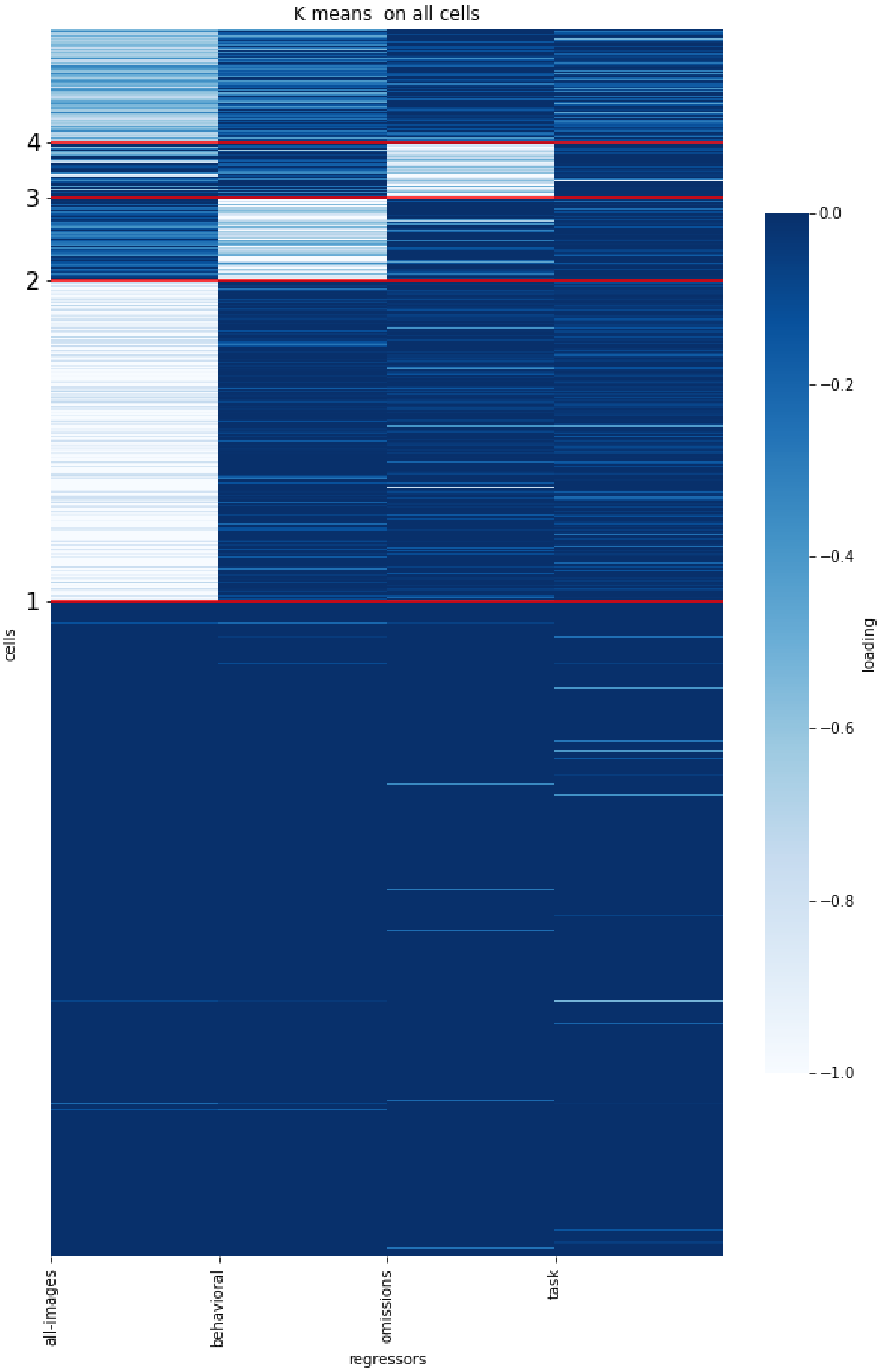
Visual Behavior - Cluster Distributions

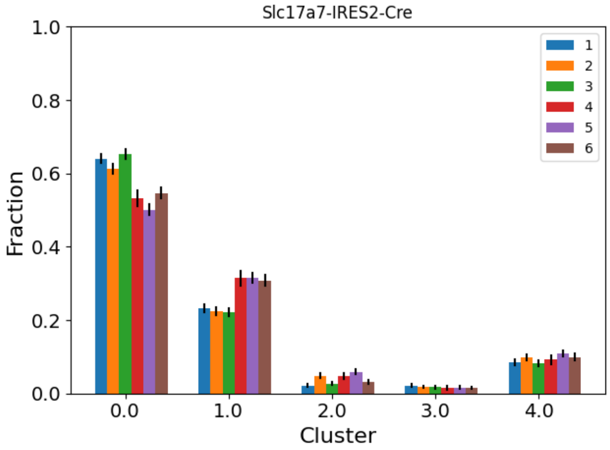
We would like to develop a clustering analysis of cells both within each session, and across sessions. However, we have a few issues that need resolution. Not all cells are tracked across all 6 sessions, and we need to determine whether those cells that are tracked are reasonable subset to examine. Second, we want to ask whether individual cells change their coding properties, or whether population changes across sessions are driven by new cells becoming active.

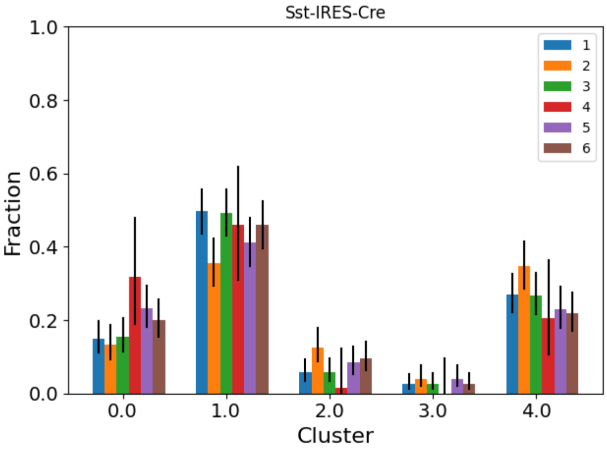
To answer these questions I am going to use the results from a simple k-means clustering with k set to 5 with all cells across all sessions clustered jointly. Here is a visualization of the clusters across cells.

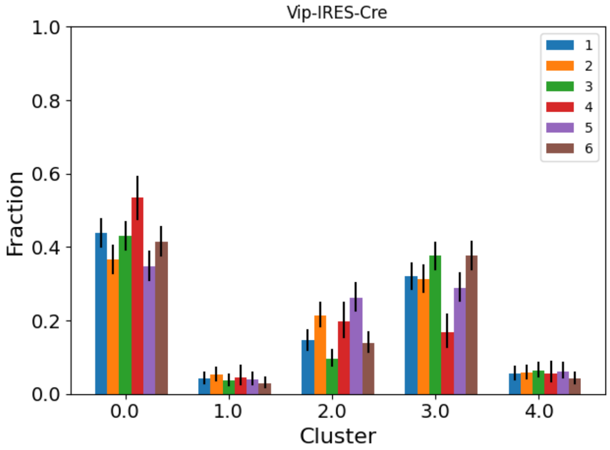


Do we see similar distribution of clusters across sessions?

To answer this question, we can plot the fraction of cells in each cluster across each session. If a cell has multiple retakes of a given session, I picked one at random. The black lines are 95% Multinomial proportion confidence intervals.

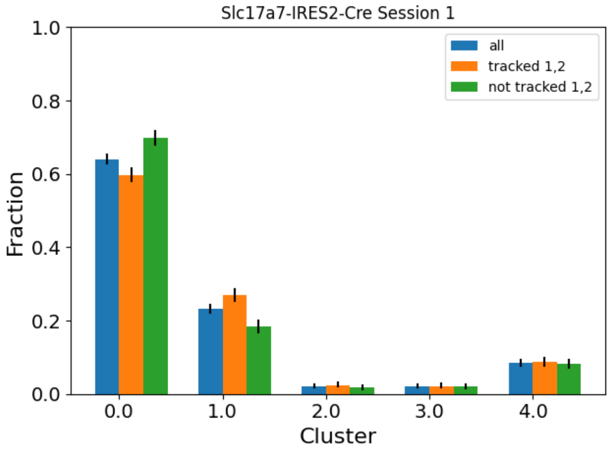


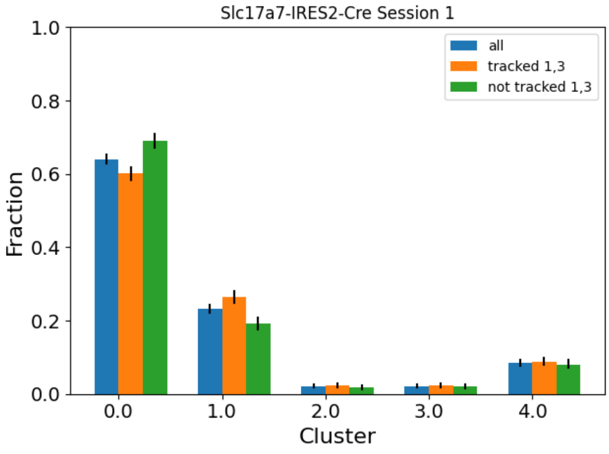


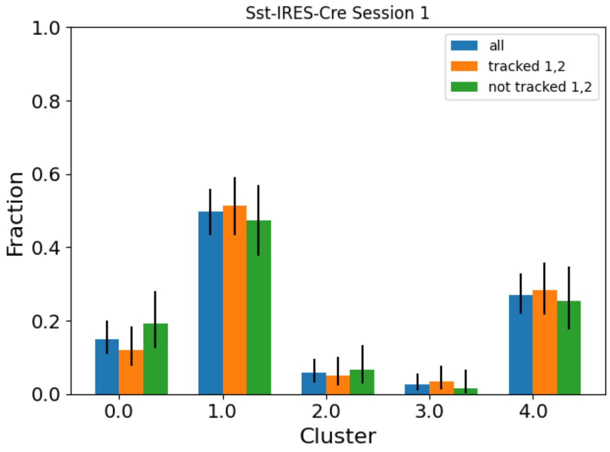


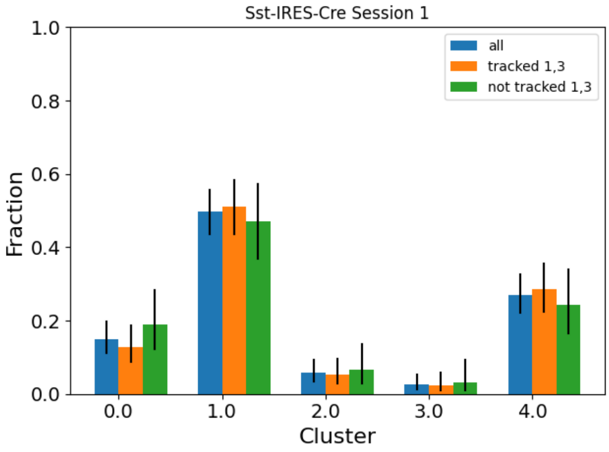
Are cells tracked across sessions different from non-tracked cells?

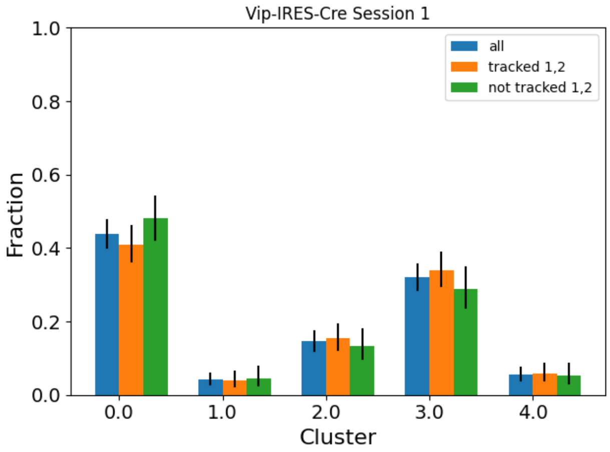
To answer this question I can plot the fraction of cells in each cluster for different subsets of cells. Here we can look at cells that are tracked from session 1 to the other familiar sessions

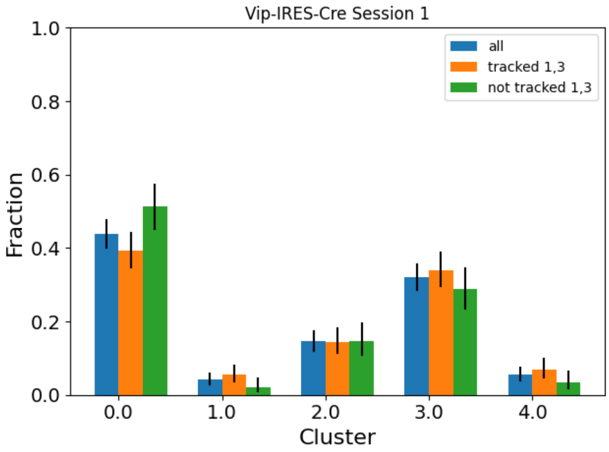




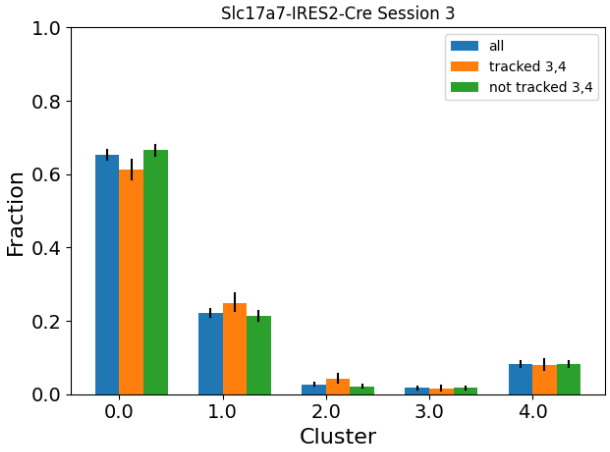


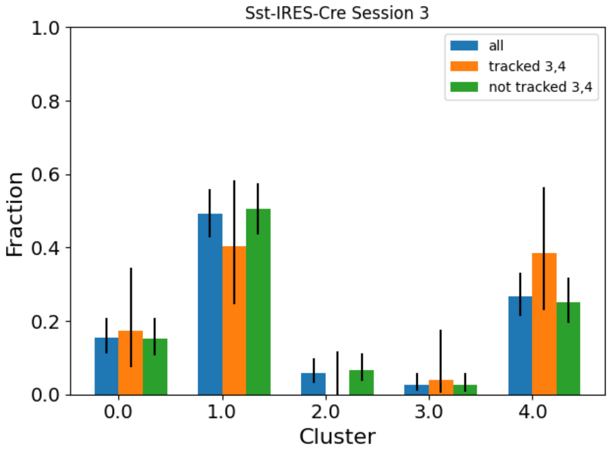


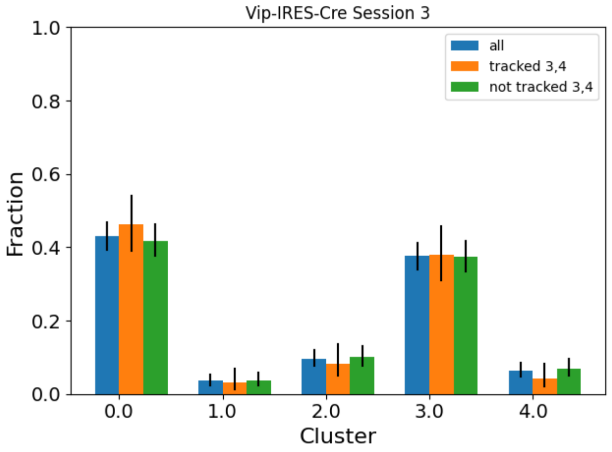




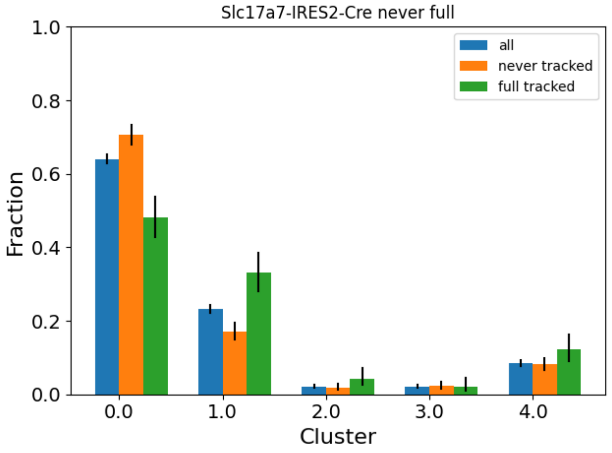
Now we can look across familar to novel

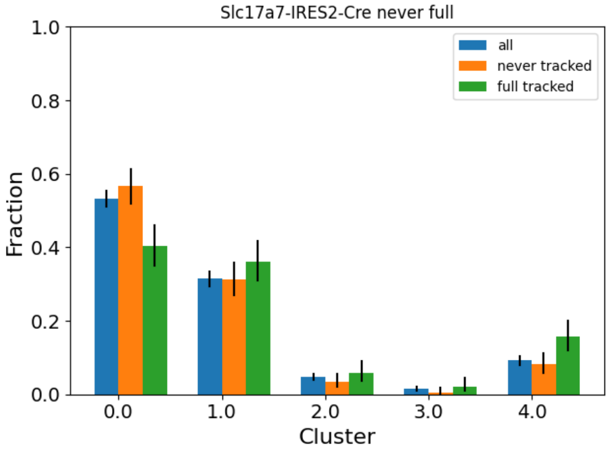


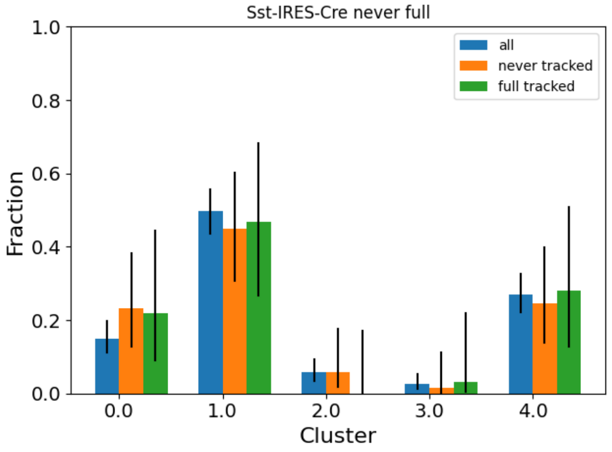


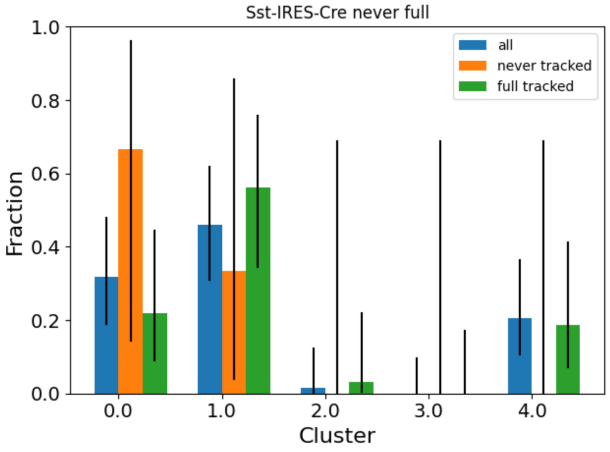


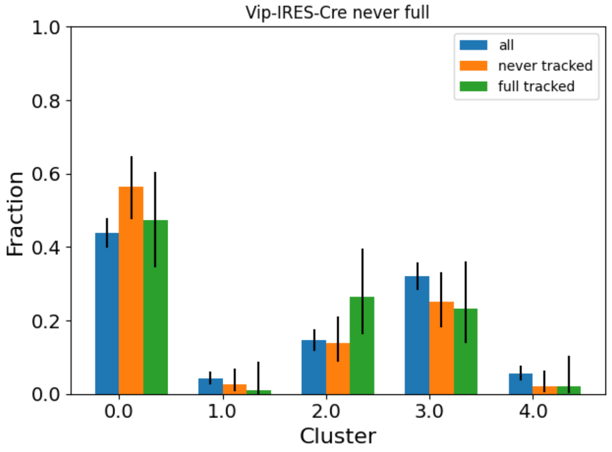
For a larger effect we can compare between cells that are tracked across all 6 sessions, and those that are tracked for just one session. Left column is session 1, right column is session 4

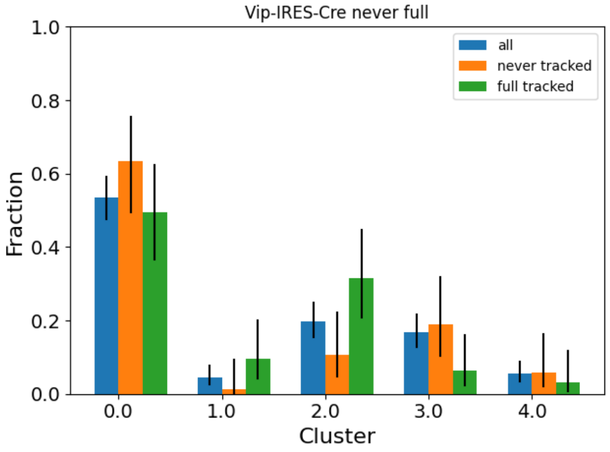




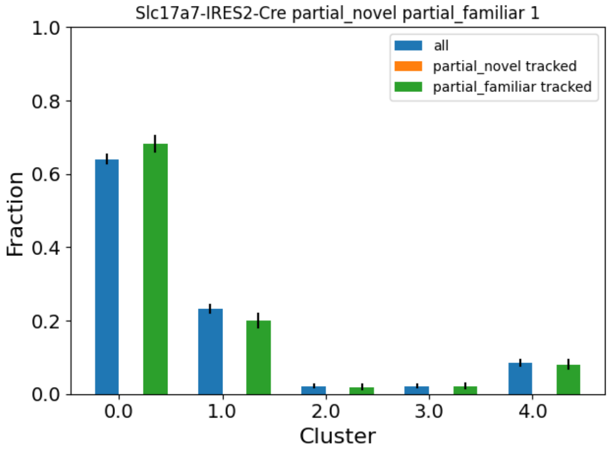


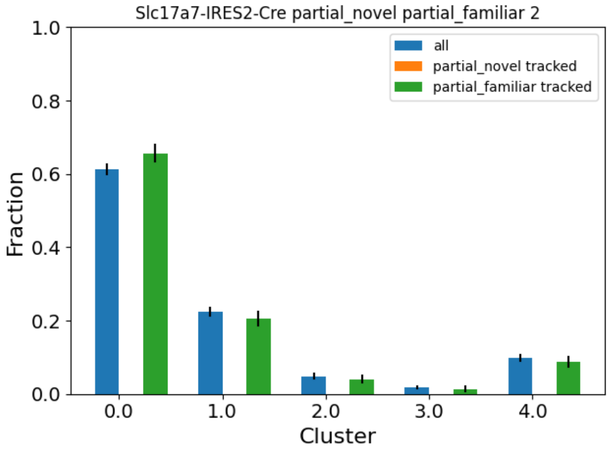


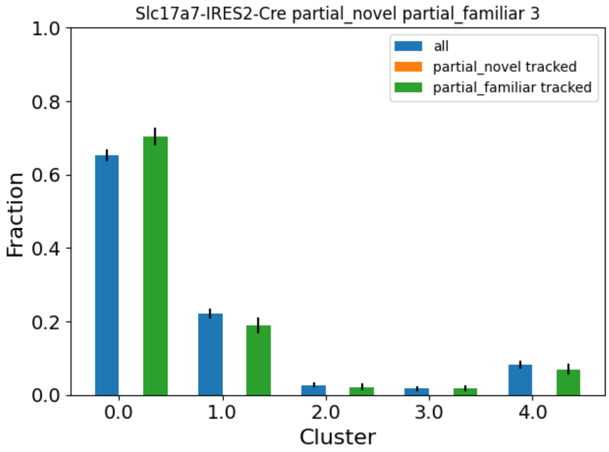


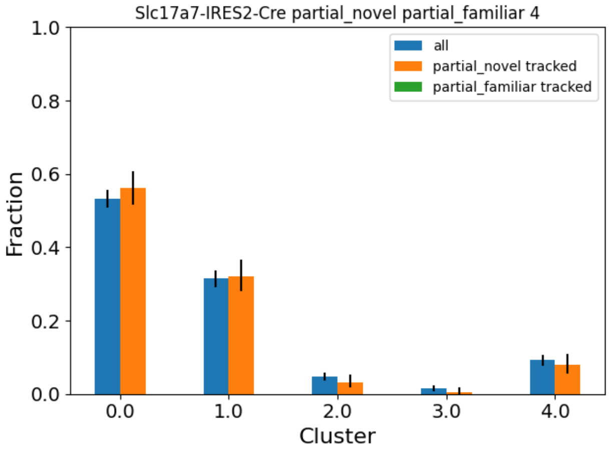


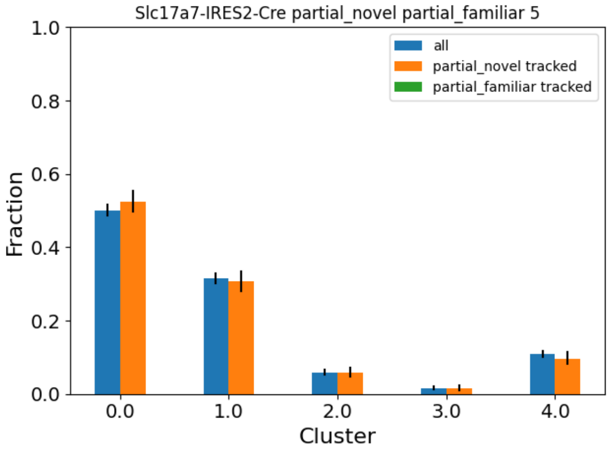
Marina suggested looking at cells that are novel only or familiar only. I defined “full_novel” as cells in all novel sessions and no familiar sessions, and “partial_novel” as cells in some novel sessions and no familiar sessions.

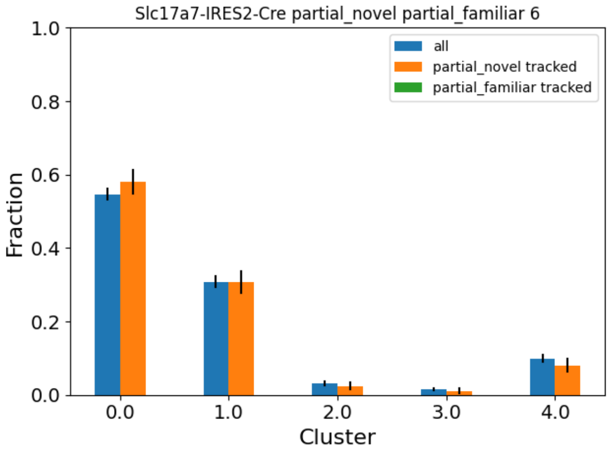


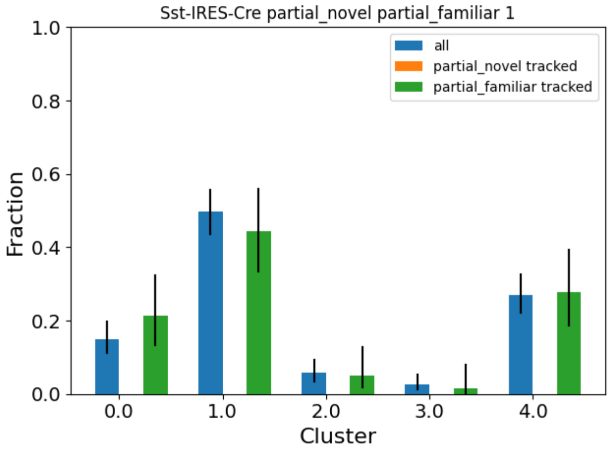


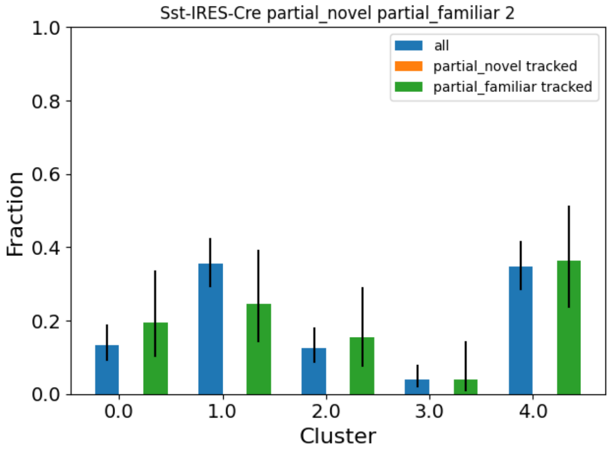


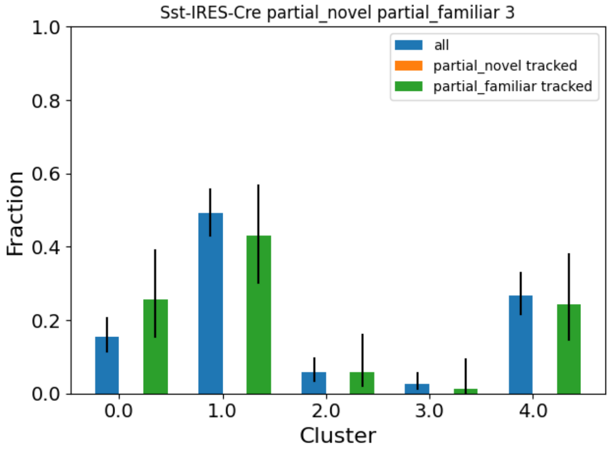


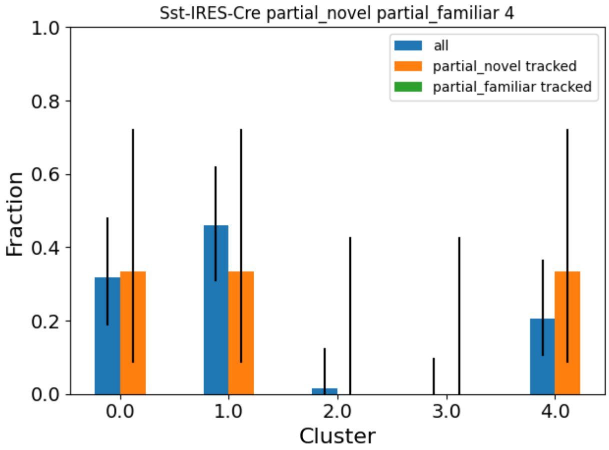


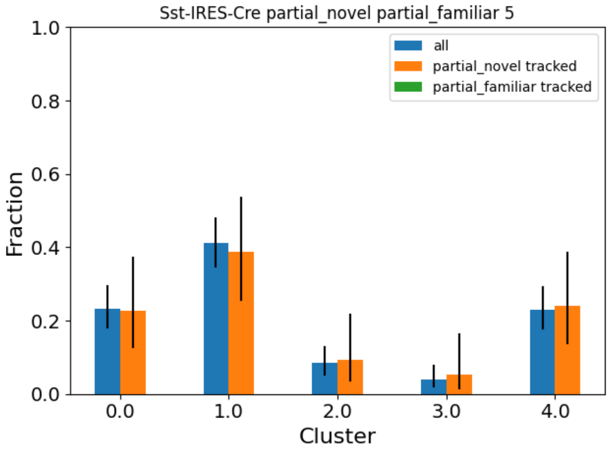


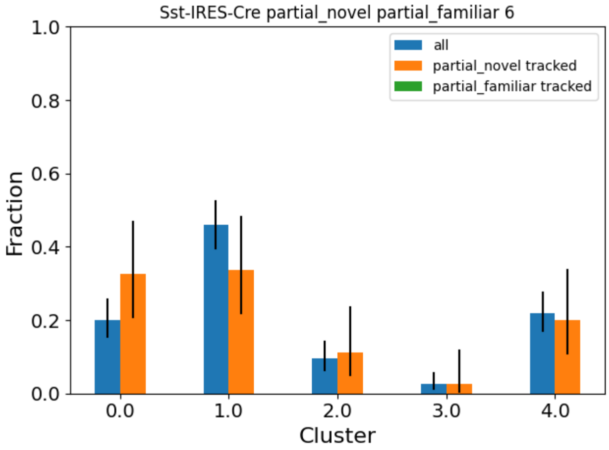

Sst-IRES-Cre partial_novel partial_familiar 1
all
partial_novel tracked
partial_familiar tracked
Fraction
0.0
0.2
0.4
0.6
0.8
1.0
0.0
1.0
2.0
3.0
4.0
Cluster

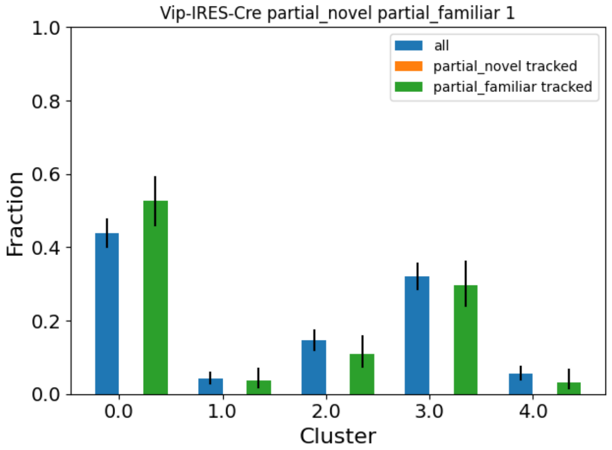

Sst-IRES-Cre partial_novel partial_familiar 2
all
partial_novel tracked
partial_familiar tracked
Fraction
0.0
0.2
0.4
0.6
0.8
1.0
0.0
1.0
2.0
3.0
4.0
Cluster


Sst-IRES-Cre partial_novel partial_familiar 3
all
partial_novel tracked
partial_familiar tracked
Fraction
0.0
0.2
0.4
0.6
0.8
1.0
0.0
1.0
2.0
3.0
4.0
Cluster

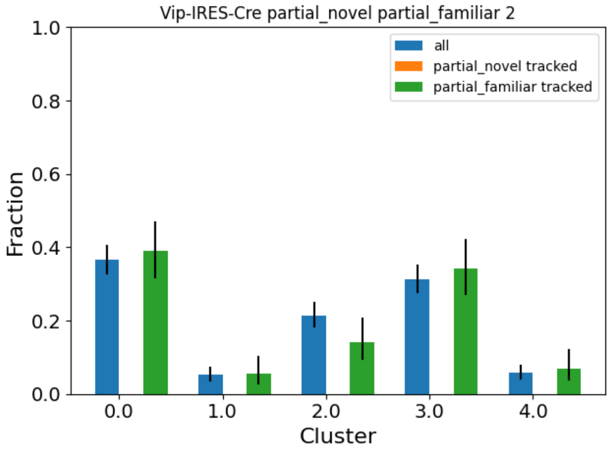

Sst-IRES-Cre partial_novel partial_familiar 4
all
partial_novel tracked
partial_familiar tracked
Fraction
0.0
0.2
0.4
0.6
0.8
1.0
0.0
1.0
2.0
3.0
4.0
Cluster


Sst-IRES-Cre partial_novel partial_familiar 5
all
partial_novel tracked
partial_familiar tracked
Fraction
0.0
0.2
0.4
0.6
0.8
1.0
0.0
1.0
2.0
3.0
4.0
Cluster

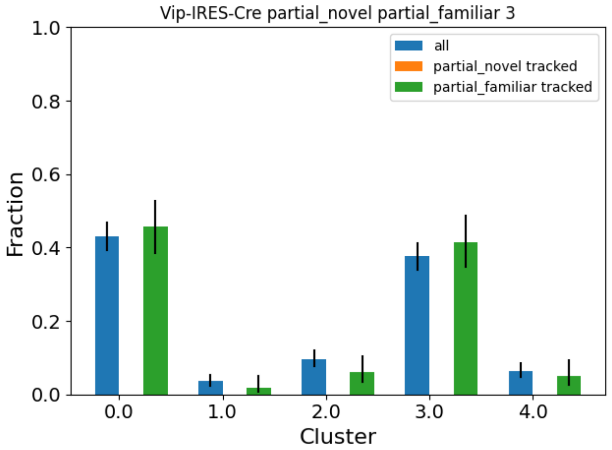

Sst-IRES-Cre partial_novel partial_familiar 6
all
partial_novel tracked
partial_familiar tracked
Fraction
0.0
0.2
0.4
0.6
0.8
1.0
0.0
1.0
2.0
3.0
4.0
Cluster



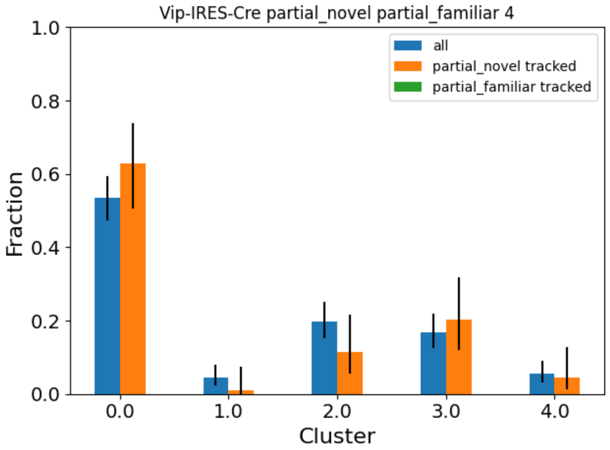
Vip-IRES-Cre partial_novel partial_familiar 1



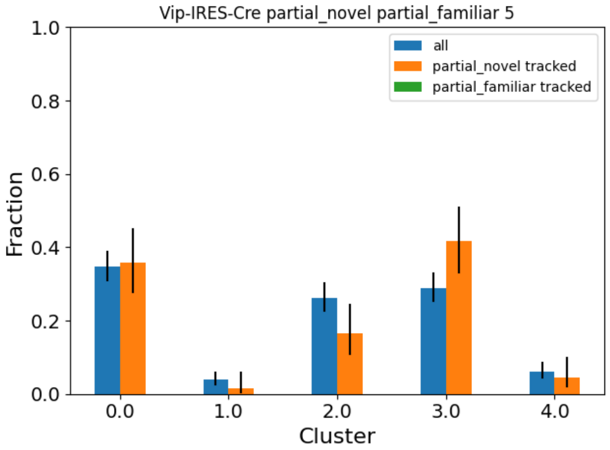
Vip-IRES-Cre partial_novel partial_familiar 2



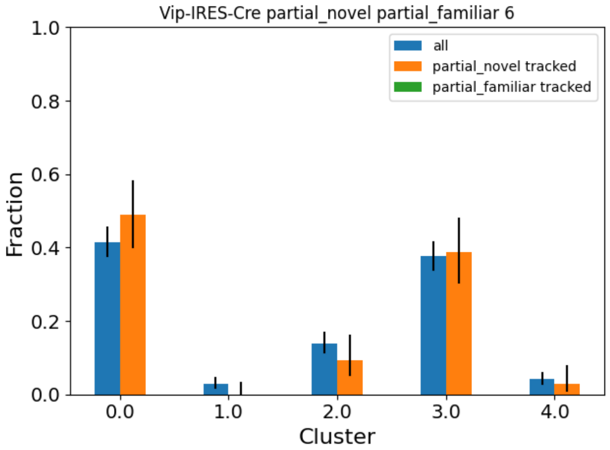
Vip-IRES-Cre partial_novel partial_familiar 3



Vip-IRES-Cre partial_novel partial_familiar 4



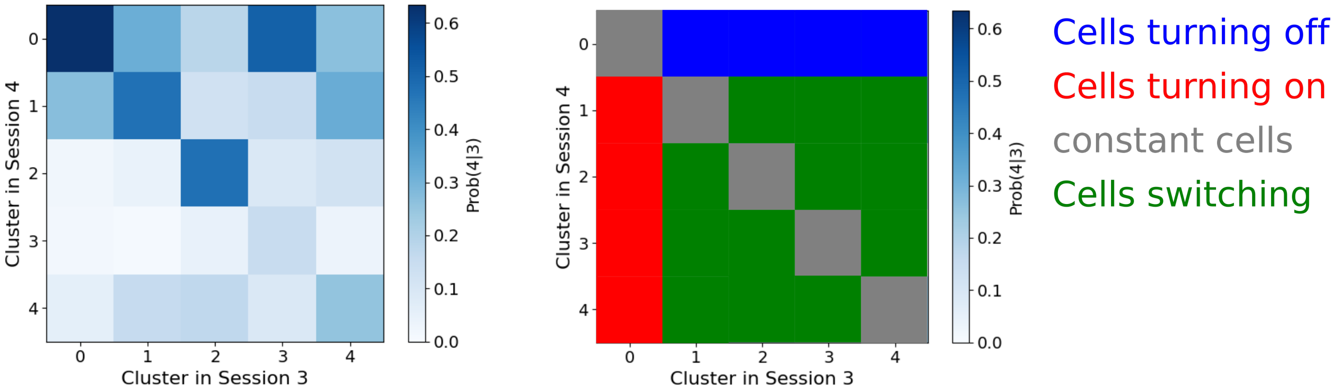
Vip-IRES-Cre partial_novel partial_familiar 5



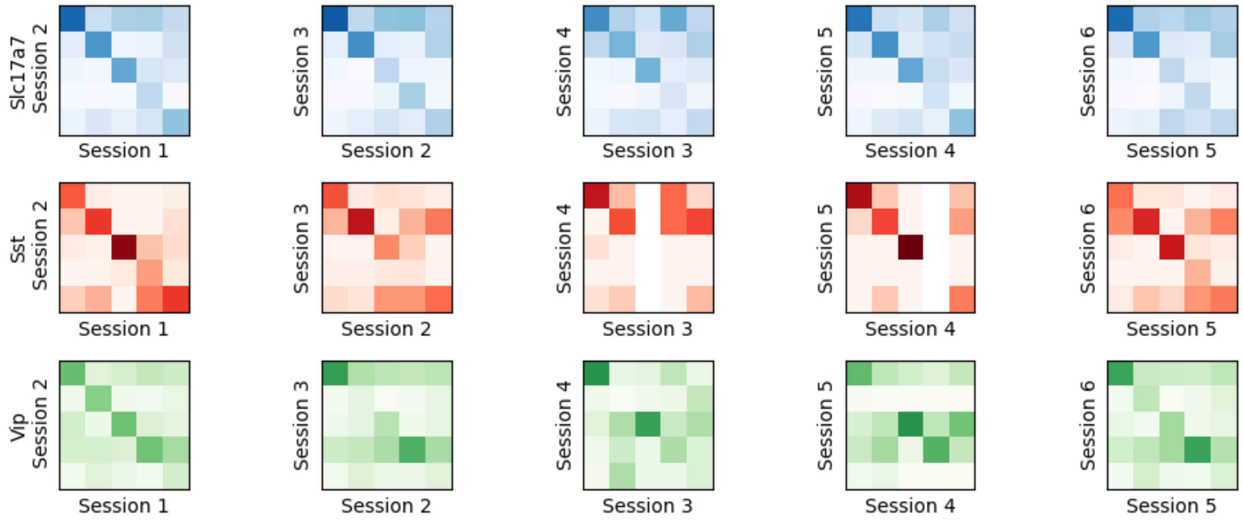
Vip-IRES-Cre partial_novel partial_familiar 6

Are changes in cluster distribution driven by new cells, or changes in coding from existing cells?

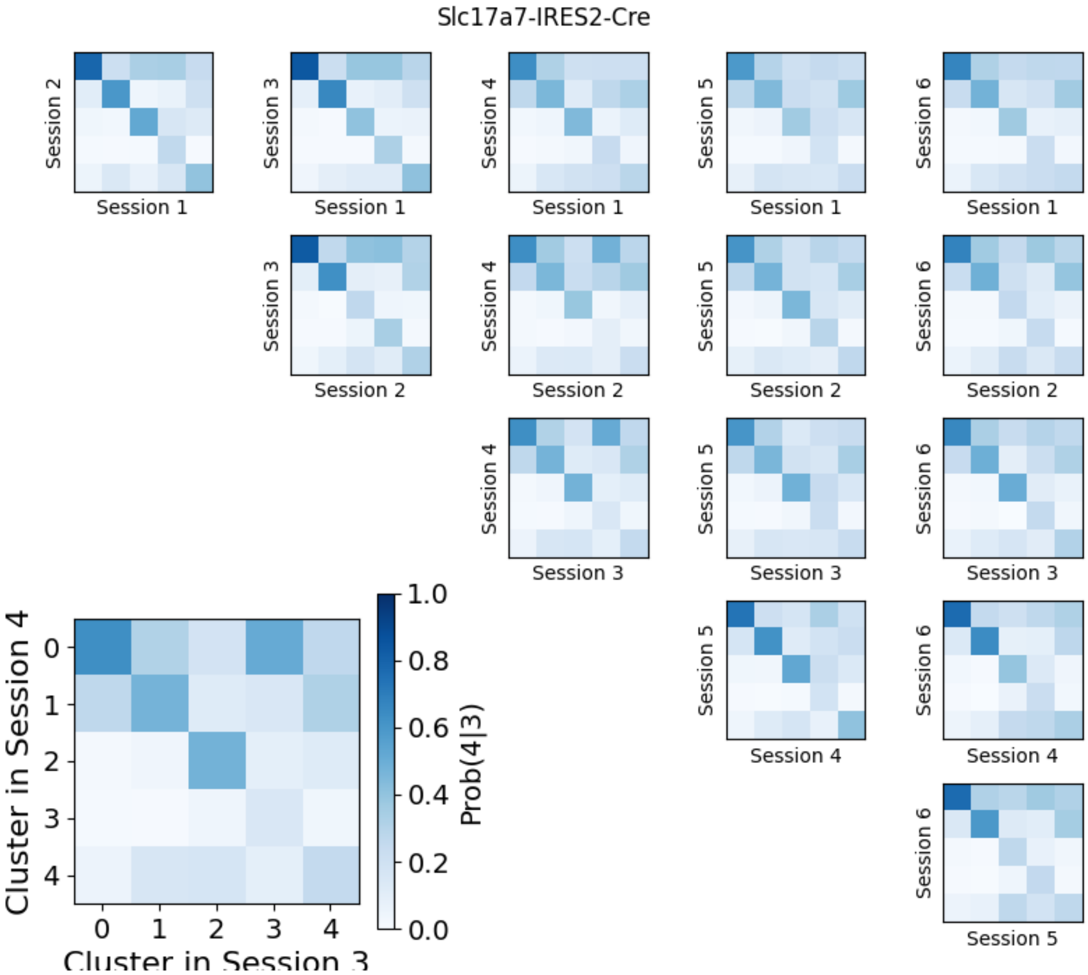
To examine this question I can compute the transition matrix, T, that shows the probability of moving from one cluster to a new cluster in the next session. Here each column is normalized to sum to 1, so this is a transition matrix such that if x_i is the distribution of clusters on session i then $x_{i+1} = Tx_i$. Here is an example for excitatory cells from session 3 to session 4. Note that only cells that are tracked from 3 to 4 go into this matrix, and so it could be missing cells that were not active in session 3 and then were active in session 4. However we have a large fraction of cells in cluster 0, the non-active cluster, so this worry is partially resolved.

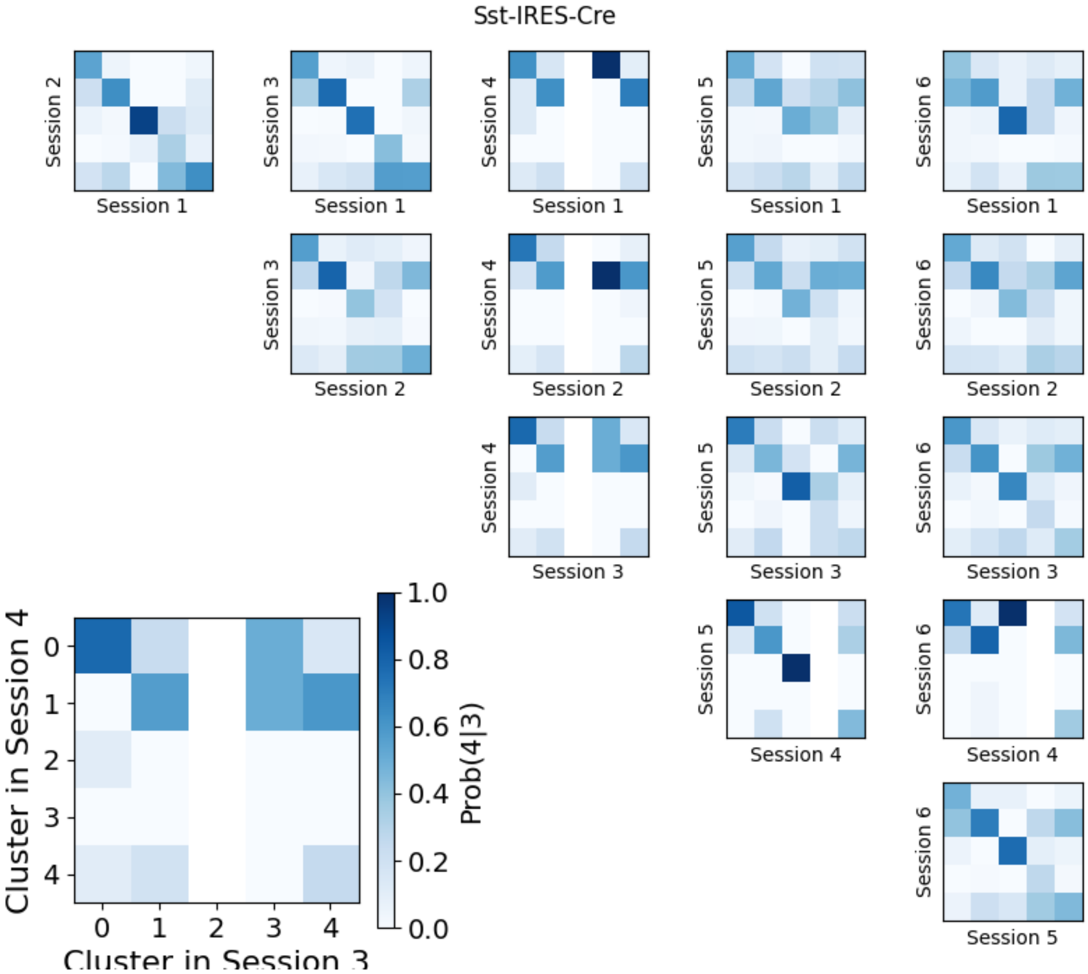


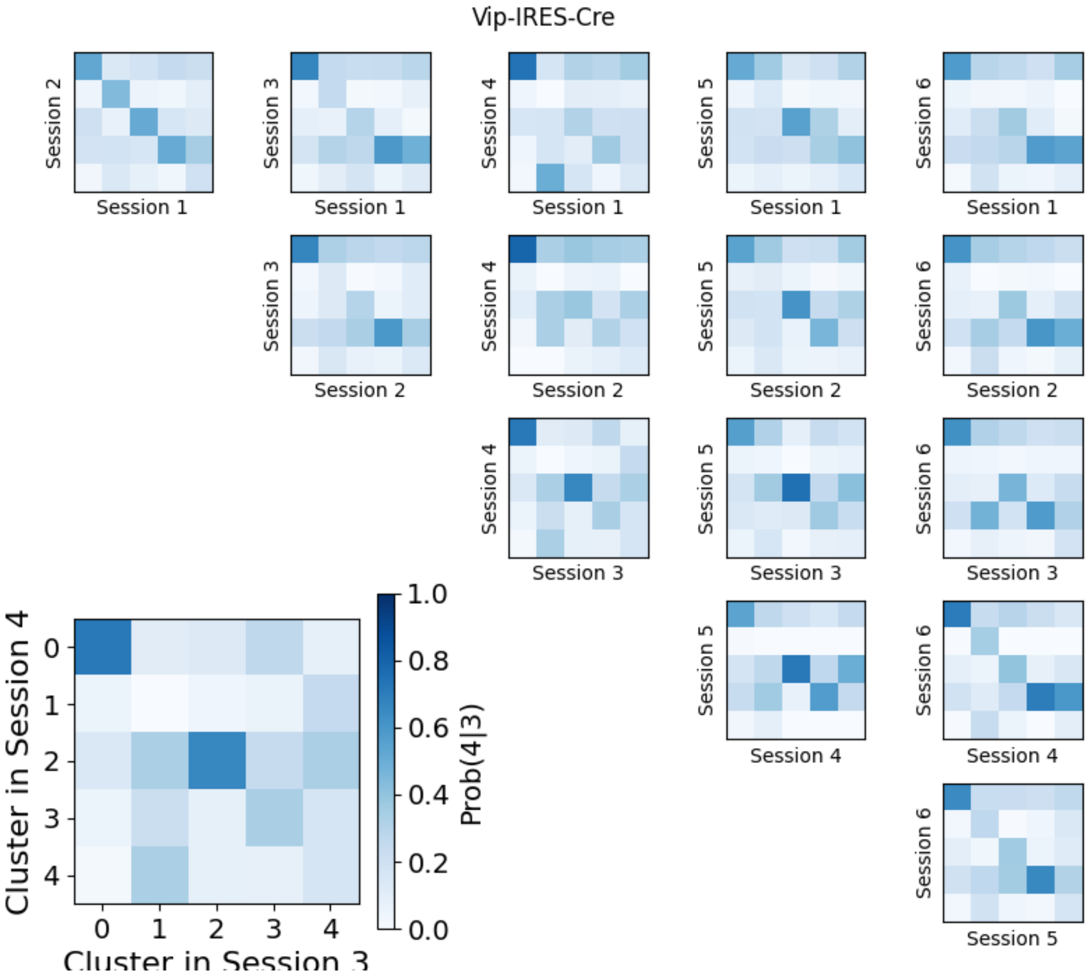
Note that any off-diagonal values in the submatrix for the transitions between cluster 1-4 are cells “switching” their coding. Where cells in the first column or first row are cells “turning off” or “turning on”. Next I am plotting the transition matrix between each stage progression for each cre-line



We can make these transition matrices for any pair of sessions, even those not “in-order”.


Slc17a7-IRES2-Cre
Session 2
Session 1
Session 3
Session 1
Session 4
Session 1
Session 5
Session 1
Session 6
Session 1
Session 3
Session 2
Session 4
Session 2
Session 5
Session 2
Session 6
Session 2
Session 4
Session 3
Session 5
Session 3
Session 6
Session 3
Session 5
Session 4
Session 6
Session 4
Session 6
Session 5
Cluster in Session 4
0
1
2
3
4
Cluster in Session 3
1.0
0.8
0.6
0.4
0.2
0.0
Prob(4|3)


Sst-IRES-Cre
Session 2
Session 1
Session 3
Session 1
Session 4
Session 1
Session 5
Session 1
Session 6
Session 1
Session 3
Session 2
Session 4
Session 2
Session 5
Session 2
Session 6
Session 2
Session 4
Session 3
Session 5
Session 3
Session 6
Session 3
Session 5
Session 4
Session 6
Session 4
Session 6
Session 5
Cluster in Session 4
0
1
2
3
4
Cluster in Session 3
1.0
0.8
0.6
0.4
0.2
0.0
Prob(4|3)


Vip-IRES-Cre
Session 2
Session 1
Session 3
Session 1
Session 4
Session 1
Session 5
Session 1
Session 6
Session 1
Session 3
Session 2
Session 4
Session 2
Session 5
Session 2
Session 6
Session 2
Session 4
Session 3
Session 5
Session 3
Session 6
Session 3
Session 5
Session 4
Session 6
Session 4
Session 6
Session 5
Cluster in Session 4
0
1
2
3
4
0
1
2
3
4
Cluster in Session 3
1.0
0.8
0.6
0.4
0.2
0.0
Prob(4|3)

Can we accurately predict cluster distributions using the transition matrices?

We can use the transition matrix to estimate the distribution of cells in a session given different populations of cells to estimate the transition matrix, and the input distribution. Specifically, using a subset a of all cells that were recorded in session i: $x_{a,i}$, and a transition matrix from a (potentially different) subset of cells b that were tracked across sessions i and $i+1$: $T_{b,i,i+1}$, we can estimate the distribution of cells in session $i+1$:

$$x_{c,i+1} = T_{b,i,i+1} x_{a,i} \tag{1}$$

Key analysis decisions:

- Choice of cells. We can select different subsets of cells for the input distribution, the output distribution to compare against, and for the transition matrix.

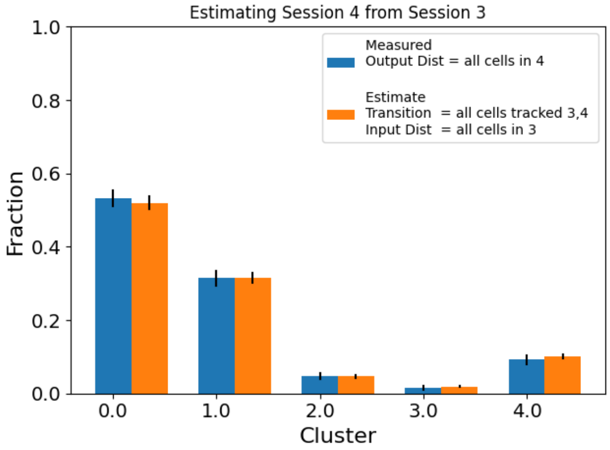
- Some transition matrices have columns of NaNs because no cells were recorded in that cluster in the input session. This problem is more pronounced when we subselect cells to make the transition matrix. I decided to clean the transition matricies to include an even probability distribution from that cluster to all other clusters. I think this is the most conservative estimate because it assumes the most entropy. A slightly less conservative approach might be to use the average column from the rest of the matrix.

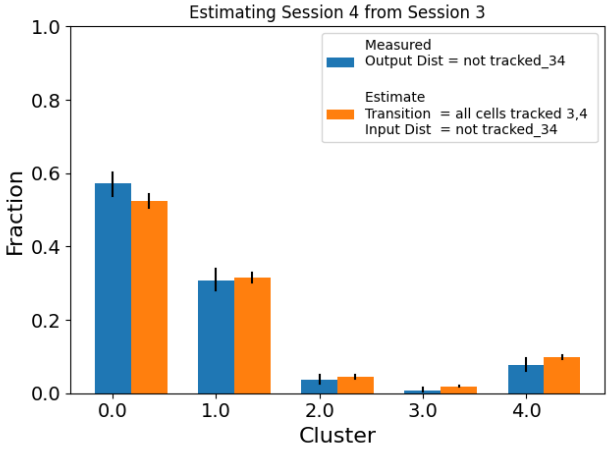
- Confidence intervals for our estimates. I decided to propagate the 95% confidence intervals from the input distribution forward. This approach captures our uncertainty about the input distribution, but does not capture our distribution about the transition matrix. The transition matrix uncertainty becomes worse the more we subselect cells. I think the better thing to do would be a bootstrap analysis where we sample from the transition cells to make the transition matrix and then get an output distribution uncertainty from that. However I think that is too much work for now. In general this means our estimate uncertainty distributions are too narrow.

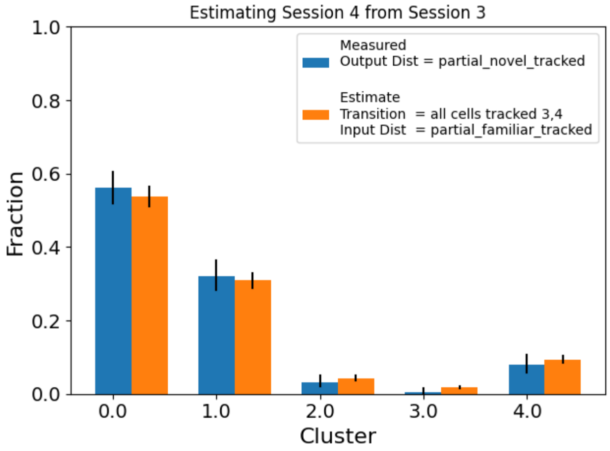
For each cre-line I present three different cell selection plots. First, the most inclusive, includes all cells in the input and output distributions, and makes the transition matrix with all cells that are tracked across both sessions. Next, I make the input and output distributions with all cells that are NOT used to make the transition matrix. Note that the input and output distributions can include the same cell. Finally, the most restrictive, the input and output distributions are made with non-overlapping populations that were not tracked across the transition. Therefore the three cell subsets are all non-overlapping.

Conclusion In general, we can accuratly predict the distribution of cells across the session 3 to session 4 transition. For excitatory cells, across all three cell selection approaches, our estimates are very close. For the inhibitory lines, low cell counts become an issue. For SST cells, the uncertainty intervals on both the measured and estimated distributions are very large. All estimates are within the confidence interval, but this isn't reassuring given the larger confidence intervals. For VIP cells, we see larger differences between the measured and estimated distributions. For the most inclusive and most restrictive analyses, the confidence intervals appear to always overlap. Remembering that we are not including our uncertainty on the transition matrix, I think we can conclude these distributions are the same. The middle road inclusion criteria does include a difference in the non-active cells (cluster 0). However we dont see any differences in any other clusters, which suggests that the cell inclusion criteria may be including more non-tracked, weakly active cells in the measured distribution.

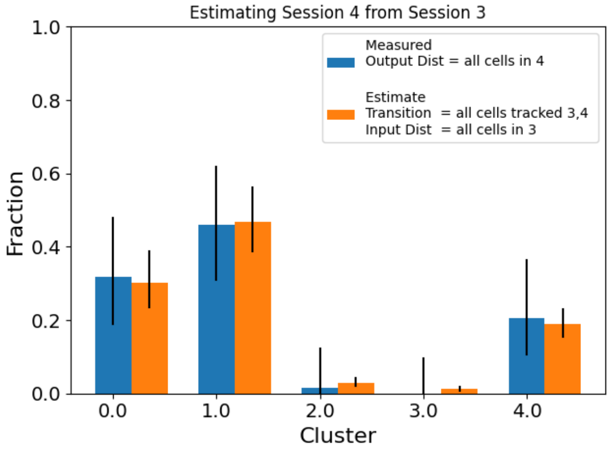
Excitatory Cells

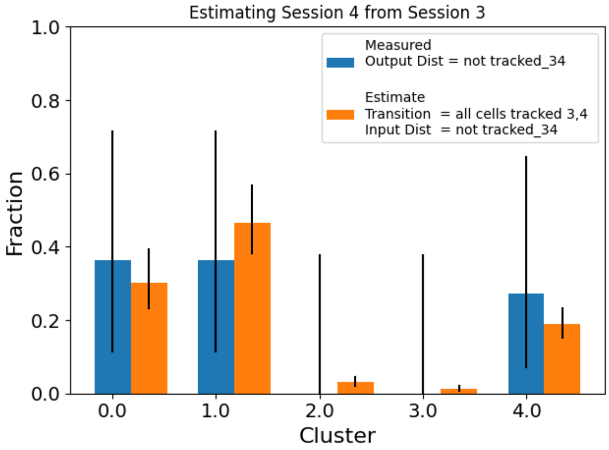


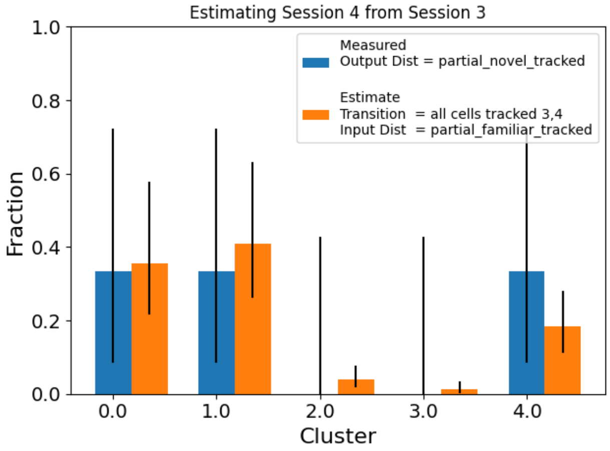




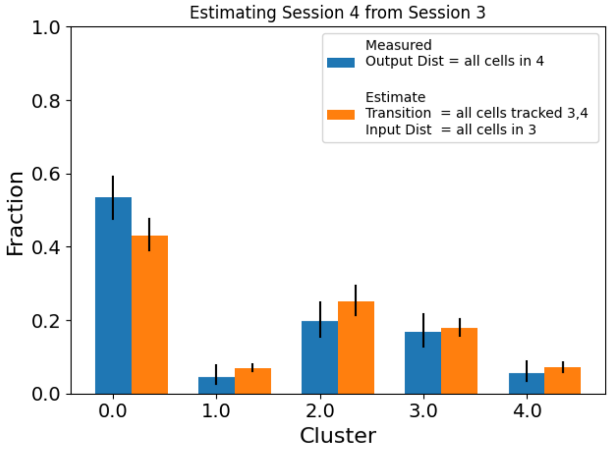
Sst Cells

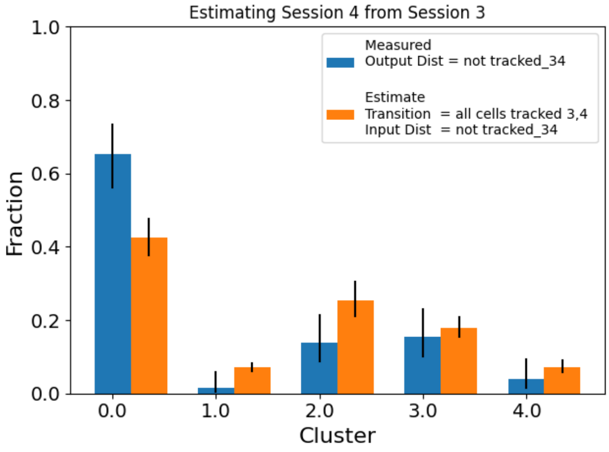


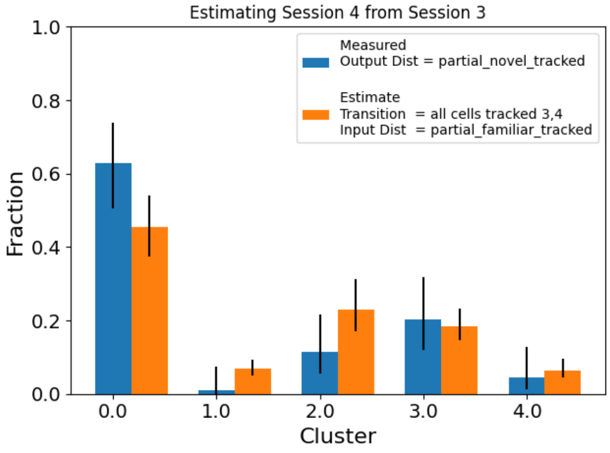




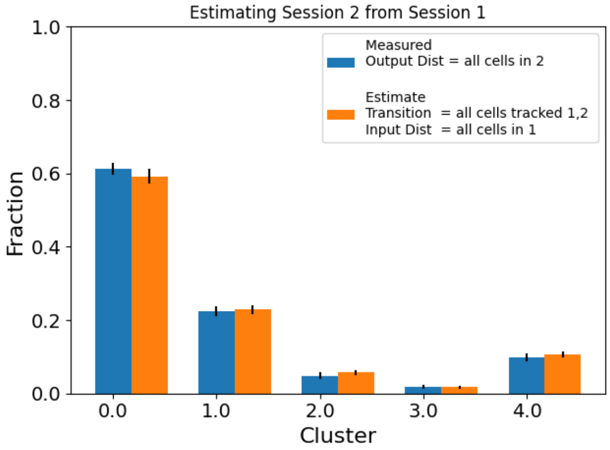
Vip Cells

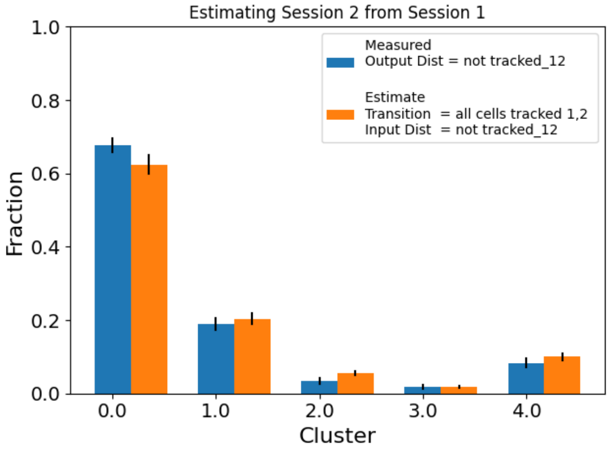

Estimating Session 4 from Session 3
Measured
Output Dist = all cells in 4
Estimate
Transition = all cells tracked 3,4
Input Dist = all cells in 3
Fraction
1.0
0.8
0.6
0.4
0.2
0.0
0.0
1.0
2.0
3.0
4.0
Cluster

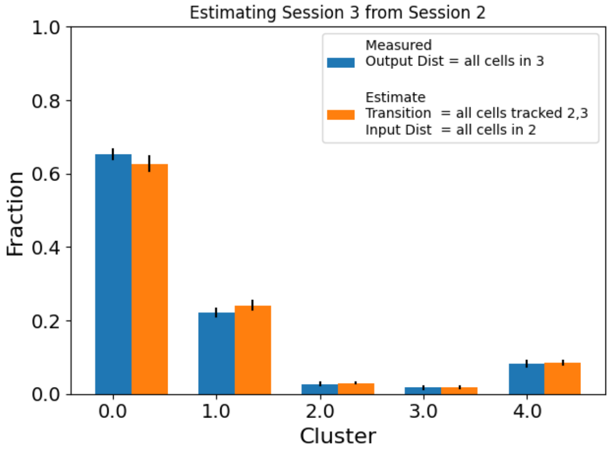

Estimating Session 4 from Session 3
Measured
Output Dist = not tracked_34
Estimate
Transition = all cells tracked 3,4
Input Dist = not tracked_34
Fraction
1.0
0.8
0.6
0.4
0.2
0.0
0.0
1.0
2.0
3.0
4.0
Cluster

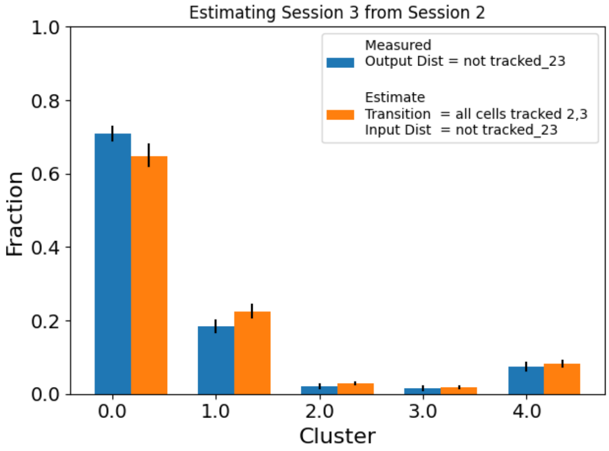

Estimating Session 4 from Session 3
Measured
Output Dist = partial_novel_tracked
Estimate
Transition = all cells tracked 3,4
Input Dist = partial_familiar_tracked
Fraction
1.0
0.8
0.6
0.4
0.2
0.0
0.0
1.0
2.0
3.0
4.0
Cluster

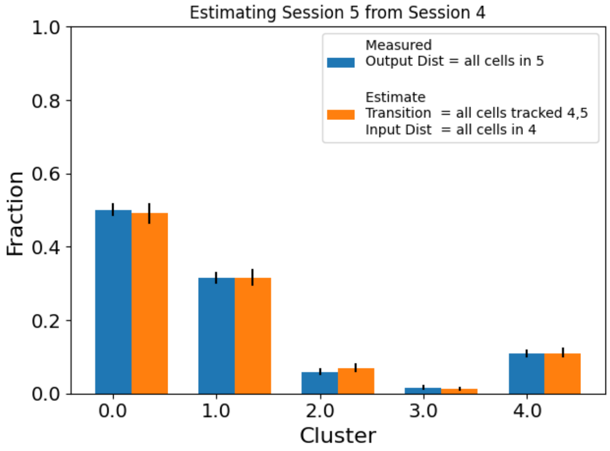
Excitatory Cells all transitions

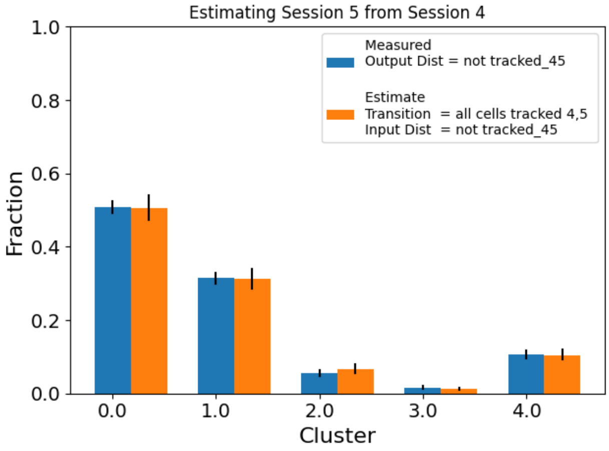


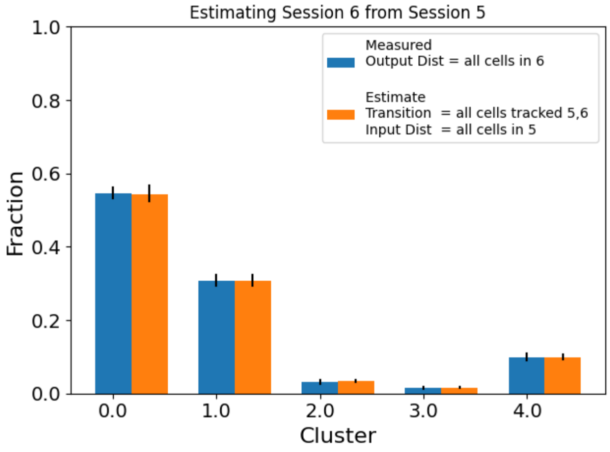


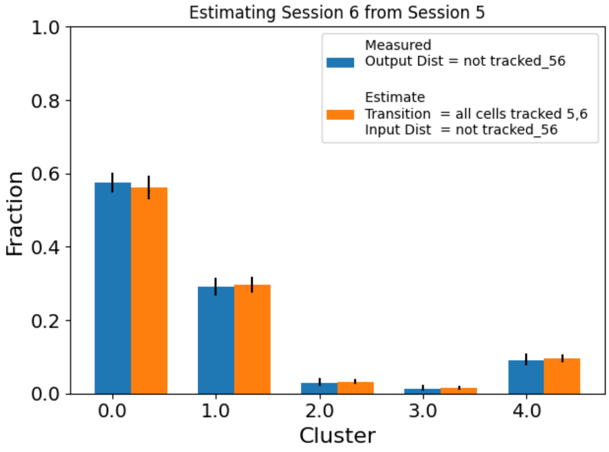




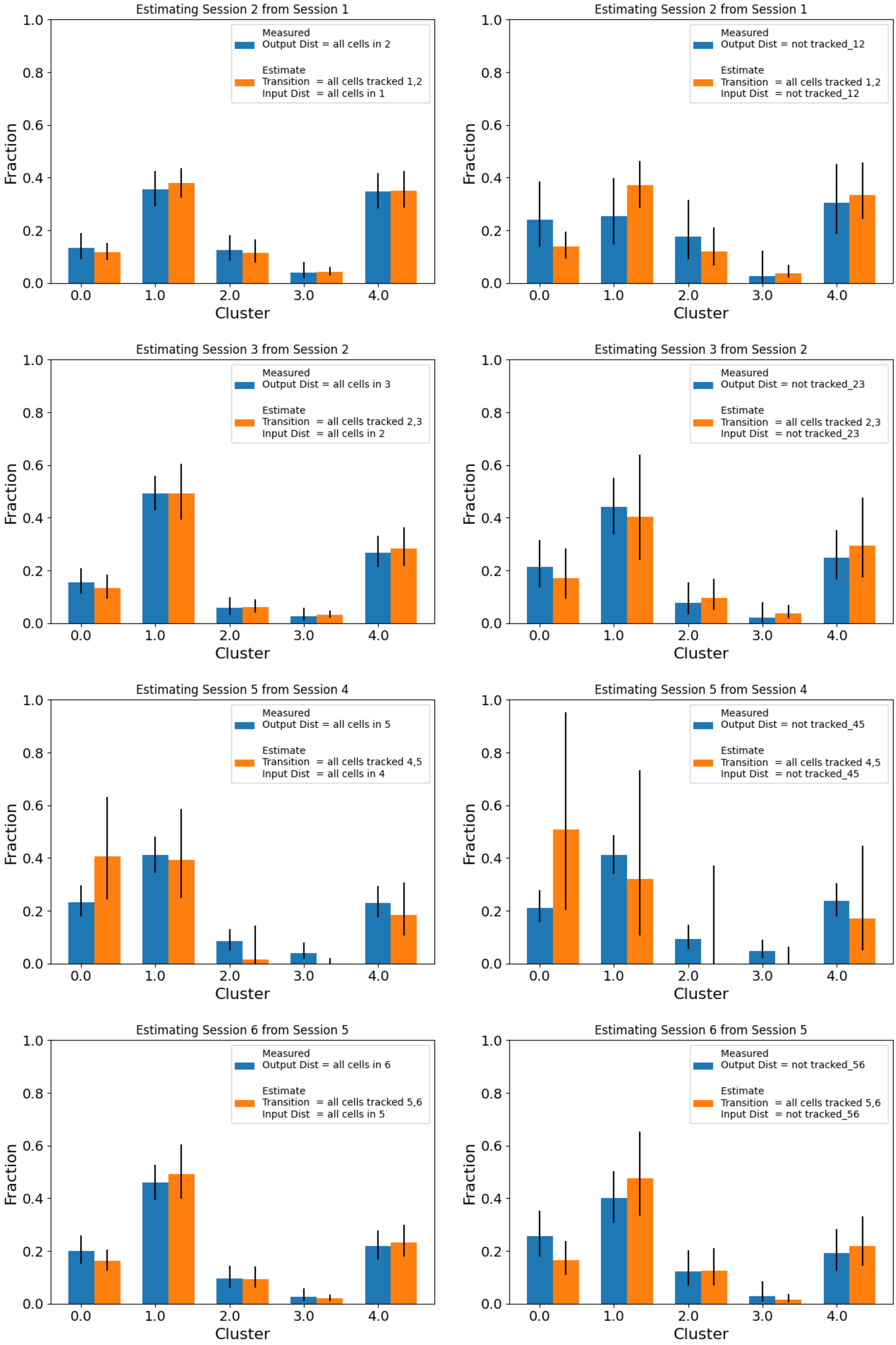




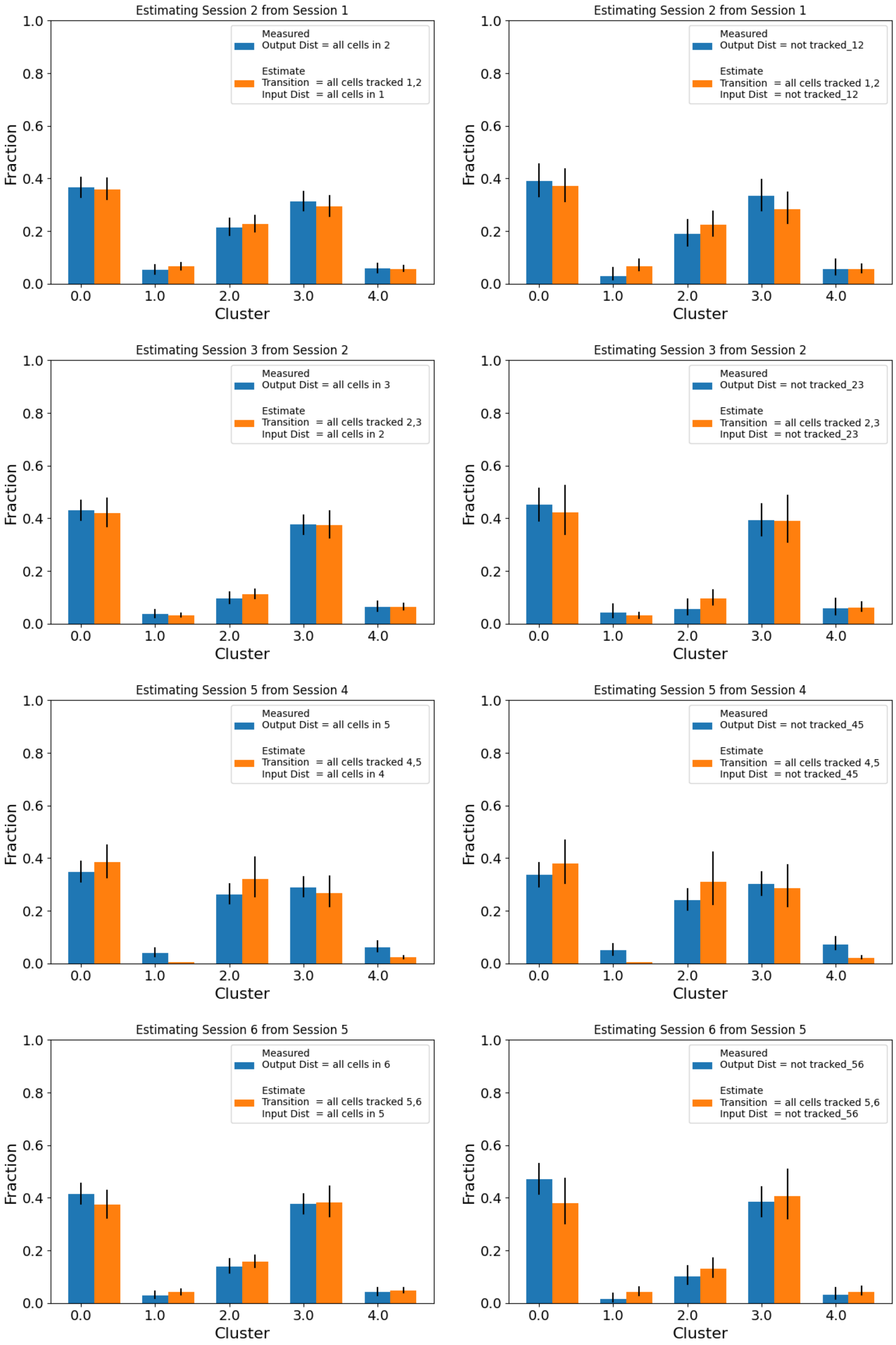




Sst Cells all transitions



Vip Cells all transitions


Estimating Session 2 from Session 1
Measured
Output Dist = all cells in 2
Estimate
Transition = all cells tracked 1,2
Input Dist = all cells in 1
Estimating Session 2 from Session 1
Measured
Output Dist = not tracked_12
Estimate
Transition = all cells tracked 1,2
Input Dist = not tracked_12
Estimating Session 3 from Session 2
Measured
Output Dist = all cells in 3
Estimate
Transition = all cells tracked 2,3
Input Dist = all cells in 2
Estimating Session 3 from Session 2
Measured
Output Dist = not tracked_23
Estimate
Transition = all cells tracked 2,3
Input Dist = not tracked_23
Estimating Session 5 from Session 4
Measured
Output Dist = all cells in 5
Estimate
Transition = all cells tracked 4,5
Input Dist = all cells in 4
Estimating Session 5 from Session 4
Measured
Output Dist = not tracked_45
Estimate
Transition = all cells tracked 4,5
Input Dist = not tracked_45
Estimating Session 6 from Session 5
Measured
Output Dist = all cells in 6
Estimate
Transition = all cells tracked 5,6
Input Dist = all cells in 5
Estimating Session 6 from Session 5
Measured
Output Dist = not tracked_56
Estimate
Transition = all cells tracked 5,6
Input Dist = not tracked_56
Fraction
Cluster