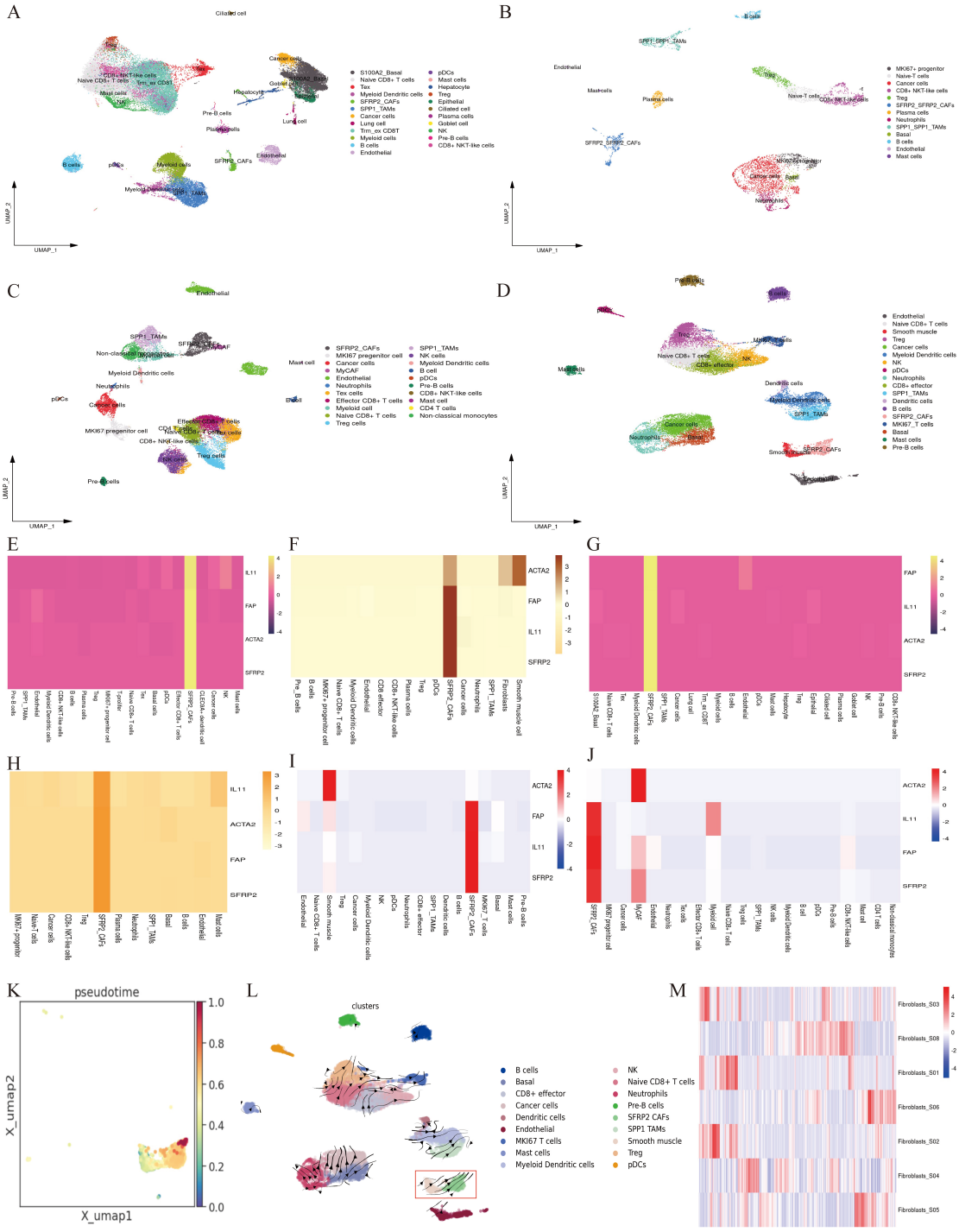
# Supplementary materials

# SFRP2 Cancer Associated Fibroblasts Drive Tumorigenesis in Head and Neck Squamous Cell Carcinoma

**Supplementary Table 1**

| Dataset | Total samples | Matched samples (Primary vs Metastasis) | Metastasis samples |
|---|---|---|---|
| GSE181919 | 23 | 8 | 4 |
| GSE188737 | 14 | 14 | 7 |
| GSE234933 | 52 | 0 | 12 |
| GSE173468 | 15 | 2 | 1 |
| GSE164690 | 18 | 0 | 0 |
| GSE215403 | 12 | 0 | 0 |
| GSE181300 | 8 (paraffin) | 0 | 0 |

## Supplementary Figure1

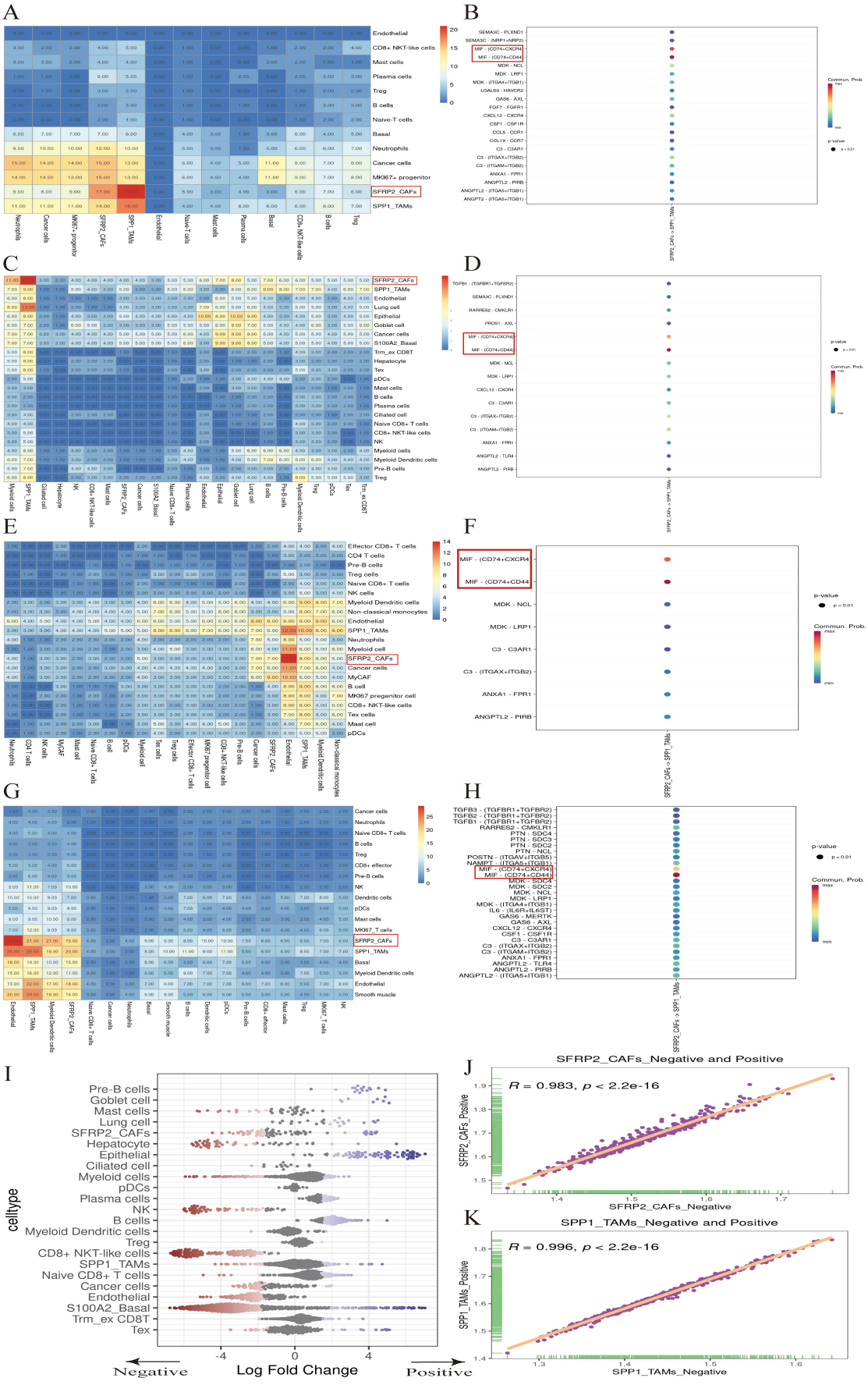



**Supplementary Figure1 legend**

**A)** In the study of single-cell objects from advanced head and neck squamous cell carcinoma (HNSCC) GSE234933, 23 distinct cell types were identified. **B)** In the study of single-cell objects from advanced head and neck squamous cell carcinoma (HNSCC) GSE173468, 13 distinct cell types were identified. **C)** In the study of single-cell objects from advanced head and neck squamous cell carcinoma (HNSCC) GSE164690, 21 distinct cell types were identified. **D)** In the study of single-cell objects from advanced head and neck squamous cell carcinoma (HNSCC) GSE215403, 18 distinct cell types were identified. **E-J)** Heatmaps depict the expression of four markers, namely

SFRP2, FAP, ACTA2, and IL11, across the datasets GSE18191, GSE188737, GSE234933, GSE173468, GSE215403 and GSE164690. **K)** UMAP representation highlighting the starting cell (blue) with the lowest SFRP2 expression, and the terminal cells (red) as predicted by Palantir in GSE215403. **L)** Velocities derived from the dynamical model for GSE215403 HNSCC single cell dataset are visualized as streamlines in a UMAP-based embedding. **M)** Heatmap depict the abundance of seven Fibroblasts subtypes in meta cohort.

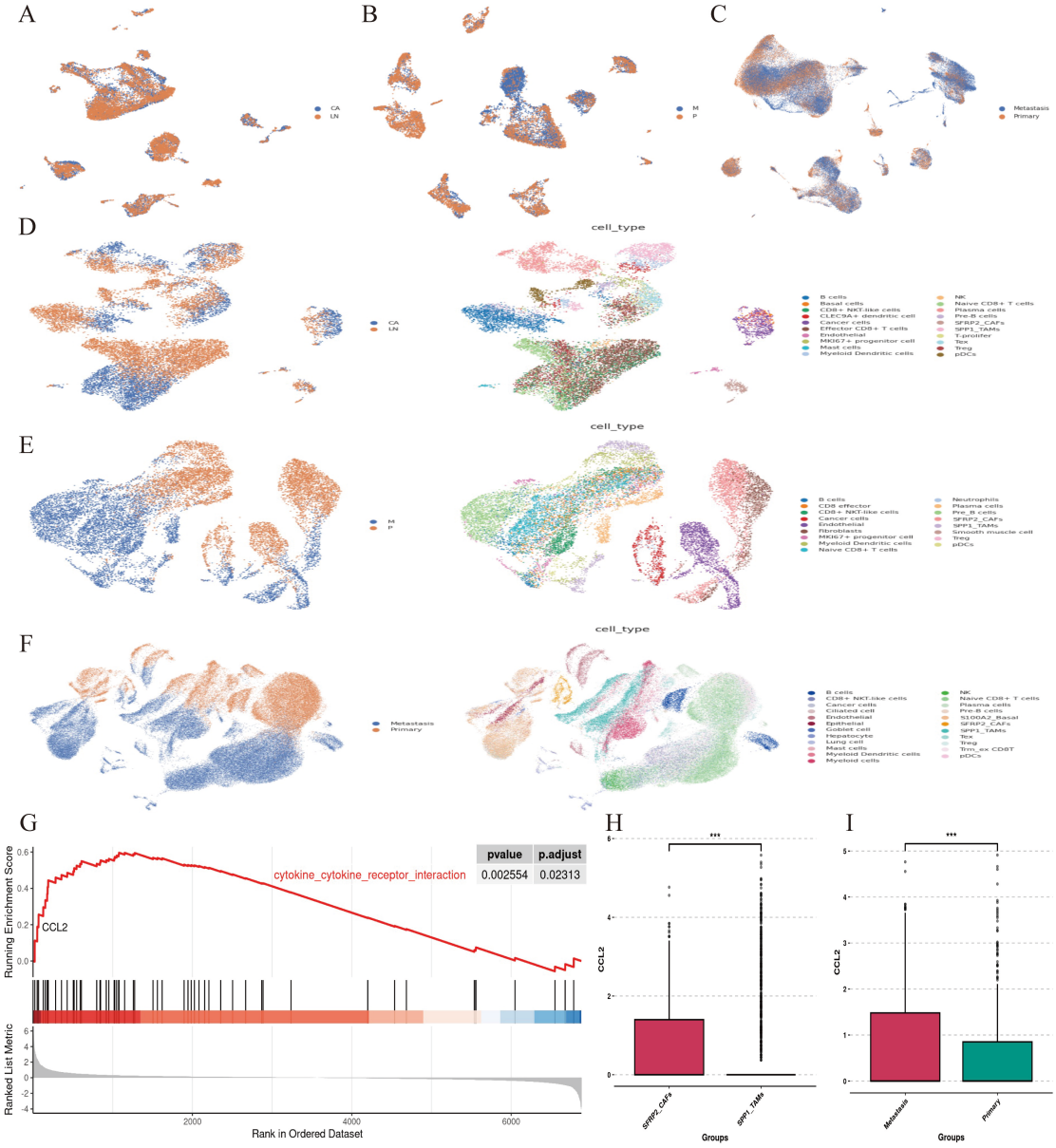
# Supplementary Figure2



**Supplementary Figure2 legend**

**A-B)** In GSE173468, cell communication revealed that SFRP2_CAFs and SPP1_TAMs play a dominant role in the tumor microenvironment of HNSCC metastases (the "communication score" between SFRP2_CAFs and SPP1_TAMs = 21, higher than the "communication score" between any other pair of cells). The MIF-CD74 ligand-receptor interaction predominated in this cell communication. **C-D)** In GSE234933, cell communication revealed that SFRP2_CAFs and SPP1_TAMs play a dominant role in the tumor microenvironment of HNSCC metastases (the "communication score" between SFRP2_CAFs and SPP1_TAMs = 15, higher than the "communication score" between any other pair of cells). The MIF-CD74 ligand-receptor interaction predominated in this cell communication. **E-F)** In GSE164690, HNSCC single cell database with no metastatic lesions, cell communication revealed that SFRP2_CAFs crosstalk with Endothelial play a dominant role in the tumor microenvironment of HNSCC. The interaction facilitated by the MIF-CD74 ligand-receptor pair remains a discernible feature within the crosstalk between SFRP2_CAFs and SPP1_TAMs. **G-H)** In GSE215403, another no metastatic lesions, cell communication revealed that SFRP2_CAFs crosstalk with Endothelial still play a dominant role in the tumor microenvironment of HNSCC. The MIF-CD74 ligand-receptor interaction continues to be observed within the crosstalk between SFRP2-CAFs and SPP1-TAMs. **I)** Beeswarm plot of the distribution of log fold change across HPV negative and HPV positive in neighborhoods containing cells from different cell type clusters. Differential abundance neighborhoods at FDR 10% are colored. **J)** The correlation ($R = 0.983$, $p < 2.2e-16$) between HPV negative SFRP2_CAFs genes' expression and HPV positive SFRP2_CAFs genes' expression. **K)** The correlation ($R = 0.996$, $p < 2.2e-16$) between HPV negative SPP1_TAMs genes' expression and HPV positive SPP1_TAMs genes' expression.

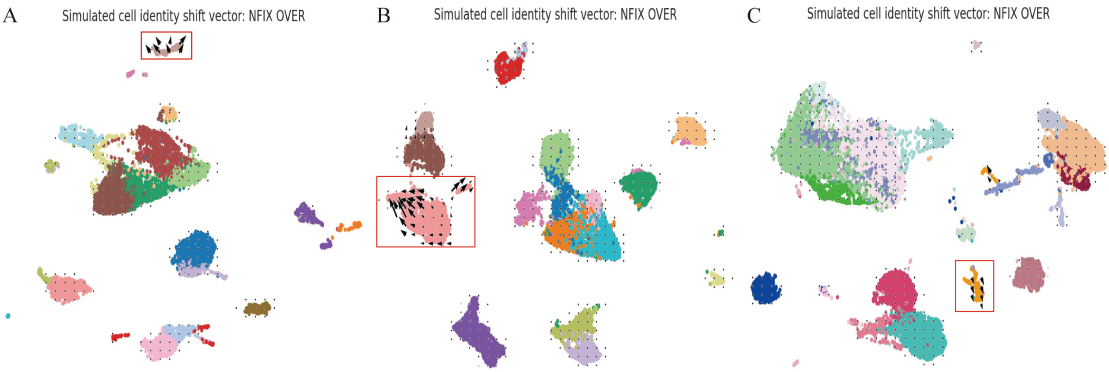
## Supplementary Figure3




## Supplementary Figure3 legend

**A-C)** The single-cell datasets GSE181919, GSE188737, and GSE234933, after batch effect removal, show that CA (primary site) and LN (lymph node metastatic site), as well as P (primary site) and M (metastatic site), Metastasis and Primary cannot be distinctly separated. **D-F)** The single-cell datasets GSE181919, GSE188737, and GSE234933, after fine-tuned application, primary and metastatic site can be well separated. **G)** CCL2 exhibited a significant positive regulatory effect of cytokine-cytokine-receptor signaling pathways in SFRP2_CAFs, GSE234933, p = 0.002554. **H)** CCL2 expression in GSE234933, focusing on SFRP2_CAFs and SPP1_TAMs (* < 0.05; ** < 0.01; *** < 0.001, **** < 0.0001). **I)** CCL2 expression in GSE234933, focusing on Metastasis and Primary (* < 0.05; ** < 0.01; *** < 0.001, **** < 0.0001).

## Supplementary Figure4




**Supplementary Figure4 legend**

**A-C)** CellOracle simulation cell-state transitions in the NFIX over expression simulation; The resulting cell-state transition vectors were summarized and projected onto a force-directed graph; red box; left to right; GSE181919, GSE188737 and GSE234933.

**Supplementary Western blot Figure_1**

Figure4A

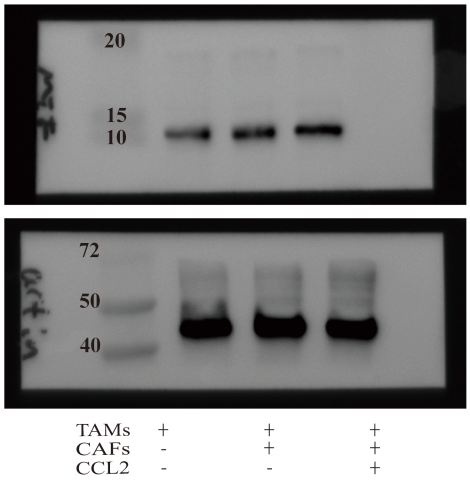
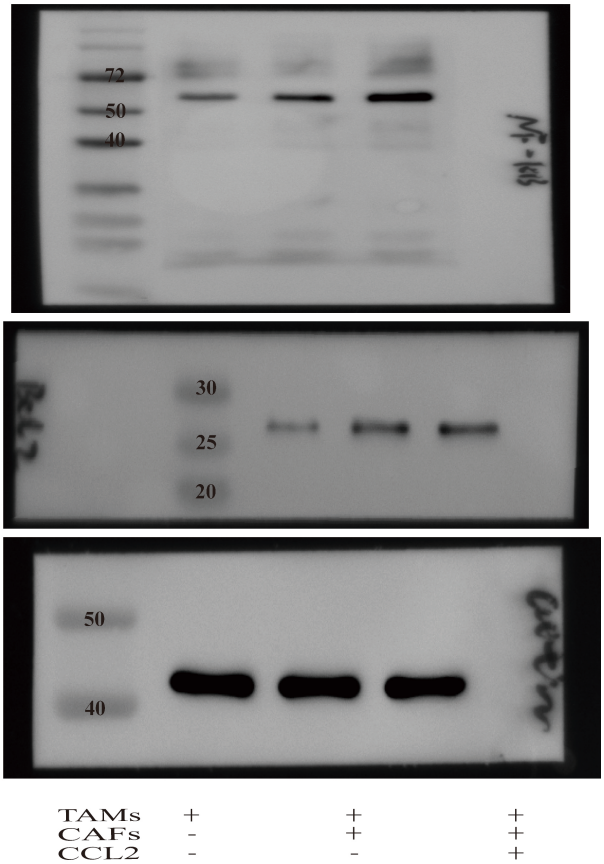



Figure4O



**Supplementary Western blot Figure_2**

Figure5B

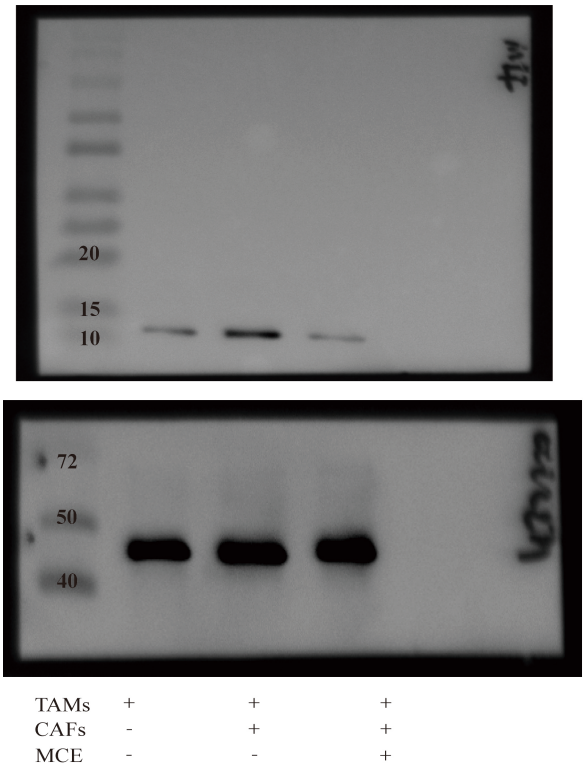



Figure6N