

Multiome data (RNA+ATAC), Qinyu Zhang @ David Bryder lab

Meeting 2022-02-08

Focus on integration

Focus on integrating/merging the two samples, agreed upon in our meeting last week.

Focus on integration

Focus on integrating/merging the two samples, agreed upon in our meeting last week.

The two samples:

1. EPCR
2. Total (aka Viable_RNA)

There are two conceptual approaches to integrating these multi-modal data.

There are two conceptual approaches to integrating these multi-modal data.

- Modality-based
 - Integrate the samples modality-wise
 - (EPCR RNA with Total RNA, and EPCR ATAC with Total ATAC)
 - then integrate the result's modalities (RNA & ATAC) with each other.
 - (joint-RNA with joint-ATAC)

There are two conceptual approaches to integrating these multi-modal data.

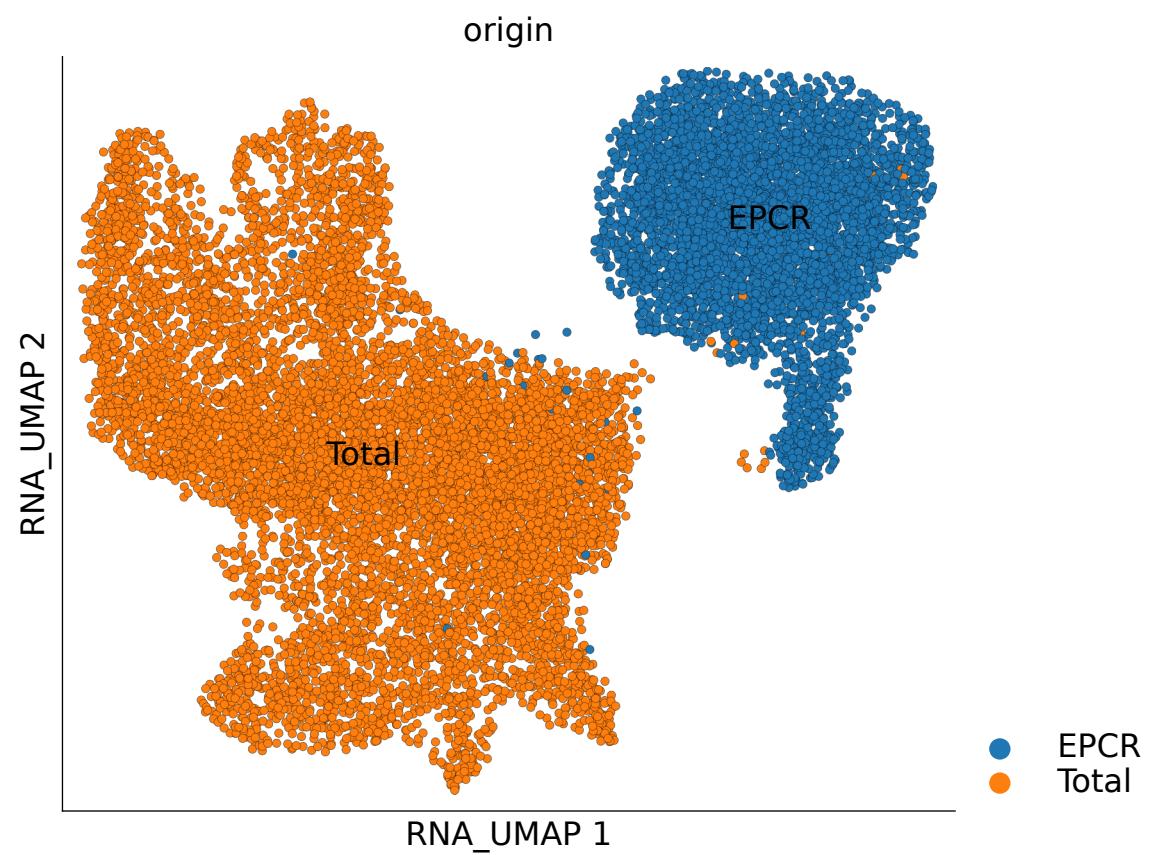
- Modality-based
 - Integrate the samples modality-wise
 - (EPCR RNA with Total RNA, and EPCR ATAC with Total ATAC)
 - then integrate the result's modalities (RNA & ATAC) with each other.
 - (joint-RNA with joint-ATAC)
- Sample-based
 - Integrate the modalities sample-wise,
 - (EPCR RNA with EPCR ATAC, and Total RNA with Total ATAC)
 - and then integrate the two samples with each other.
 - (EPCR-joint with Total-joint)

There are two conceptual approaches to integrating these multi-modal data.

- Modality-based
 - Integrate the samples modality-wise
 - (EPCR RNA with Total RNA, and EPCR ATAC with Total ATAC)
 - then integrate the result's modalities (RNA & ATAC) with each other.
 - (joint-RNA with joint-ATAC)
- Sample-based
 - Integrate the modalities sample-wise,
 - (EPCR RNA with EPCR ATAC, and Total RNA with Total ATAC)
 - and then integrate the two samples with each other.
 - (EPCR-joint with Total-joint)

For now I went with the first approach, modality-based. Specifically, integrating EPCR RNA with Total RNA.

Not yet integrated (baseline/starting point)



I tried a couple of approaches:

I tried a couple of approaches:

- Integration through projection (EPCR on total)

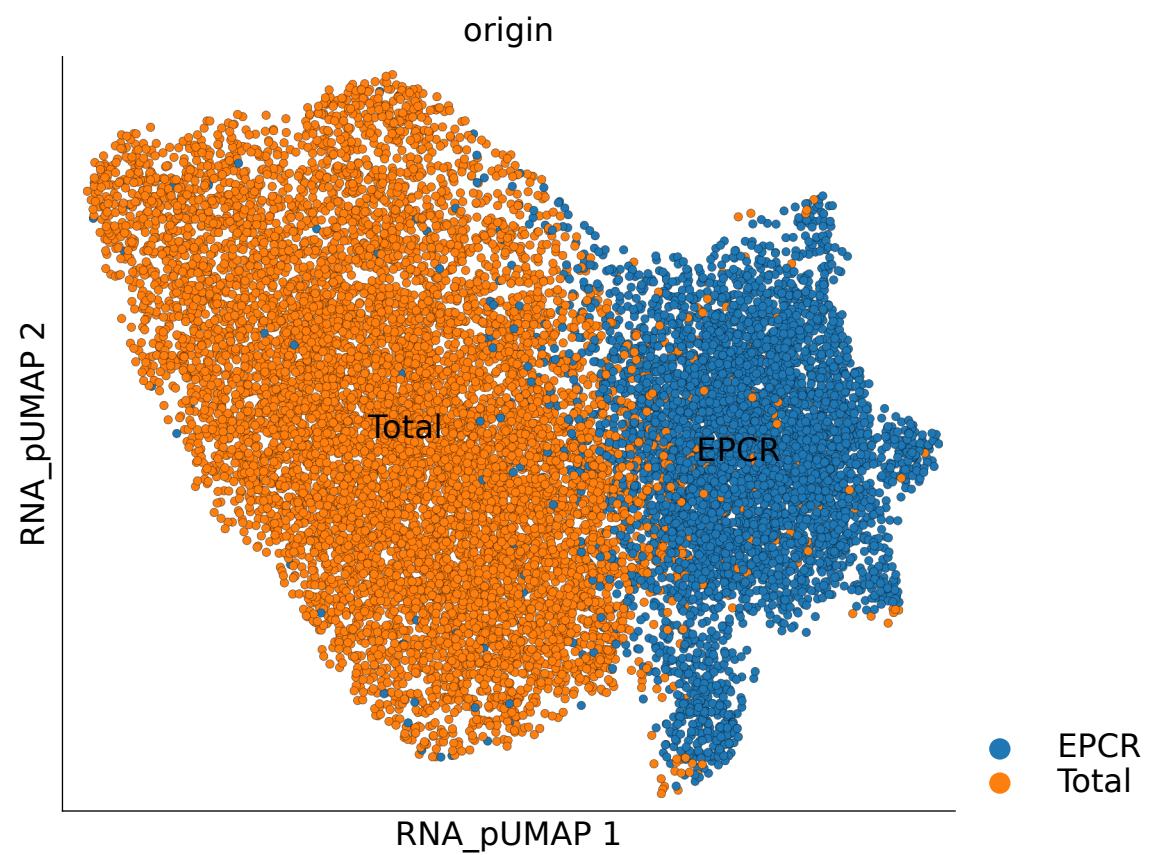
I tried a couple of approaches:

- Integration through projection (EPCR on total)
- Integration through Partial PCA

I tried a couple of approaches:

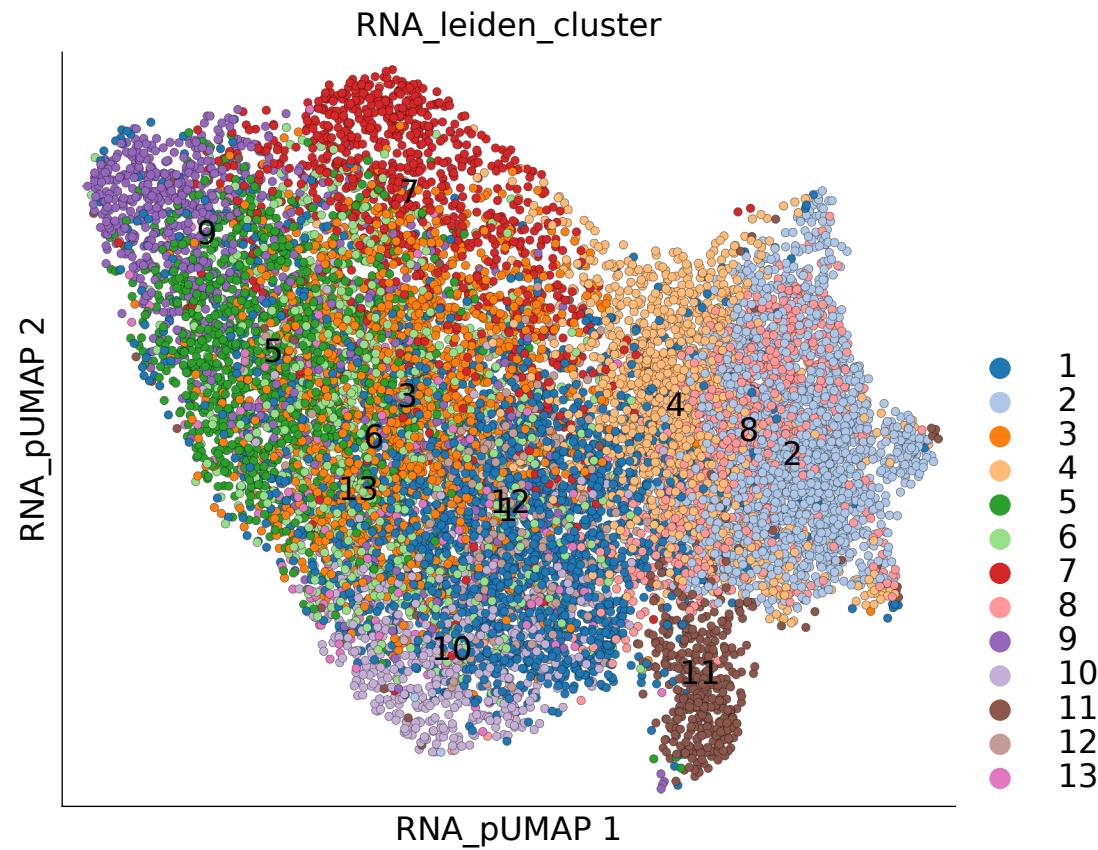
- Integration through projection (EPCR on total)
- Integration through Partial PCA
- Integration with Harmony

Partial PCA

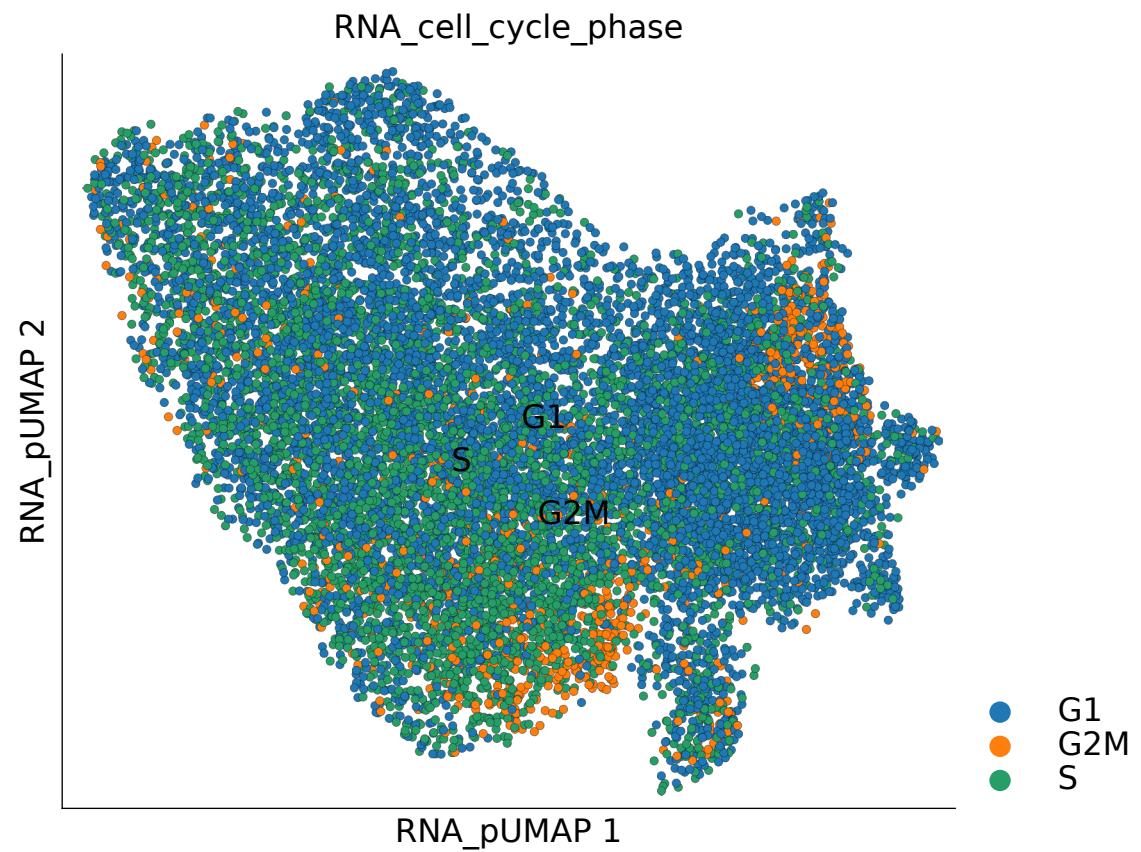


Not very good. Barely merged.

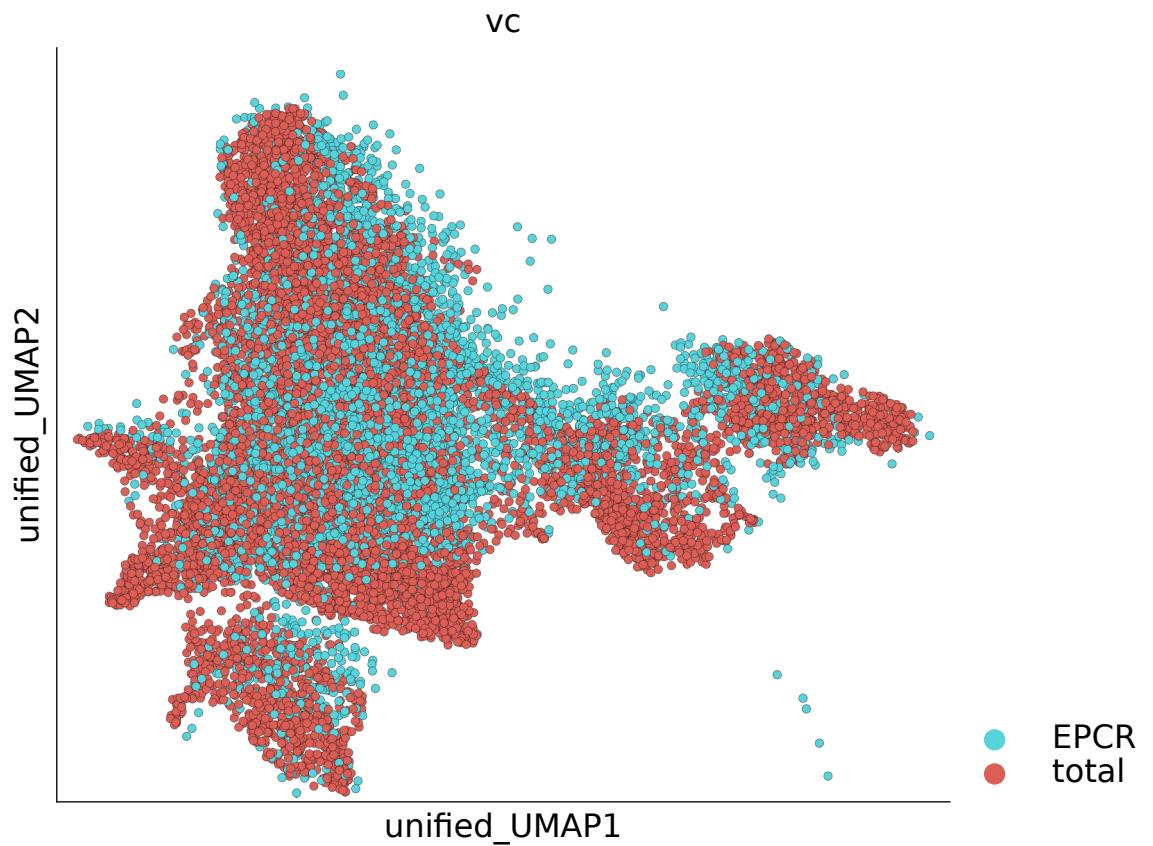
Partial PCA, clustering



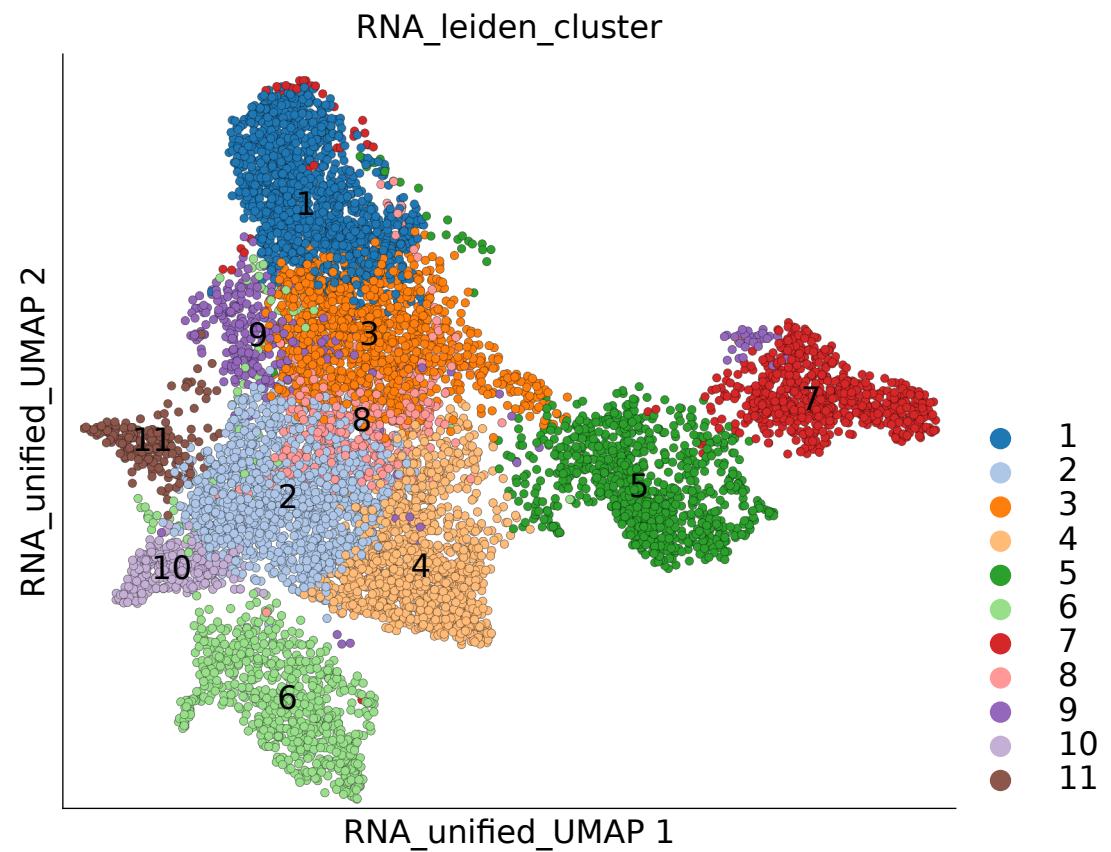
Cell cycle stage



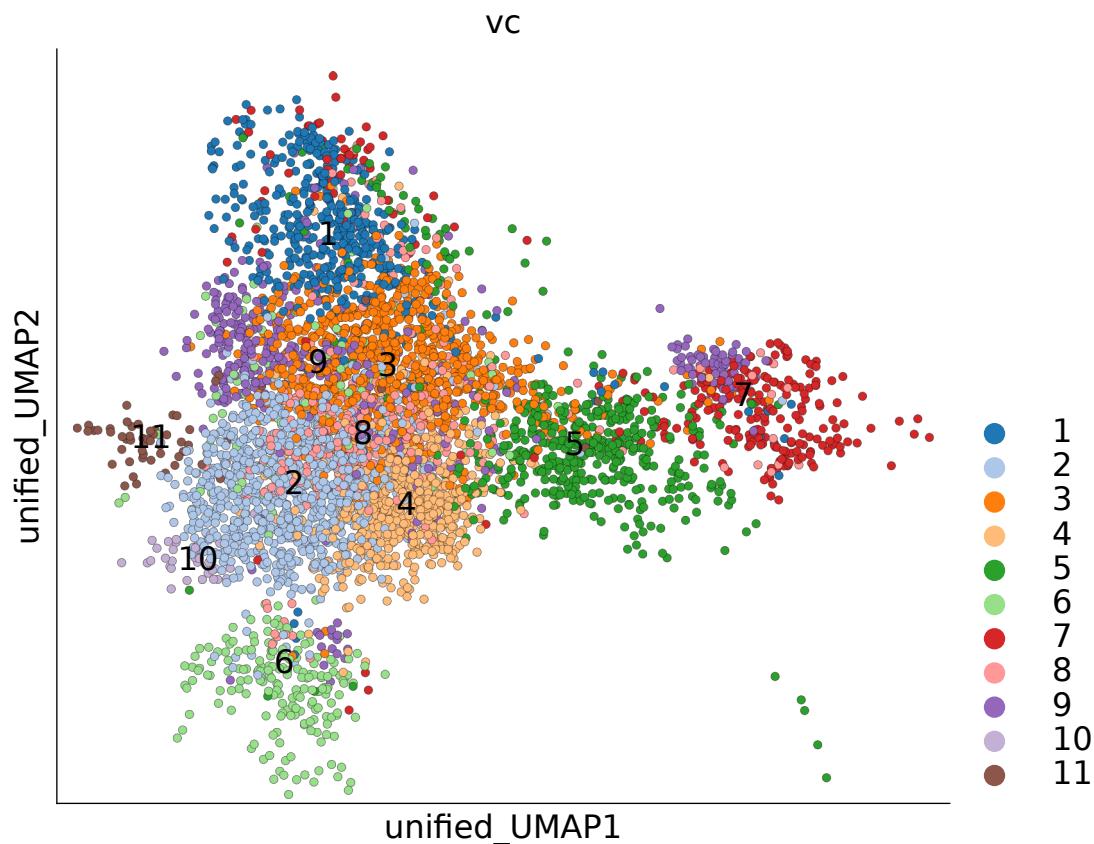
Projection (scarf), "unified UMAP"



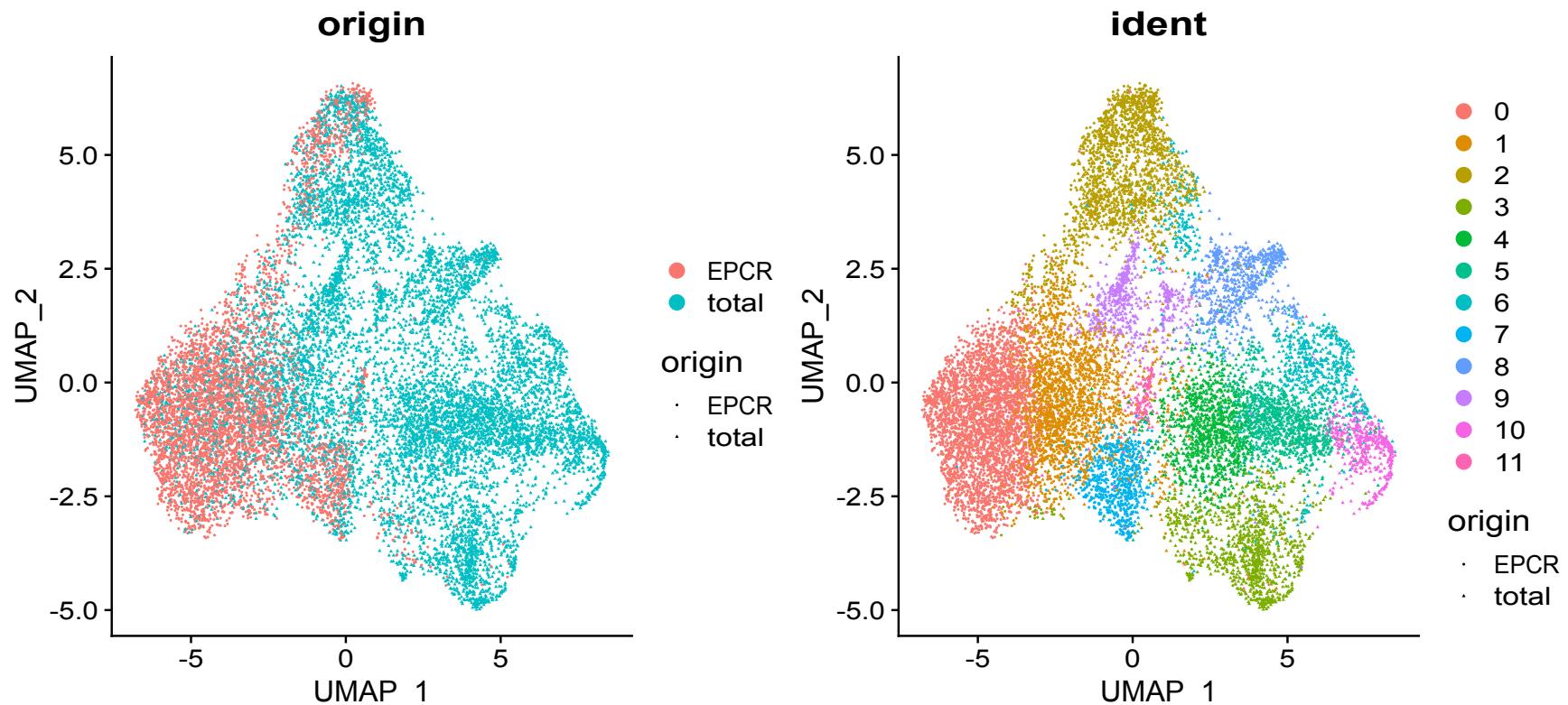
Unified UMAP with Total clusters



Unified UMAP with Total clusters, only EPCR cells shown



Harmony



Discussion

Projection gave a more evenly distribution of EPCR cells over Total cells, while Harmony placed them mostly to one end of the data. Biologically, this seems to make more sense than from the projection approach, considering that EPCR cells should be mostly immature cells while Total should be more diverse. Based on this, I would say we go with Harmony.

Discussion

Projection gave a more evenly distribution of EPCR cells over Total cells, while Harmony placed them mostly to one end of the data. Biologically, this seems to make more sense than from the projection approach, considering that EPCR cells should be mostly immature cells while Total should be more diverse. Based on this, I would say we go with Harmony.

Of course, there are a lot of integration methods (and also conceptual approaches as I mentioned before). We can try others, but from our discussions so far I think the Harmony result align with the biological interpretation, which should be good to base further analysis on.

