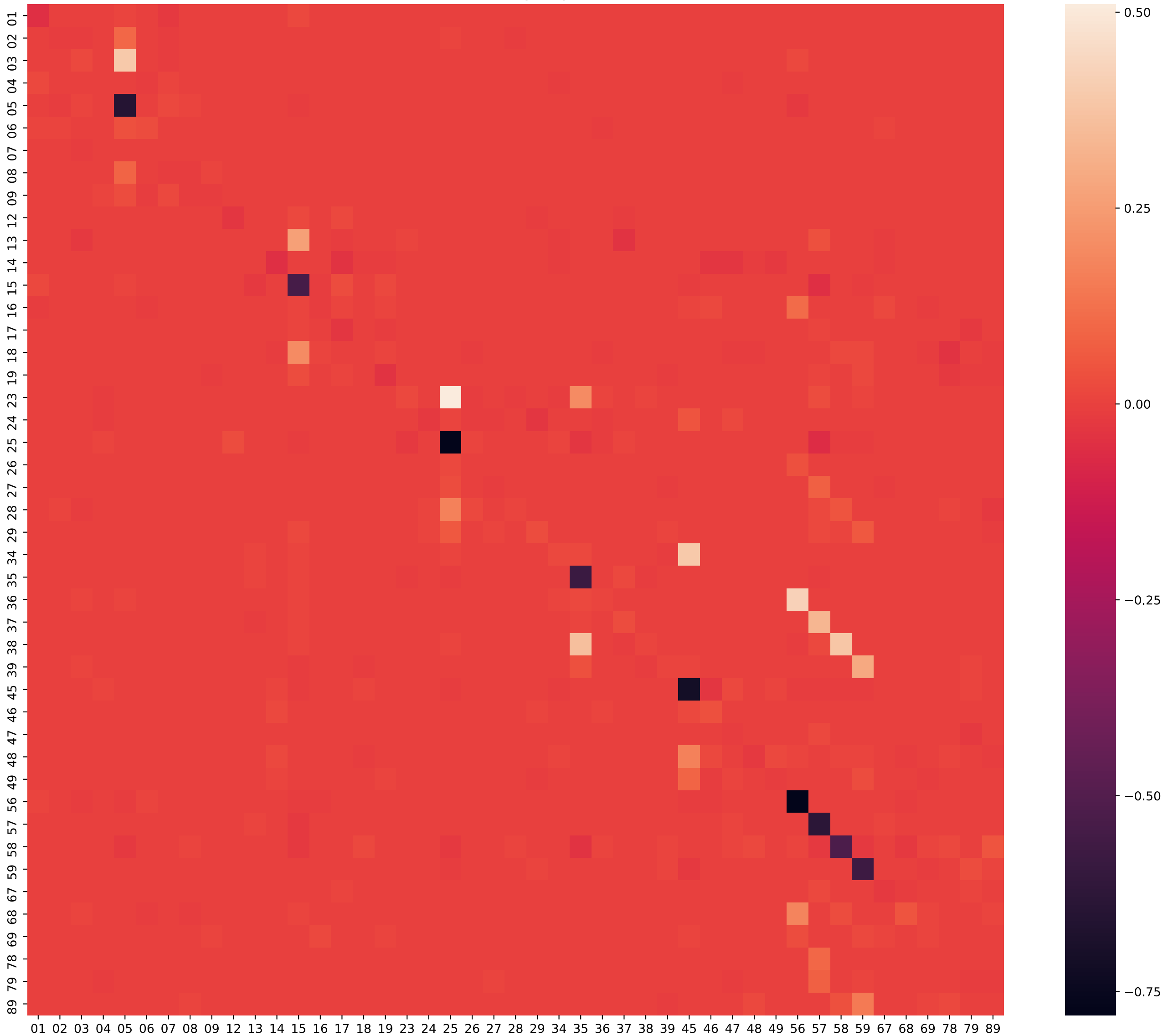
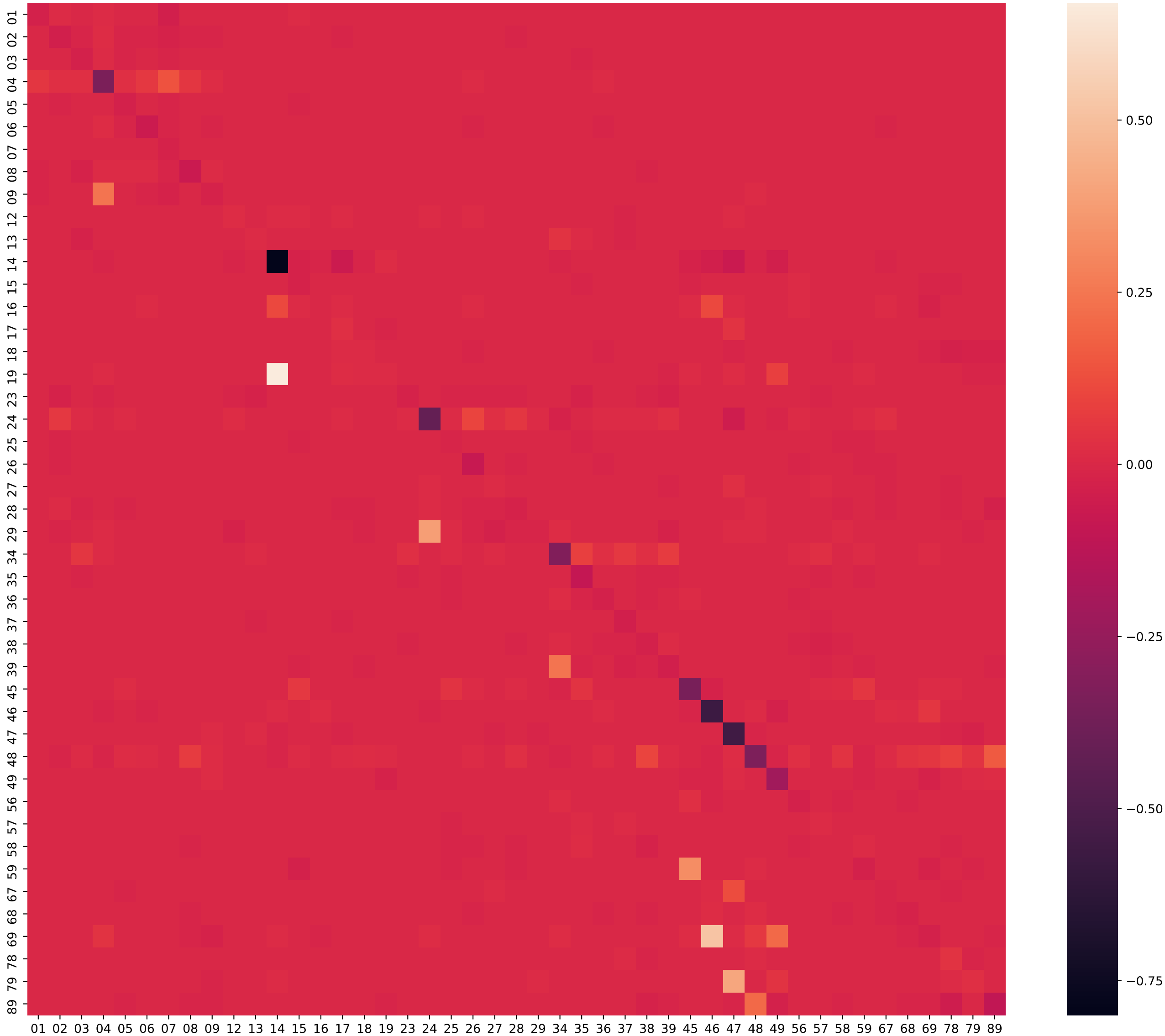




Ablation of Unit 1 which originally detects Number 5

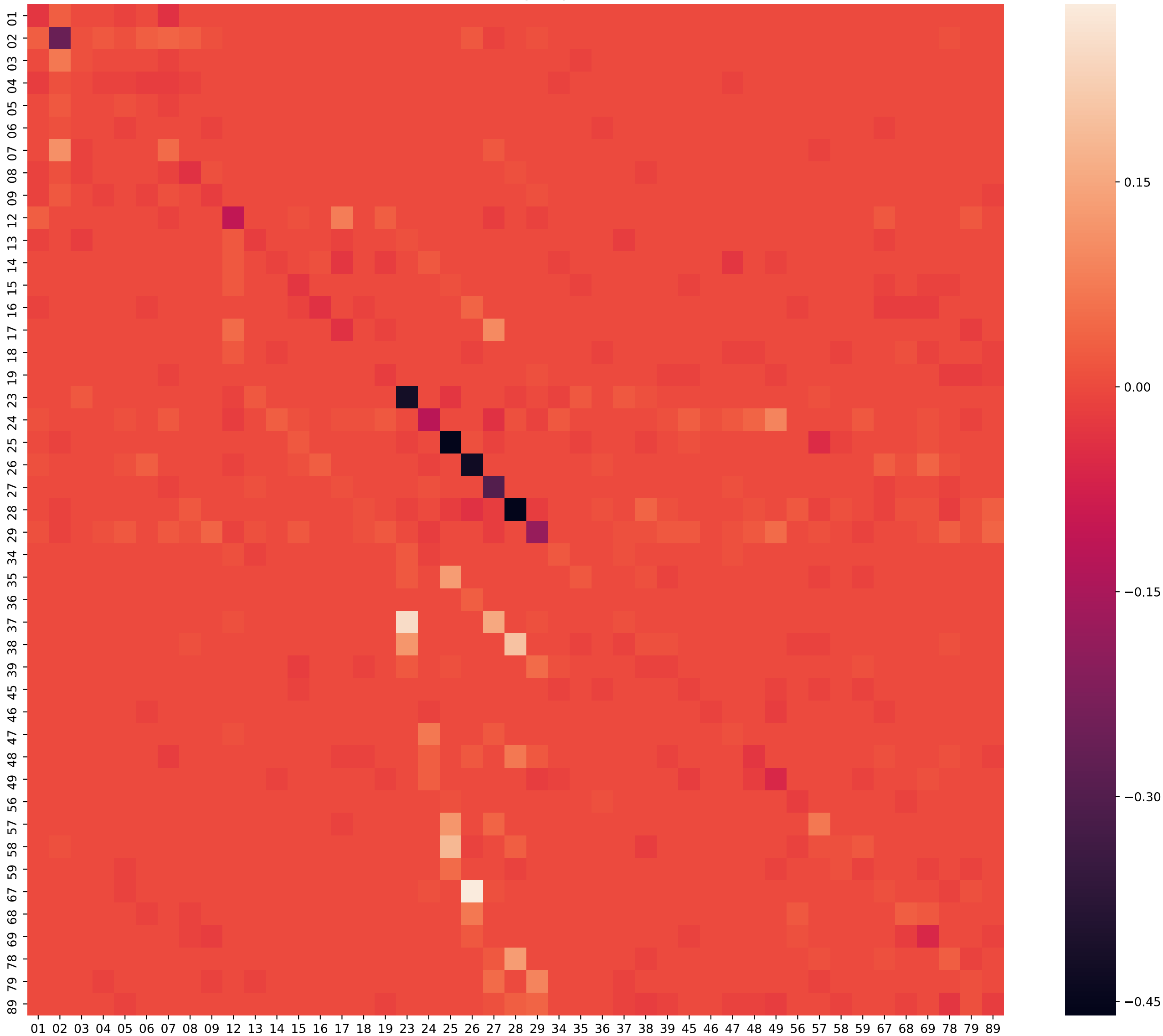


Ablation of Unit 2 which originally detects Number 4

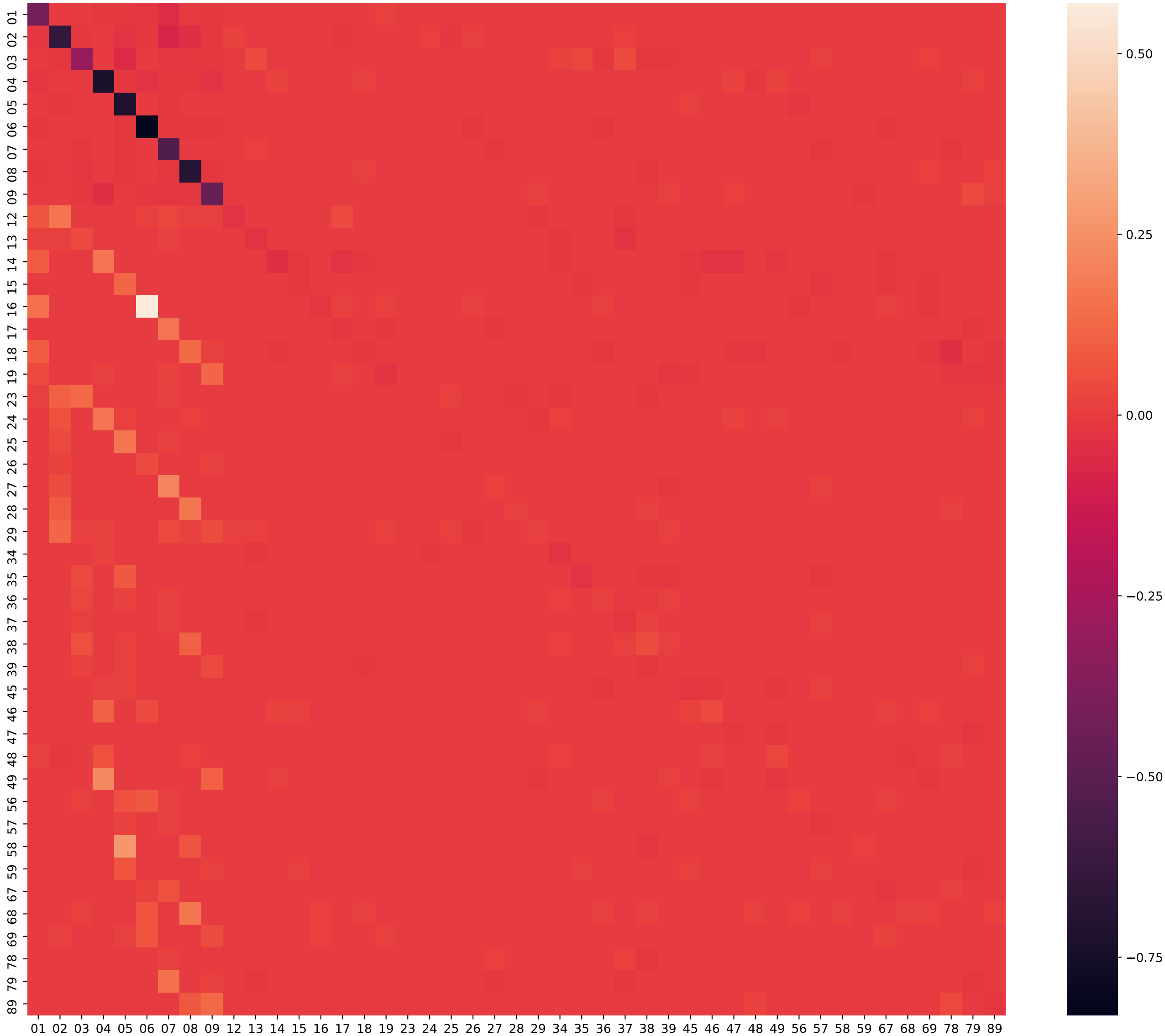


Heatmap visualization of the correlation matrix for 89 variables. The color scale ranges from -0.75 (dark purple) to 0.50 (light yellow). The diagonal is dark purple (-0.75). The matrix shows a block-like structure with high positive correlations (yellow) in the top-left and bottom-right blocks, and negative correlations (purple) in the off-diagonal blocks.

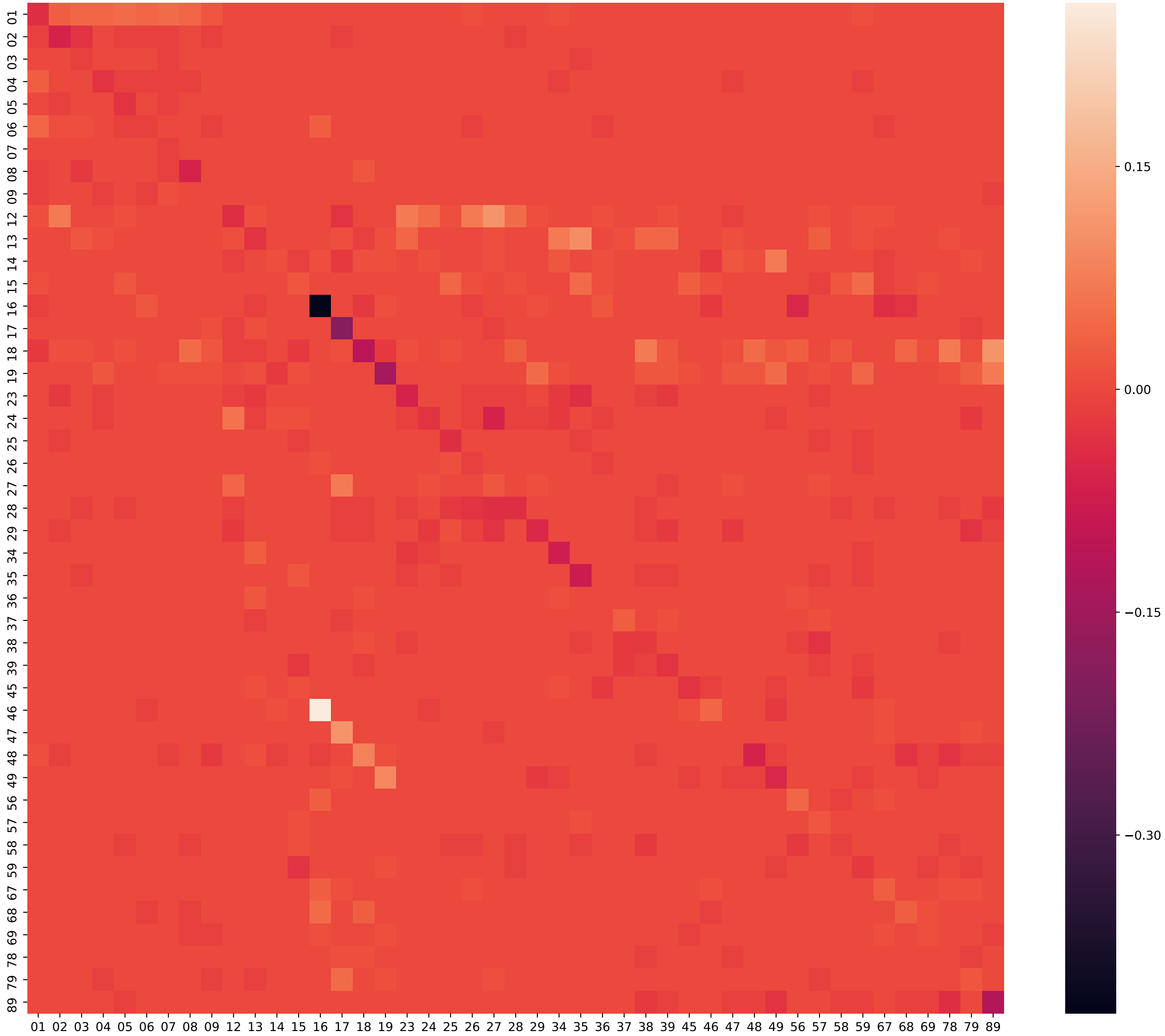
Ablation of Unit 4 which originally detects Number 2



Ablation of Unit 5 which originally detects Number 0



Ablation of Unit 6 which originally detects Number 1







Ablation of Unit 8 which originally detects Number 8

