculate in this manuscript.



Representative immunostaining of dorsal cortical NSC apical end-feet with Jag1 antibody at E12.5 and E18.5, co-immunostained with β-catenin antibody (arrows).

In page 9, the authors indicate that NBNs showed less transcriptional dynamics over time. However, their PCA data shows that these are all bunched together, except for those in gliogenesis.

We thank the reviewer for the comment as the analysis may not have been clear. In the PCA where the NBNs show less transcriptional dynamics (Figure 2E), the analysis is an orthogonal PCA where the first two PCs which account for temporal variance have been removed. We removed the first two PCs as these were strongly dictated by the NSC populations and masked the variance within the NBNs which are more similar to each other than they are to NSCs. However, we do observe transcriptional dynamics in NBNs in Fig 2M when PC1 and PC2 are included.

The authors report that genes including Dlx1, Dlx5 and Dlx2 separated NSCs (page 9). Expression data for Dlx1,2 and 5 are contrary to public ISH (Genepaint) and scRNAseq (Humous) datasets in terms of expression timing along the cell lineage in the embryonic mouse neocortex. In fact, Dlx genes are well known as markers of subpallial lineage.

We thank the reviewer for highlighting this point and we agree the Dlx genes are commonly used as markers of subpallial lineage. We observe the expression of Dlx genes in NSCs and BPs from our bulk-RNA sequencing data. To investigate this further, we analyzed our single-cell data and found Dlx transcripts as well as other inhibitory markers expressed in a few neurogenic NSCs (revised Figure EV6E). We addressed the expression of Dlx genes in the Linnarsson dataset (La Manno et al., 2021) comparing their radial glia with our NSCs and found similar low expression of Dlx genes (revised Figure EV8F).

Interestingly, recent studies in human cortex using clonal analyses highlights the possibility of a local dorsal origin of inhibitory neurons (Delgado et al., 2022). We speculate that the Dlx+ NSCs we observe in our C1 data and the Linnarsson dataset could represent mouse counterparts of the human local, dorsal inhibitory neuron progenitors shown by (Delgado et al., 2022). Further experimental analysis and accurate lineage tracing will be needed to clarify this.

Data provided in Figs 3E and 3I are interesting and very relevant to the study, but these figures are excessively complex and poorly intuitive. The manuscript would greatly benefit from a significant improvement at this level. Is there an order or sense in the location of items in space? I don't understand the colored scale "time-cell type". Does it refer to the little frames for the plots? Not clear.

We apologize that the original Figure 3E and 3I were difficult to follow. In the revised manuscript we have completely rewritten this section in the paper and reworked Figure 3. In addition, we added a supplementary methods text to better explain how the analysis was done.

To summarize briefly: Figure 3E shows the core regulatory network predicted by ISMARA (Balwierz et al., 2014) for the NSCs, with each node corresponding to a TF binding motif, the curve in each rectangular node of this network showing the motif activity profile across developmental time, and the edges showing predicted regulatory interactions between pairs of core TF motifs. TF motif nodes are color-coded depending on whether they are most active during the expansion phase (red), the neurogenesis phase (green), or the gliogenesis phase (blue). This is now explained in the revised manuscript. We also describe more clearly the observation that almost all predicted regulatory interactions involve neurogenic TFs targeting their motifs associated with the expansion phase are repressive or target TF motifs associated with gliogenesis are most activating. Note that the motifs were sorted in Figure 3I (now Figure 3H) such that motifs most active during the expansion phase are on the left, motifs most active during neurogenesis and gliogenesis on the right. The colors in new Figure 3H indicate whether the motifs vary mostly along PC1 of Figure 3F (and are thus distinguished mostly by cell type) or along PC2 (and are thus distinguished mostly across developmental time).

Finally, in the revised manuscript, we now also note that the full ISMARA analysis of these data is available through the URL (https://ismara.unibas.ch/NeuroStemX) linking to the ISMARA website.

"The single-cell transcriptomes of highly variable genes (HVGs) revealed a low heterogeneity within the NSCs during expansion and gliogenesis (Figure 4A)." Why was this restricted to HVGs? How were these defined? Even if these genes carry most of cellular variability, the authors should present a PCA analysis of all cells with all their genes. In Fig 4D, the authors should show UMAPs, for cluster analysis and feature plots, as it is well known that UMAPs represent transcriptional proximity much better than tSNE plots.

We used HVGs for single cell mRNA sequencing analyses following (Luecken and Theis, 2019) and first used by (Brennecke et al., 2013). Focusing on HVGs accounts for the intrinsic technical noise in single cell mRNA sequencing. (Klein et al., 2015) also showed that the downstream analyses of single cell sequence data are more robust when using an exact choice of number of HVGs. The consensus in the field is that HVG numbers varying between 200-2400 does not affect the dimensional representation in the PCA space. To test this with our data, we have now repeated the PCA with 500, 1000 and 2000 HVGs and found the dimensional representation and clustering remains the same.

We have also repeated our analyses for all cell types and show UMAPs in revised Figures EV6, EV7. With UMAP visualization, we also identify 5 clusters for NSCs, 3 for BPs and 2 for NBNs. As requested by the reviewers, we have included the new UMAP plots, feature plots and analyses for NSCs, BPs and NBNs in the revised manuscript (Figures 5, EV6, EV7). We also performed Slingshot analyses and elucidate the neurogenic trajectories from NSCs to BPs to NBNs (Figure 5).

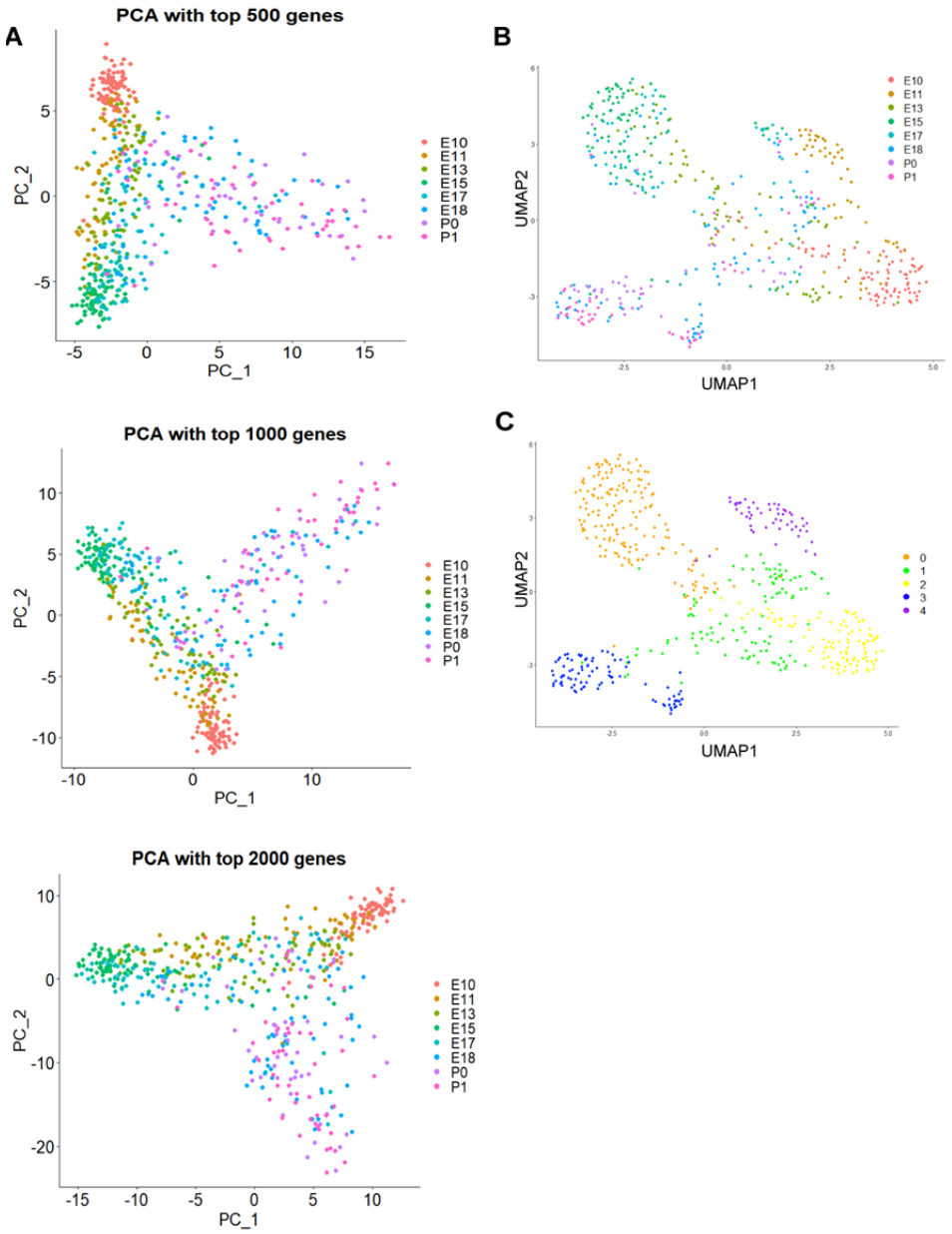


Figure: PCA for NSCs with top 500, 1000 and 2000 highly variable genes.

(A) PCA of NSCs with top 500, 1000 and 2000 highly variable genes (HVGs) to test for differences in dimensionality.

(B) UMAP visualization of clustering of NSCs based on time points.

(C) UMAP visualization of NSCs based on clusters, 5 clusters similar to the results shown with the previous PCA analysis.

"the single cell transcriptomes reflected the heterogeneity of the population at the respective time point (Figure 4B)." This is not clear in Fig 4B. This panel does not show heterogeneity of NSCs at each time point, only between time points.

We thank the reviewer to point out the typo. The comment should have referred to Figure 4A. Cellular heterogeneity is shown in Figure 4A, while Figure 4B shows the average of the single NSC sequences projected onto the bulk data. We have corrected this in the revised text.

Related to my main comment above, do clusters for NSCs, BPs and NBNs segregate similarly well when analyzed together? This is very important for the value of these findings as a resource, because other future/existing studies mostly will not be performed using these reporter mouse lines, so the usefulness of this study is only as good as the independency from the use of these mouse lines.

This is a very interesting question. As proposed, we have analyzed the three cell types together and visualized these by UMAP. In this complex analysis, we identify 8 clusters in total, four clusters for NSCs, two for BPs and two for NBNs. We identify robust temporal segregation of the NSCs, which is driven by their putative fate potential. NSCs segregate in four major clusters- NSC1 (expansion), NSC2 (late, astrocytic), NSC3 (late, oligodendrocytic) and NSC4 (intermediate, neurogenic). BPs segregate into two clusters, BP1 enriched in Bcl11b+ cells (Ctip2+, deep neuron layer marker) and a BP2 cluster enriched in Pou3f2+ cells (Brn2+, upper neuron layer marker). The NBNs divide into two clusters (NBN1 and NBN2), similar to the individual NBN analysis, the clustering is not driven by cortical layering markers. We have included the new UMAP plots, example feature plots, and analyses in the revised manuscript as a new main Figure 5. We also performed Slingshot analysis for pseudo-time analysis and elucidate the neurogenic trajectories from NSCs to BPs to NBNs (Figure 5). We have identified the markers for each cluster and all the lists are updated in EV Table 6.

In Figure 4H there is no feature map for a gene characterizing cluster 2

Thank you for pointing this out, this was an oversight. We have added an example feature plot for cluster 2 in Figure 4H. The gene lists for these clusters are in EV Table 4.

The authors conclude that "The scRNA-Seq data enabled a high-resolution definition of gene signatures for each cluster (cell type) of NSCs, BPs and NBNs." However, the limited single cell heterogeneity of NBNs is very surprising (2 clusters?!), potentially due to the very small numbers of cells sampled, as those studied have a significant degree of variability, as seen in PCA and t-SNE plots. The sentence "These findings demonstrate an unprecedented heterogeneity in NSCs, BPs and NBNs over time and a dynamic shift in gene expression of these cells at the population and single-cell levels" is definitely inappropriate to describe the actual findings.

We do have a limited number of time points for NBNs in our dataset and observe most of the heterogeneity in NSCs. We have reworded this sentence.

What genes and gene functions distinguish NBN clusters? Are these corresponding with known markers of upper and lower layers? Or functions of upper and lower layers neurons?

Do these results indicate that neuron subtypes are already specified at NBN stage? Or further subspecified later on? More should be interpreted in Discussion (¿?)

NBNs divide into two distinct clusters which are not determined by cortical layering markers and are instead driven by a time component. The genes separating the clusters

are shown in EV Table 4. Using Metacore analysis, we did not observe any major gene functions that distinguish the two clusters at the single cell level. This could be because of the limited time points collected, or because we have NBNs and not mature neurons. This may also mean the neuronal subtype is definitively specified later, during maturation. Interestingly, the TFs of the cortical neurons in subsequent layers have been shown to have a negative regulatory role on the expression of each other to refine the final fate. We have reinforced this point in the discussion in the revised manuscript.

"This suggested that the transcriptional program that defines cortical neuron subtypes is initiated in NSCs long before their exit from cell cycle." This conclusion is glaringly different from previous studies (i.e. Telley 2019), and should be highlighted much more and discussed appropriately. It would be good to overlap the datasets from NSCs, BPs and NBNs on expression of individual genes, to compare relative expression levels and their temporal dynamics across cell types.
We thank the reviewer for the comment and totally agree that the suggestive transcriptional program active in NSCs is an interesting finding. We reanalyzed the Linnarsson dataset (La Manno et al., 2021) and made similar observations to those we found with our C1 data (as described above). We also performed CCA integration analyses on the Linnarsson forebrain and dorsal forebrain cells and our C1 data using KNN for cluster identification and identified 10 clusters (Figure EV8B). We also include feature plots as examples in Figure EV8F to show the enrichment of known markers including Dcx, Tubb3, suggesting the start of neuronal transcriptional programs in NSCs is prior to their birth. We also show expression of relevant genes including Bcl11b (Ctip2), Pou3f2 (Brn2) in both datasets.

In Discussion, the authors state (second paragraph) that "Here we posed the questions of how gene expression changes in stem cells, progenitors, and newly formed neurons over time and whether we can uncover distinct traits and patterns within specific cell-types defined based not on an ad hoc identification using a selection of RNA transcripts". However, this is precisely what was done here using the reporter mouse lines.
We have reworded this sentence in the discussion.

Referee #2:

In the present study, Mukhtar et al. performed RNA-seq of murine cortical NSCs, BPs, and NBNs from E10.5 to P0 and uncovered the transcriptional heterogeneity of cortical neural lineages. They identified potential intercellular interactions and signaling pathways that may play critical roles in cortical development. Moreover, they found transcriptional programs were highly dynamic over time and respond differently to signals. In general, this study is of great value for understanding biological processes underlying cortical development and neurological disorders. In its current form,

however, there are several points that need to be addressed prior to publication.

Major concerns:
1. In Figure 2B, the author claimed that NSCs in the expansion phase preferentially expressed blood cell-related genes Hbb-bh1, Hba-x, and Hbb-y. what would the expression of these genes in the scRNA-seq dataset look like? As a control for sample quality, it would be helpful to show the expression of housekeeping genes and conduct an integration analysis with the publically available dataset to see if similar signatures could be observed in the published dataset.
We thank the reviewer for their constructive comments. We observed the expression of Hbb sub-units in the bulk-RNA sequences. Brown *et al*., 2016 have previously observed neuronal Hbb expression (Brown et al., 2016). In response to the reviewer to address this question further, and analyze the expression of Hbb genes, we investigated our single cell data. We found Hbb genes are expressed by NSCs in clusters 2 and 4, which correspond to the early E10.5 and E11.5 time points (revised Figure EV6B).
In the revised manuscript, we repeated the single cell analyses and now visualize the cells by UMAP clusters in addition to the previously shown PCA. When we analyze NSCs, BPs, and NBNs together, they segregate in 8 clusters (revised Figure 5). We performed CCA integration analyses on Linnarsson radial glia cells (La Manno et al., 2021) and identified 5 NSC clusters (revised Figure EV8G). We then compared their radial glia with our NSCs and found our NSCs fall within their radial glia cell definition (revised Figure EV8I). We now show some examples of relevant feature plots of genes such as Crabp2, Hmga2, Bcl11b (Ctip2), Pou3f2 (Brn2), Aqp4, Olig2 etc. as observed previously in NSCs, expressed in both datasets (revised Figure EV8J). In addition, in the revised manuscript, we show feature plots of Hba-x, Hbb-bh1 and Hbb-y detected in both datasets (revised Figure EV8K).

2. It is not very clear how the regulatory networks in Figure 3E and I were constructed, please provide the supporting statistical calculation evidence in the supp. material. conventional methods widely used by Seurat and Scanpy?
As indicated in our answer to reviewer #1 above, we agree that in the initial manuscript, we failed to properly explain the analysis that was done in Figure panels 3E and 3I (now Figure 3H). In the revised manuscript, not only we rewrote the text for the corresponding section, but also reworked the figure and added a supplementary methods section explaining the analysis that was done. Please see response to reviewer 1. The methods that were used here are mainly based on the Motif Activity Response Analysis that was developed in the van Nimwegen lab (Balwierz et al., 2014; Consortium et al., 2009).

3. Figure 3E is interesting in the sense of showing the different potential of NSC but quite confusing. What is the exact value of each axis? Why was the number of points in each panel different?
In the revised manuscript, we completely reworked the Figure 3 and the relevant text. We have made clear what is shown along each axis in Figure 3E. We refer the reviewer to the extensive answer given to reviewer #1 above. To answer the reviewer's specific

question about number of points in each panel: the 3 curves shown for each regulatory network node of Figure 3H correspond to the activities of the TF motif across developmental time for the 3 cell types, i.e., with NSCs in green, BPs in red, and NBNs in purple. We note that the number of available time points is different for NSCs, BPs, and NBNs. In the revised manuscript, we have included a link to the ISMARA analyses for the entire dataset (https://ismara.unibas.ch/NeuroStemX).

4. In Figure4, the Mukhtar et al. uncovered heterogeneity within NSCs, BPs, and NBNs from scRNA-seq data. NSCs were clustered into 5 sub-groups and showed higher cellular diversity, while 2 NBN sub-groups were identified. Are there any associations between NSCs and NBN subclusters? How do NSCs contribute to the diversity of NBNs?

In the revised manuscript, we repeated our single cell analyses and in accordance with the reviewer's suggestions, visualize the cells by UMAP in addition to the previously shown PCA. When we analyze NSCs, BPs, and NBNs together, they segregate in 8 clusters (revised Figure 5). We have now performed trajectory analyses using Slingshot, to determine lineage relationships among these cells (revised Figure 5E-G).

We aimed to determine the neurogenic lineage starting from NSCs to NBNs, through BPs. We observed NSCs going into BPs, which in turn generate two distinct clusters of NBNs following a time component. We have included this data into a new figure in the revised manuscript (Figure 5A, B, E-G). The 5 NSC clusters were identified and discriminated by time and not on the basis of neuron subtype specific genes. Although, in Figure EV9, when we compare the temporal distribution of deep layer neuron versus upper layer neuron gene expressing NSCs, BPs and NBNs, we observe that NSCs move from deep to both deep and upper layer neuron marker expression and then to upper layer neuron marker expression. This indicates that early NSCs express the deepest layer neuron markers and very few express upper neuron markers. Later, during corticogenesis, both types NSCs are present - those expressing deep layer neuron markers and upper layer neuron markers. We observed a similar trend in the BP populations, with a more restricted deep layer neuron expression early, and a broader deep and upper layer neuron marker expression during the course of corticogenesis. In NBNs, we mostly observe upper layer marker expression possibly because we collected single NBN cells from E16.5 onwards when upper layers are being formed. These findings suggest some intrinsic programs are active in these cells, but further functional assays are required to understand the crosstalk between the intrinsic and extrinsic programs in controlling neuronal fate.

5. In Figure 5D, how was the scale bar for each image computed, the size of cells from each image varies.

We used Fiji to import the metadata of the scale bar from the raw Zeiss .czi files. We have noticed that NSCs and BPs are relatively smaller than NBNs. This was evident also by FACS.

Minor points:
1. In Figure 1 C and E, it would be helpful if the scale bar for both heatmaps could be consistent.
We have updated these in the revised manuscript.

2. On page 12, please correct "Figure S5A" to "EV5A"
Thank you for pointing out this typo. We have updated this in the revised manuscript.

3. In the legend of Figures 5D, F, and H, a description of the scale bar should be added.
This was an oversight on our part. We have added this in the revised manuscript.

4. In Figures 6B, D, and Figure 7B-D, it would be helpful to add a figure legend beside the graphs to show what each color represents?
We have included the information in the revised figure legends.

5. In Figure EV1 E and F, the scale bar is missing.
We have added these in the revised manuscript.

6. In Figure EV2 C and Figure EV3 D, please provide a description of the statistical test in the figure legend.
We have added this in the revised manuscript.

Referee #3:

This well written and beautifully illustrated study uses state of the art sequencing techniques and bioinformatics to address an important question in the field, how do neural progenitors in the developing cerebral cortex change over time during the phases of expansion, neurogenesis, and gliogenesis, and how does this relate to the diversity of cells in the mature cerebral cortex.
The authors use GFP reporters to isolate mouse NSCs, BPs, and NBNs for bulk and single cell sequencing and generate a data-set which will be a valuable resource for the community.
The most interesting part of the study is the discovery that the molecular properties of these cells change with time and, most important, that the transcriptome fingerprints of progenitors matches the cells they will differentiate into.
This last claim (data shown in Figure 5) is critical to the significance of this study and could be better supported, for example by comparing their scRNAseq data for progenitors with published single cell mouse transcriptomes for mature cell types corresponding to different cortical layers using CCA analysis (or similar) to demonstrate how the progenitor transcriptome fingerprints correlates to differentiated cell type transcriptome fingerprints and whether this supports their central hypothesis.
We thank the reviewer for their positive comments and support.

We agree with the reviewer that the expression of neuronal mRNA lineage transcripts by NSCs and BPs before the start of the neurogenic program suggests the significance of post transcriptional regulation and the need to study this in the future.

We note that already the PCA analysis of the bulk RNA-seq samples shows that not only are the gene expression profiles of NSCs during neurogenesis closest to those of the corresponding BPs, and NBNs, but also that all three cell types change their gene expression state in the same direction as development progresses (i.e., along PC2). The same phenomenon is also apparent in the single cell RNA sequence data, we have edited the text in the revision to stress this point.

We have re-analyzed our single cell sequence data for NSC, BP, and NBN together and identify 8 distinct clusters which we visualize by UMAP (revised Figure 5). We also performed trajectory inference analysis using Slingshot, identifying 3 neurogenic lineages starting from our NSCs to NBNs through BPs.

We performed CCA integration of our dataset with the Linnarsson dataset (La Manno et al., 2021) and identified 10 distinct clusters. When we visualize our C1 cells with the Linnarsson dataset post CCA integration, we find that our NSCs and NBNs fall within the expected radial glial cell and neuronal clusters respectively. We also find that our BPs fall into clusters with cells classified as neurons in the Linnarsson dataset. As their dataset does not identify BPs, but we find our BPs fall correctly within clusters expressing markers for BPs including Tbr2 (revised Figure EV8F). Our C1 sequenced cells integrate as expected into the Linnarsson dataset with our clusters maintaining their distinct groupings. While we observe distinct expression of genes for the neurogenic and gliogenic NSCs in our data, we do not see correlative expression of mature cortical subtype markers as identified in (Zeisel et al., 2018), shown below.

These lines of evidence all point towards an active transcriptional program early in progenitors/NSCs that distinguish broad neurogenic and gliogenic fates. However, the separation of mature cortical subtypes cannot be ascertained in these earlier timepoints and are likely driven by later processes in maturation and extrinsic signaling cues. We discuss this further in the revised manuscript.

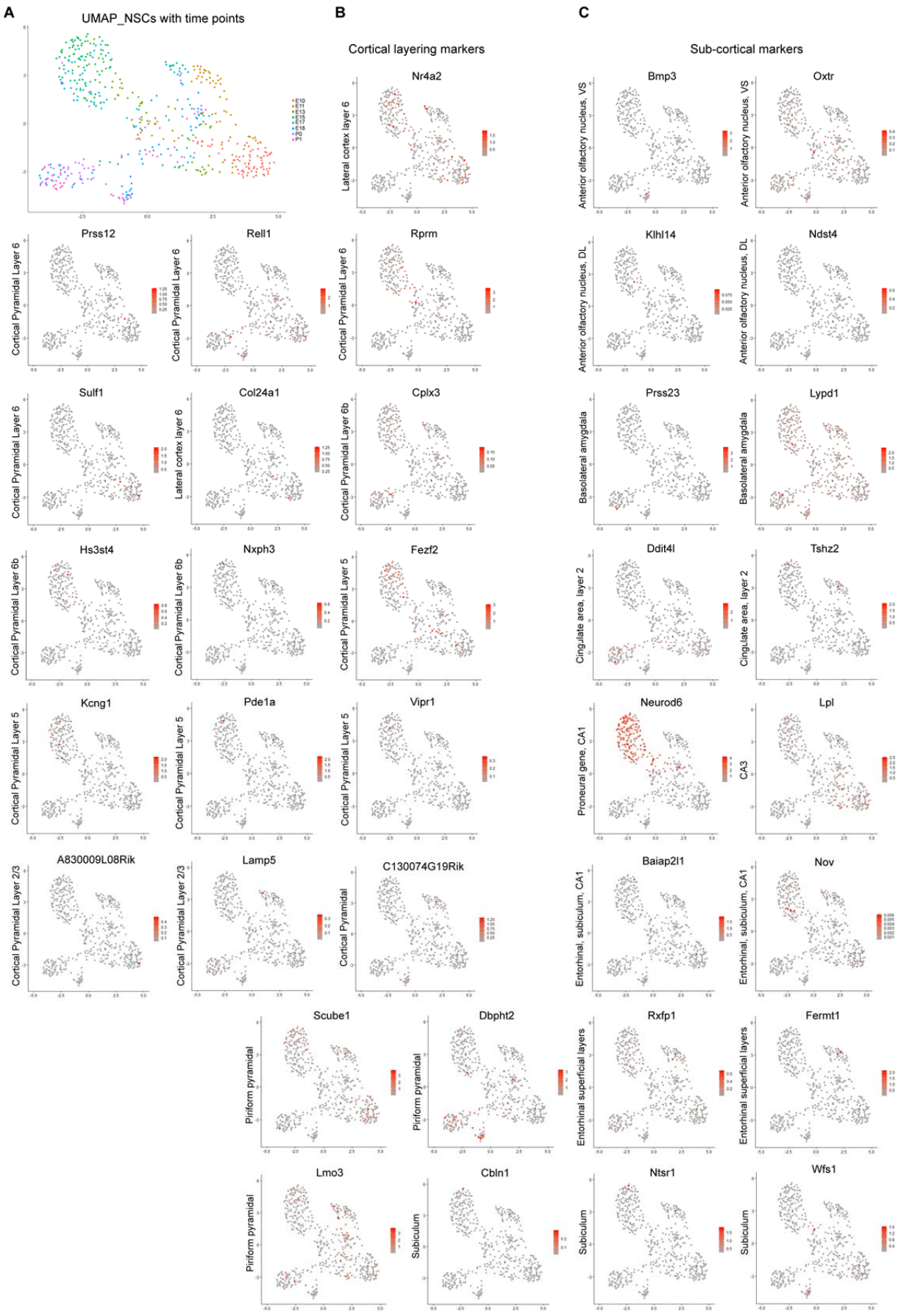

C1 data compared to Linnarsson Adolescent cell types
A
UMAP_NSCs with time points
E10
E11
E13
E15
E17
E18
P0
P1
B
Cortical layering markers
Nr4a2
Lateral cortex layer 6
Prss12
Cortical Pyramidal Layer 6
Rell1
Cortical Pyramidal Layer 6
Rprm
Cortical Pyramidal Layer 6
Sulf1
Cortical Pyramidal Layer 6
Col24a1
Lateral cortex layer 6
Cplx3
Cortical Pyramidal Layer 6b
Hs3st4
Cortical Pyramidal Layer 6b
Nxph3
Cortical Pyramidal Layer 6b
Fezf2
Cortical Pyramidal Layer 5
Kcng1
Cortical Pyramidal Layer 5
Pde1a
Cortical Pyramidal Layer 5
Vipr1
Cortical Pyramidal Layer 5
A830009L08Rik
Cortical Pyramidal Layer 2/3
Lamp5
Cortical Pyramidal Layer 2/3
C130074G19Rik
Cortical Pyramidal
C
Sub-cortical markers
Bmp3
Anterior olfactory nucleus, VS
Oxtr
Anterior olfactory nucleus, VS
Klhl14
Anterior olfactory nucleus, DL
Ndst4
Anterior olfactory nucleus, DL
Prss23
Basolateral amygdala
Lypd1
Basolateral amygdala
Ddit4l
Cingulate area, layer 2
Tshz2
Cingulate area, layer 2
Neurod6
Proneural gene, CA1
Lpl
CA3
Baiap2l1
Entorhinal, subiculum, CA1
Nov
Entorhinal, subiculum, CA1
Scube1
Piriform pyramidal
Dbpht2
Piriform pyramidal
Rxfp1
Entorhinal superficial layers
Fermt1
Entorhinal superficial layers
Lmo3
Piriform pyramidal
Cbln1
Subiculum
Ntsr1
Subiculum
Wfs1
Subiculum

Figure: C1 data compared to (Zeisel et al., 2018) Linnarsson Adolescent cell types.

(A) UMAP clustering for NSCs showing the entire developmental time course.

(B) Feature plots for C1 expression data of cortical layers markers (listed below), taken from (Zeisel et al., 2018).

(C) Feature plots for C1 expression data of sub-cortical markers (listed below), shown below taken from (Zeisel et al., 2018).

Symbol	Index	Description	Markers	Cells	Region	Likely location
TEGLU1	1	Excitatory neurons, cerebral cortex	*Myl4 Cpne4*	402	Cortex	Cingulate/Retrosplenial area, layer 6
TEGLU3	2	Excitatory neurons, cerebral cortex	*Rprm Rell1*	2734	Cortex	Cortex pyramidal layer 6
TEGLU2	3	Excitatory neurons, cerebral cortex	*Cplx3 Nxph3 Hs3st4*	274	Cortex	Cortex pyramidal layer 6b
TEGLU20	4	Excitatory neurons, cerebral cortex	*Sulf1 Prss12*	55	Cortex	Cortex pyramidal layer 6
TEGLU11	5	Excitatory neurons, cerebral cortex	*Igfbp6 C130074G19Rik*	1464	Cortex	Cortical pyramidal (poor markers)
TEGLU12	6	Excitatory neurons, cerebral cortex	*Nr4a2 Col24a1 Oprk1*	386	Cortex	Lateral cortex layer 6: gustatory, barrel field, auditory
TEGLU10	7	Excitatory neurons, cerebral cortex	*Hs3st2 Vipr1 Pde1a*	866	Cortex	Cortical pyramidal layer 5
TEGLU9	8	Excitatory neurons, cerebral cortex	*Dkkl1 Galntl6*	202	Cortex	Cingulate/Retrosplenial area, layer 5
TEGLU8	9	Excitatory neurons, cerebral cortex	*Krt12 Tcap*	3071	Cortex	Cortical pyramidal layer 4
TEGLU7	10	Excitatory neurons, cerebral cortex	*A830009L08Rik Gm12371 Lamp5*	2435	Cortex	Cortical pyramidal layer 2/3
TEGLU6	11	Excitatory neurons, cerebral cortex	*Tshz2 Ddit4l*	598	Cortex	Cingulate/Retrosplenial area, layer 2
TEGLU13	12	Excitatory neurons, cerebral cortex	*Wfs1 RP24-134N2.1 Vwc2l*	127	Cortex	Subiculum
TEGLU14	13	Excitatory neurons, cerebral cortex	*Cbln1 Ntsr1*	35	Cortex	Subiculum
TEGLU5	14	Excitatory neurons, cerebral cortex	*Rxfp1 Fermt1*	126	Cortex	Entorhinal superficial layers
TEGLU16	15	Excitatory neurons, cerebral cortex	*RP23-231J2.1 Dbpht2 Cdkl4*	61	Cortex	Piriform pyramidal
TEGLU15	16	Excitatory neurons, cerebral cortex	*Scube1 Tekt5*	186	Cortex	Piriform pyramidal
TEGLU17	17	Excitatory neurons, cerebral cortex	*Trim54 Slc30a3 Lmo3*	784	Cortex	Piriform pyramidal
TEGLU18	18	Excitatory neurons, cerebral cortex	*Abi3bp 4930426D05Rik Ndst4 Klhl14 Rgs14*	657	Cortex	Anterior olfactory nucleus, deep layer
TEGLU19	19	Excitatory neurons, cerebral cortex	*Oxtr Bmp3*	89	Cortex	Anterior olfactory nucleus and ventral striatum
TEGLU22	20	Excitatory neurons, amygdala	*Lypd1 Prss23 Kcng1*	958	Amygdala	Basolateral amygdala
TEGLU21	21	Excitatory neurons, hippocampus CA1	*Nov 4833422C13Rik Baiap2l1*	104	Hippocampus	CA1 (posterior),subiculum,entorhinal
TEGLU4	22	Excitatory neurons, cerebral cortex	*Fezf2 RP24-134N2.1 Kcng1*	664	Cortex	Cortex pyramidal layer 5, Cingulate/Retrosplenial area (superficial and deep)
TEGLU24	23	Excitatory neurons, hippocampus CA1	*Spink8 Lefty1 Neurod6*	2037	Hippocampus	CA1
TEGLU23	24	Excitatory neurons, hippocampus CA3	*Lpl Gm26644*	484	Hippocampus	CA3

References:

Balwierz, P.J., Pachkov, M., Arnold, P., Gruber, A.J., Zavolan, M., and van Nimwegen, E. (2014). ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. Genome Res *24*, 869-884.

Basak, O., and Taylor, V. (2007). Identification of self-replicating multipotent progenitors in the embryonic nervous system by high Notch activity and Hes5 expression. Eur J Neurosci *25*, 1006-1022.

Brennecke, P., Anders, S., Kim, J.K., Kolodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C.*, et al.* (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat Methods *10*, 1093-1095.

Brown, N., Alkhayer, K., Clements, R., Singhal, N., Gregory, R., Azzam, S., Li, S., Freeman, E., and McDonough, J. (2016). Neuronal Hemoglobin Expression and Its Relevance to Multiple Sclerosis Neuropathology. J Mol Neurosci *59*, 1-17.

Consortium, F., Suzuki, H., Forrest, A.R., van Nimwegen, E., Daub, C.O., Balwierz, P.J., Irvine, K.M., Lassmann, T., Ravasi, T., Hasegawa, Y.*, et al.* (2009). The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. Nat Genet *41*, 553-562.

Delgado, R.N., Allen, D.E., Keefe, M.G., Mancia Leon, W.R., Ziffra, R.S., Crouch, E.E., Alvarez-Buylla, A., and Nowakowski, T.J. (2022). Individual human cortical progenitors can produce excitatory and inhibitory neurons. Nature *601*, 397-403.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187-1201.

Knuckles, P., Vogt, M.A., Lugert, S., Milo, M., Chong, M.M., Hautbergue, G.M., Wilson, S.A., Littman, D.R., and Taylor, V. (2012). Drosha regulates neurogenesis by controlling neurogenin 2 expression independent of microRNAs. Nat Neurosci *15*, 962-969.

La Manno, G., Siletti, K., Furlan, A., Gyllborg, D., Vinsland, E., Mossi Albiach, A., Mattsson Langseth, C., Khven, I., Lederer, A.R., Dratva, L.M.*, et al.* (2021). Molecular architecture of the developing mouse brain. Nature *596*, 92-96.
Luecken, M.D., and Theis, F.J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol *15*, e8746.
Lugert, S., Vogt, M., Tchorz, J.S., Muller, M., Giachino, C., and Taylor, V. (2012). Homeostatic neurogenesis in the adult hippocampus does not involve amplification of Ascl1(high) intermediate progenitors. Nat Commun *3*, 670.
Nyfeler, Y., Kirch, R.D., Mantei, N., Leone, D.P., Radtke, F., Suter, U., and Taylor, V. (2005). Jagged1 signals in the postnatal subventricular zone are required for neural stem cell self-renewal. EMBO J *24*, 3504-3515.
Yoon, K.J., Ringeling, F.R., Vissers, C., Jacob, F., Pokrass, M., Jimenez-Cyrus, D., Su, Y., Kim, N.S., Zhu, Y., Zheng, L.*, et al.* (2017). Temporal Control of Mammalian Cortical Neurogenesis by m(6)A Methylation. Cell *171*, 877-889 e817.
Zeisel, A., Hochgerner, H., Lonnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Haring, M., Braun, E., Borm, L.E., La Manno, G.*, et al.* (2018). Molecular Architecture of the Mouse Nervous System. Cell *174*, 999-1014 e1022.
Zhang, C., Tu, H.L., Jia, G., Mukhtar, T., Taylor, V., Rzhetsky, A., and Tay, S. (2019). Ultra-multiplexed analysis of single-cell dynamics reveals logic rules in differentiation. Sci Adv *5*, eaav7959.

1st Revision - Editorial Decision **30th Aug 2022**

Dear Tanzila,

Congratulations on a great revision! Overall, the referees have been positive, however Referee #2 has asked that you include a figure legend for Fig 5 C-E, and we kindly ask that you include this in a revised version.

When you submit your revised version, please also take care of the following editorial items and add this also to your point-by-point response:

1. Up to five EV figures (from the supplemental file) can be uploaded as individual figure files and these should be labeled "Appendix Figure S1" etc. Please compile the remaining five figures into one PDF labeled "Appendix" and the legends added to the file as well. For more information, please see: https://www.embopress.org/page/journal/14602075/authorguide

2. Please confirm that there are four corresponding authors for this manuscript.

3. Please ensure that ORCID IDs are submitted for Iber and Van Nimwegen.

4. Please move the Data Availability section to the end of the Materials and Methods section

5. Please remove the Author Contribution section from the manuscript and rather use the free text boxes in EJP to include this information

6. Please review our new policy on conflict of interests on the EMBO author guide website and update the title of this section to: Disclosure and competing interests statement.

7. For the 7 EV tables, please rename as "Dataset EV1" etc., remove legends from the manuscript and add to the files.

8. We encourage the publication of source data; particularly for electrophoretic gels and blots and graphs, with the aim of making primary data more accessible and transparent to the reader. It would be great if you could provide me with a PDF file per figure that contains an Excel spreadsheet with the original data used to generate the graphs. The PDF files should be labeled with the appropriate figure/panel number. The PDF files will be published online with the article as supplementary "Source Data" files.

9. We include a synopsis of the paper (see http://emboj.embopress.org/). Please provide me with a general summary statement and 3-5 bullet points that capture the key findings of the paper.

10. We also need a summary figure for the synopsis. The size should be 550 wide by 200-440 high (pixels). You can also use something from the figures if that is easier.

11. Figure 4B and Table EV3 are not referred to in the manuscript, please add these in the correct order.

12. Please remove the Resource Table from the manuscript and upload as a separate file.

13. Please remove "Contact for resource sharing" section.

14. Our publisher has also done their pre-publication check on your manuscript. When you log into the manuscript submission system, you will see the file "Data Edited Manuscript file". Please take a look at the word file and the comments regarding the figure legends and respond to the issues.

Thank you for the opportunity to consider your work for publication, I look forward to your revision. If you have any questions, please feel free to reach out.

Kind regards,

Kelly

Kelly M Anderson, PhD
Editor
The EMBO Journal
k.anderson@embojournal.org

Further information is available in our Guide For Authors: https://www.embopress.org/page/journal/14602075/authorguide

Use the link below to submit your revision:

Link Not Available

Referee #2:

The revised manuscript by Mukhtar and colleagues has uncovered gene expression profiles and cell type diversities of the cortical neural lineages. The authors have added substantial analyzing data which addresses most of the concerns brought up by the reviewers. They conducted CCA integration analysis with publicly available scRNA-seq data of radial glia cells and observed similar expression patterns of signature genes including hemoglobin subunits HBB and HBA. They have added a description referring to the regulatory network prediction analysis by ISMARA and reorganized Fig3 accordingly. They constructed a trajectory from NSCs to NBNs and revealed potential associations between NSCs, BPs, and NBNs. Overall, the results are informative and the data will be a valuable source for studying the developing mouse cerebral cortex.
I only have a few minor concerns with the revised manuscript as detailed below:

1. In the revised Fig5 C-E, it would be helpful to add a figure legend to show what each color represents.

Referee #3:

In summary this paper provides a valuable resource for investigating molecular mechanisms of cortex development and has been further improved in this revision.
The revised manuscript includes new comparison to independent 'Linnarsson' embryonic mouse cortex data strengthening it's general applicability as a resource and provides interesting speculation about underlying mechanisms for future research.

The authors performed the requested editorial changes.

2nd Revision - Editorial Decision **26th Sep 2022**

Dear Tanzila,

Congratulations on an excellent manuscript, I am pleased to inform you that your manuscript has been accepted for publication in The EMBO Journal. Thank you for your comprehensive response to the referee concerns and for providing detailed source data. It has been a pleasure to work with you to get this to the acceptance stage.

There remain a few formatting issues to please attend to: Please add 5 keywords to the manuscript, add a table of contents to the appendix, and clarify which version of the DAS should be used (with or without NeuroStemX data).

I will begin the final checks on your manuscript before submitting to the publisher next week. Once at the publisher, it will take about 3 weeks for your manuscript to be published online. As a reminder, the entire review process, including referee concerns and your point-by-point response, will be available to readers.

I will be in touch throughout the final editorial process until publication. In the meantime, I hope you find time to celebrate!

Yours sincerely,

Kelly

Kelly M Anderson, PhD
Editor
The EMBO Journal
k.anderson@embojournal.org

EMBO Press Author Checklist

Corresponding Author Name: Verdon Taylor	
Journal Submitted to: The Embo Journal	
Manuscript Number: EMBOJ-2022-111132R	

USEFUL LINKS FOR COMPLETING THIS FORM

The EMBO Journal - Author Guidelines
EMBO Reports - Author Guidelines
Molecular Systems Biology - Author Guidelines
EMBO Molecular Medicine - Author Guidelines

Reporting Checklist for Life Science Articles (updated January 2022)

This checklist is adapted from Materials Design Analysis Reporting (MDAR) Checklist for Authors. MDAR establishes a minimum set of requirements in transparent reporting in the life sciences (see Statement of Task: 10.31222/osf.io/9sm4x). Please follow the journal's guidelines in preparing your manuscript.

Please note that a copy of this checklist will be published alongside your article.

Abridged guidelines for figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- ideally, figure panels should include only measurements that are directly comparable to each other and obtained with the same assay.
- plots include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if n<5, the individual data points from each experiment should be plotted. Any statistical test employed should be justified.
- Source Data should be included to report the data underlying figures according to the guidelines set out in the authorship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements.
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.
- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple χ2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Please complete ALL of the questions below.
Select "Not Applicable" only when the requested information is not relevant for your study.

Materials

Newly Created Materials	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
New materials and reagents need to be available; do any restrictions apply?	Yes	Materials and methods

Antibodies	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
For **antibodies** provide the following information: - Commercial antibodies: RRID (if possible) or supplier name, catalogue number and or/clone number - Non-commercial: RRID or citation	Yes	Materials and methods

DNA and RNA sequences	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Short novel DNA or RNA including primers, probes: provide the sequences.	Yes	Materials and methods

Cell materials	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Cell lines: Provide species information, strain. Provide accession number in repository **OR** supplier name, catalog number, clone number, and/**OR** RRID.	Not Applicable	
Primary cultures: Provide species, strain, sex of origin, genetic modification status.	Not Applicable	
Report if the cell lines were recently **authenticated** (e.g., by STR profiling) and tested for mycoplasma contamination.	Not Applicable	

Experimental animals	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Laboratory animals or Model organisms: Provide species, strain, sex, age, genetic modification status. Provide accession number in repository **OR** supplier name, catalog number, clone number, **OR** RRID.	Yes	Materials and methods
Animal observed in or captured from the field: Provide species, sex, and age where possible.	Not Applicable	
Please detail **housing and husbandry conditions**.	Yes	Materials and methods

Plants and microbes	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Plants: provide species and strain, ecotype and cultivar where relevant, unique accession number if available, and source (including location for collected wild specimens).	Not Applicable	
Microbes: provide species and strain, unique accession number if available, and source.	Not Applicable	

Human research participants	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If collected and within the bounds of privacy constraints report on age, sex and gender or ethnicity for all study participants.	Not Applicable	

Core facilities	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If your work benefited from core facilities, was their service mentioned in the acknowledgments section?	Yes	Acknowledgements

Design

Study protocol	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If study protocol has been **pre-registered, provide DOI in the manuscript**. For clinical trials, provide the trial registration number **OR** cite DOI.	Not Applicable	
Report the **clinical trial registration number** (at ClinicalTrials.gov or equivalent), where applicable.	Not Applicable	

Laboratory protocol	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Provide DOI OR other citation details if **external detailed step-by-step protocols** are available.	Yes	Materials and methods

Experimental study design and statistics	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Include a statement about **sample size** estimate even if no statistical methods were used.	Not Applicable	
Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. **randomization procedure**)? If yes, have they been described?	Not Applicable	
Include a statement about **blinding** even if no blinding was done.	Not Applicable	
Describe **inclusion/exclusion criteria** if samples or animals were excluded from the analysis. Were the criteria pre-established? If sample or data points were omitted from analysis, report if this was due to attrition or intentional exclusion and provide justification.	Not Applicable	
For every figure, are **statistical tests** justified as appropriate? Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it. Is there an estimate of variation within each group of data? Is the variance similar between the groups that are being statistically compared?	Yes	Materials and methods, Results

Sample definition and in-laboratory replication	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
In the figure legends: state number of times the experiment was **replicated** in laboratory.	Yes	Figure legends
In the figure legends: define whether data describe **technical or biological replicates**.	Yes	Figure legends

Ethics

Ethics	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Studies involving **human participants**: State details of **authority granting ethics approval** (IRB or equivalent committee(s), provide reference number for approval.	Not Applicable	
Studies involving **human participants**: Include a statement confirming that **informed consent** was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	Not Applicable	
Studies involving **human participants:** For publication of **patient photos**, include a statement confirming that consent to publish was obtained.	Not Applicable	
Studies involving experimental **animals**: State details of **authority granting ethics approval** (IRB or equivalent committee(s), provide reference number for approval. Include a statement of compliance with ethical regulations.	Yes	Materials and methods
Studies involving **specimen and field samples:** State if relevant **permits** obtained, provide details of authority approving study; if none were required, explain why.	Not Applicable	

Dual Use Research of Concern (DURC)	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Could your study fall under dual use research restrictions? Please check biosecurity documents and list of **select agents and toxins** (CDC): https://www.selectagents.gov/sat/list.htm	Not Applicable	
If you used a select agent, is the security level of the lab appropriate and reported in the manuscript?	Not Applicable	
If a study is subject to dual use research of concern regulations, is the name of the **authority granting approval and reference number** for the regulatory approval provided in the manuscript?	Not Applicable	

Reporting

The MDAR framework recommends adoption of discipline-specific guidelines, established and endorsed through community initiatives. Journals have their own policy about requiring specific guidelines and recommendations to complement MDAR.

Adherence to community standards	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
State if relevant guidelines or checklists (e.g., **ICMJE, MIBBI, ARRIVE, PRISMA**) have been followed or provided.	Not Applicable	
For **tumor marker prognostic studies,** we recommend that you follow the **REMARK** reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	Not Applicable	
For **phase II and III randomized controlled trials**, please refer to the **CONSORT** flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	Not Applicable	

Data Availability

Data availability	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Have **primary datasets** been deposited according to the journal's guidelines (see 'Data Deposition' section) and the respective accession numbers provided in the Data Availability Section?	Yes	Materials and methods
Were **human clinical and genomic datasets** deposited in a public access-controlled repository in accordance to ethical obligations to the patients and to the applicable consent agreement?	Not Applicable	
Are **computational models** that are central and integral to a study available without restrictions in a machine-readable form? Were the relevant accession numbers or links provided?	Yes	Materials and methods
If publicly available data were reused, provide the respective **data citations in the reference list**.	Yes	Materials and methods