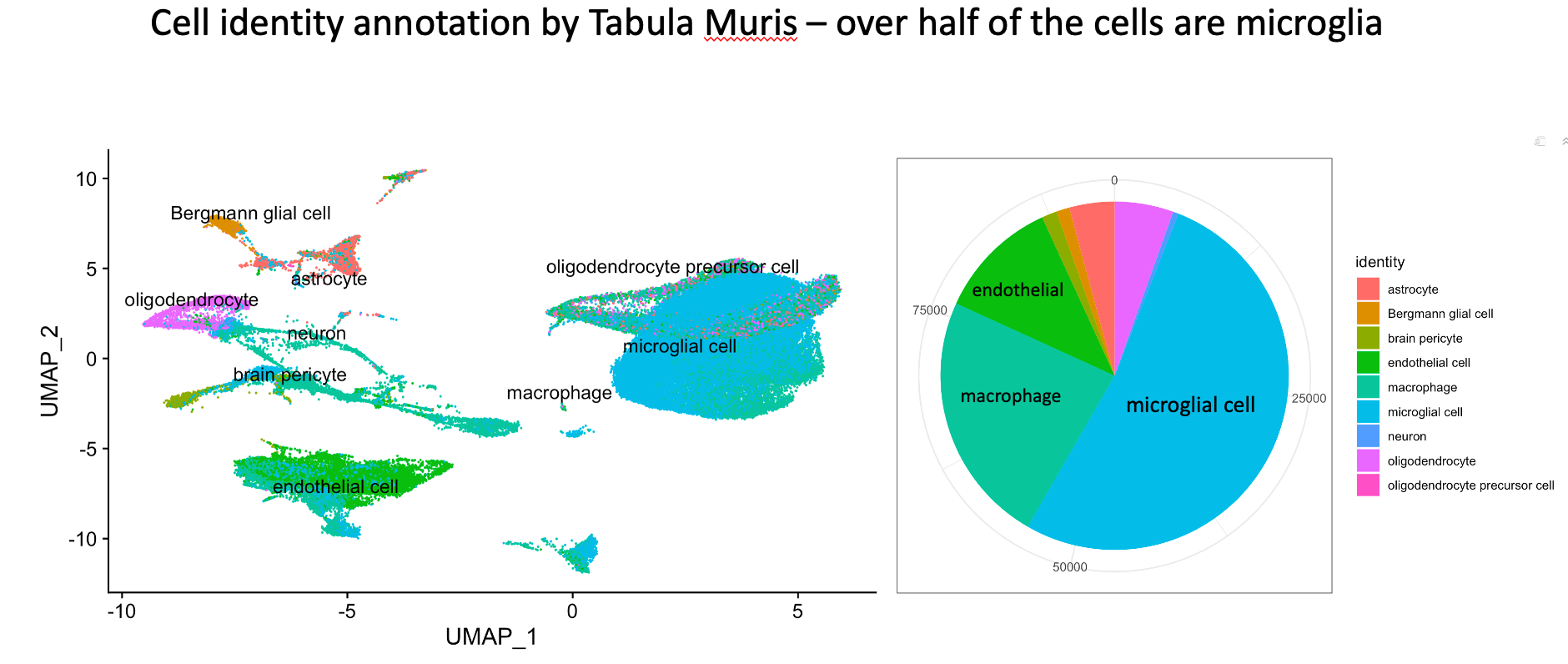
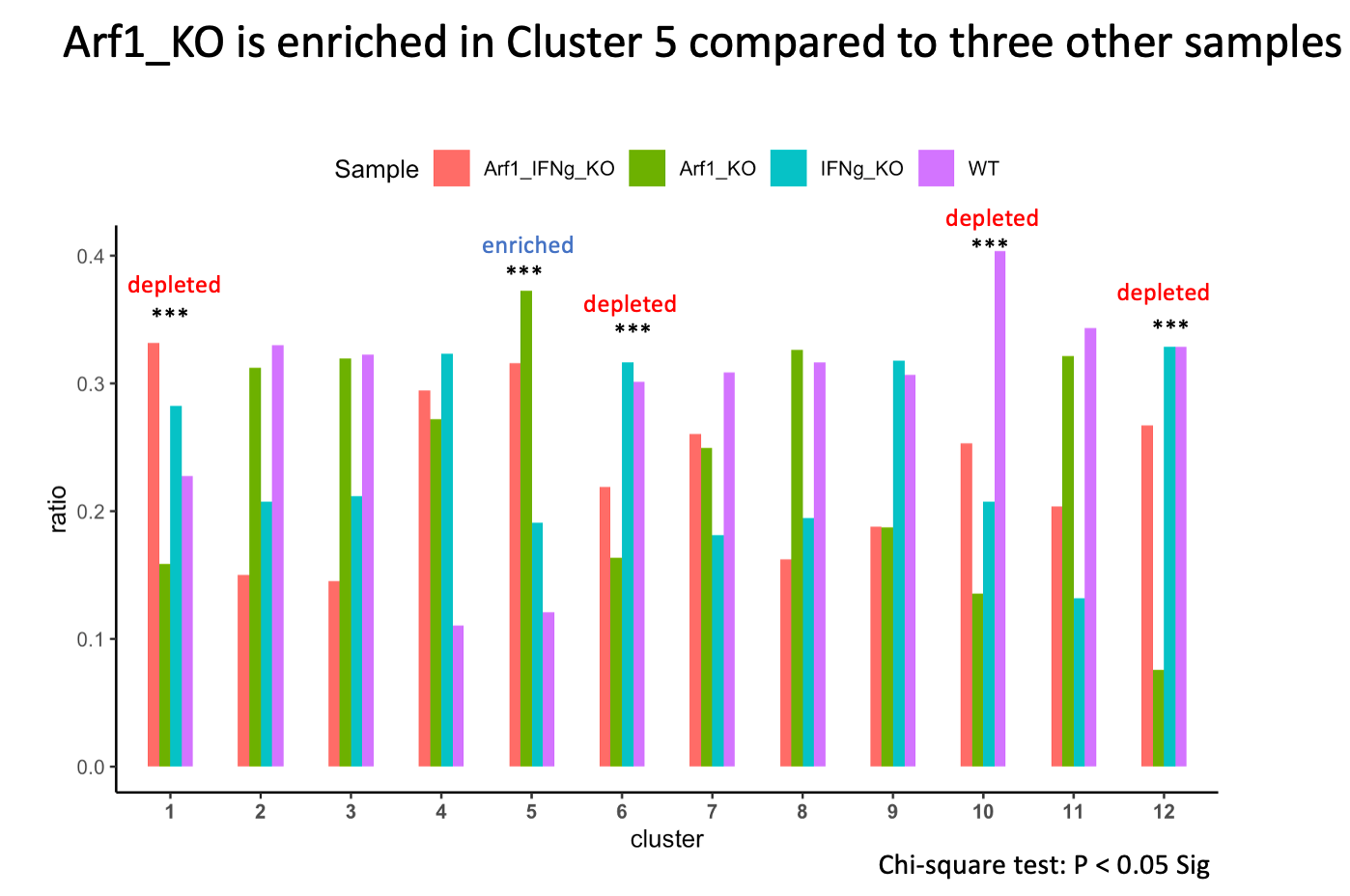
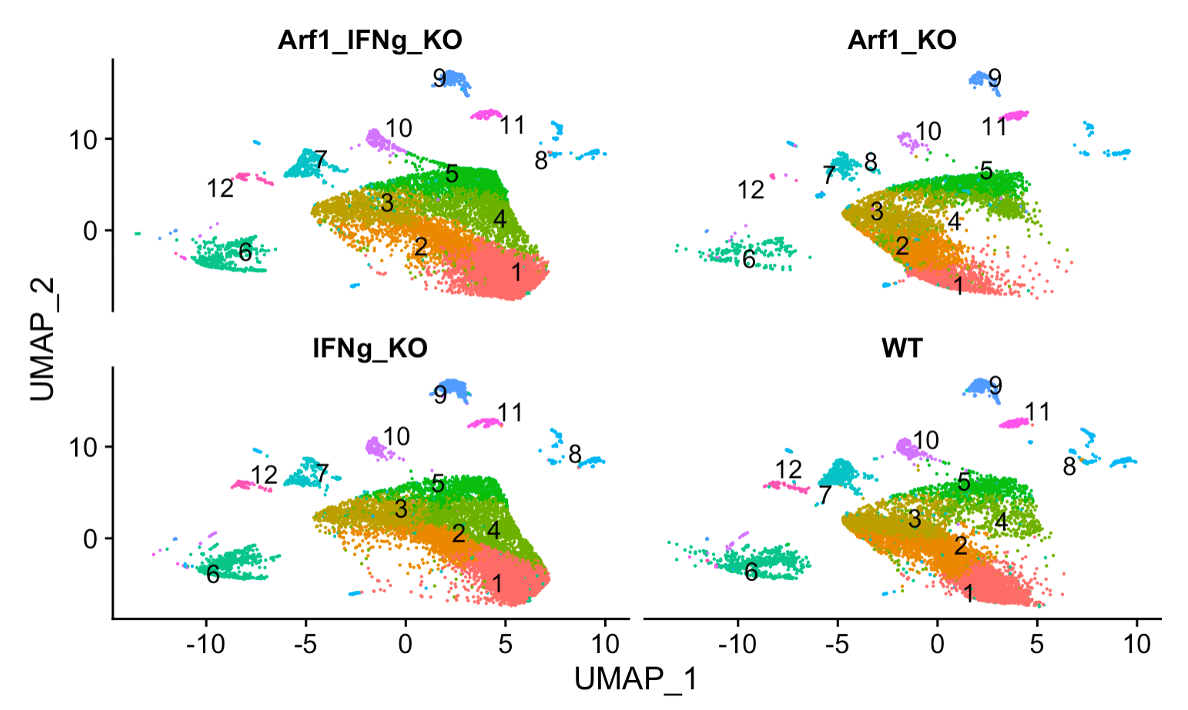
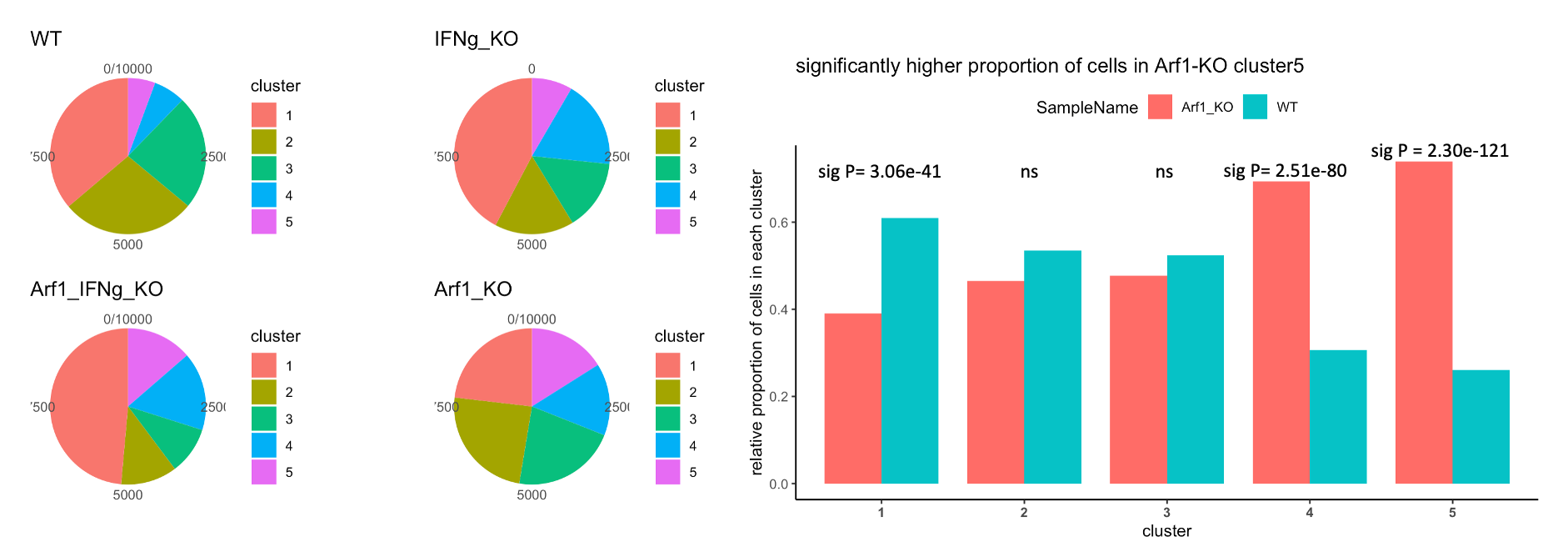
In this analysis, we used UMAP to create a map of merged cells from four samples of mouse brain. Tabula Muris was used to annotate cell identities. Cells annotated as microglial cells accounted for more than over 50% of total cell populations. Microglia were extracted, retransformed for all subsequent analysis. Microglia are separated into 12 clusters. Model-based analysis of single cell transcriptomics (MAST) was used to compute differentially expressed genes for each cluster compared to the rest of the clusters. Genes with an adjusted p value of 0.05 or below and an absolute log2 fold change of 0.5 or above is annotated as significant.



Chi-squared test was used test for any disproportional distribution of cell numbers in each cluster. Comparing Arf1\_KO to WT sample, cluster4, and 5 are enriched while 1, 6, 7, 9, 10, 12 are depleted. Comparing Arf1\_KO to WT/Arf1\_IFNg\_KO and IFNg\_KO, only cluster5 is enriched while cluster1, 6, 10 and 12 are depleted. (Table: enrichment\_depletion\_of\_cells\_in\_Arf1\_chi-squared.csv)





Since cluster1,2,3,4 and 5 accounts for 83% of the total microglia and there five are clustered together, we examined the proportion of cells in each of these clusters. In particular we found the relative fraction of Arf1\_KO cells goes up from cluster1 to cluster5 compared to WT.

# In order to access the gene expression patterns for each cluster and identify clusters that might represent potential neurodegenerative and inflammatory state, we compared the differentially expressed genes in each cluster to those previously published in other studies. Those are Experimental autoimmune encephalomyelitis (EAE) model for multiple sclerosis (Kraseman et al., 2017), Mutant [SOD1](https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/sod1) [Microglia](https://www.sciencedirect.com/topics/immunology-and-microbiology/microglia) RNA-Seq Data Relative to Control for Amyotrophic lateral sclerosis (Chiu et al., 2013), homeostatic [Microglia](https://www.sciencedirect.com/topics/immunology-and-microbiology/microglia) to DAM for Alzheimer’s disease (Keren-Shaul et al. 2017) and 24 month to 4 month mouse for aging (Holtman et al., 2015). Signed value is calculated for each gene as sign(logFC) \* log10(Pvalue) \* abs(logFC). The signed value for Arf1-WT genes and cluster5-clusterAll are compared to the genes in the studies mentioned above.

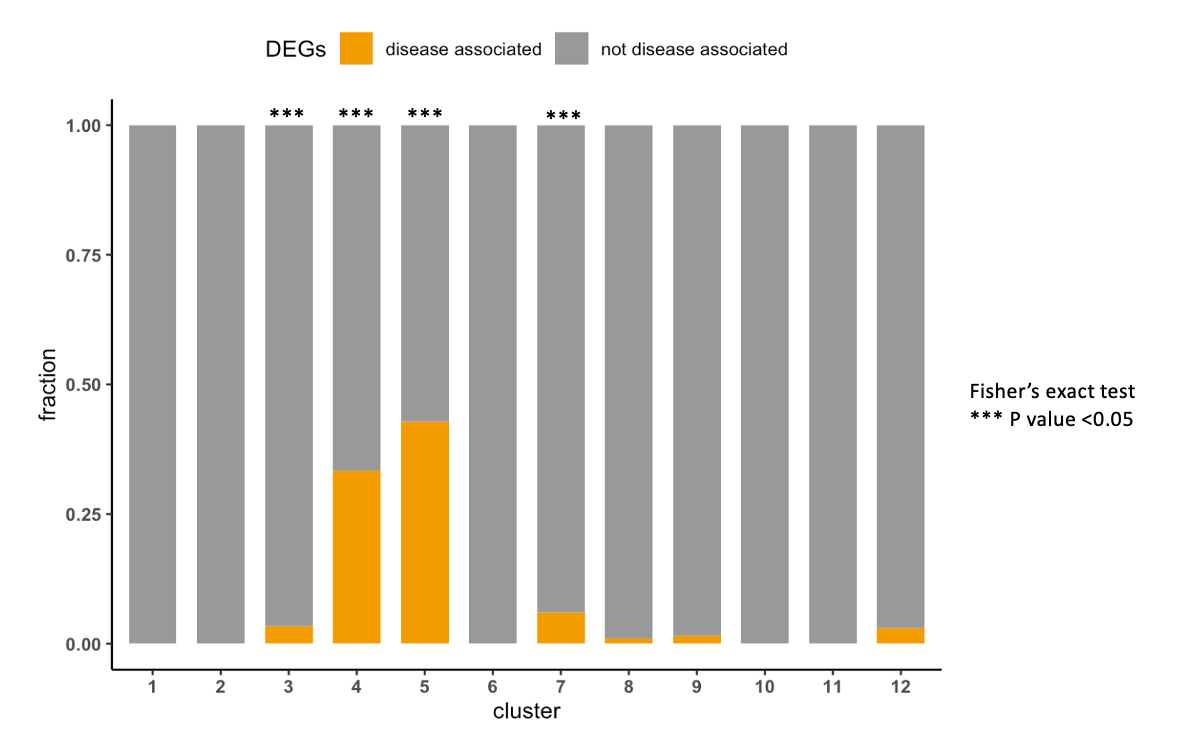
# 

# Genes that fall on the top right quadrant indicates concordant up-regulation with previously identified disease state genes. Genes that fall on the bottom left quadrant indicates concordant down-regulation with previously identified disease state genes. DEGs from both Arf1-WT contrast and Cluster5-ClusterAll contrast showed a high number of genes that showed similar signed value signatures as those disease state genes.

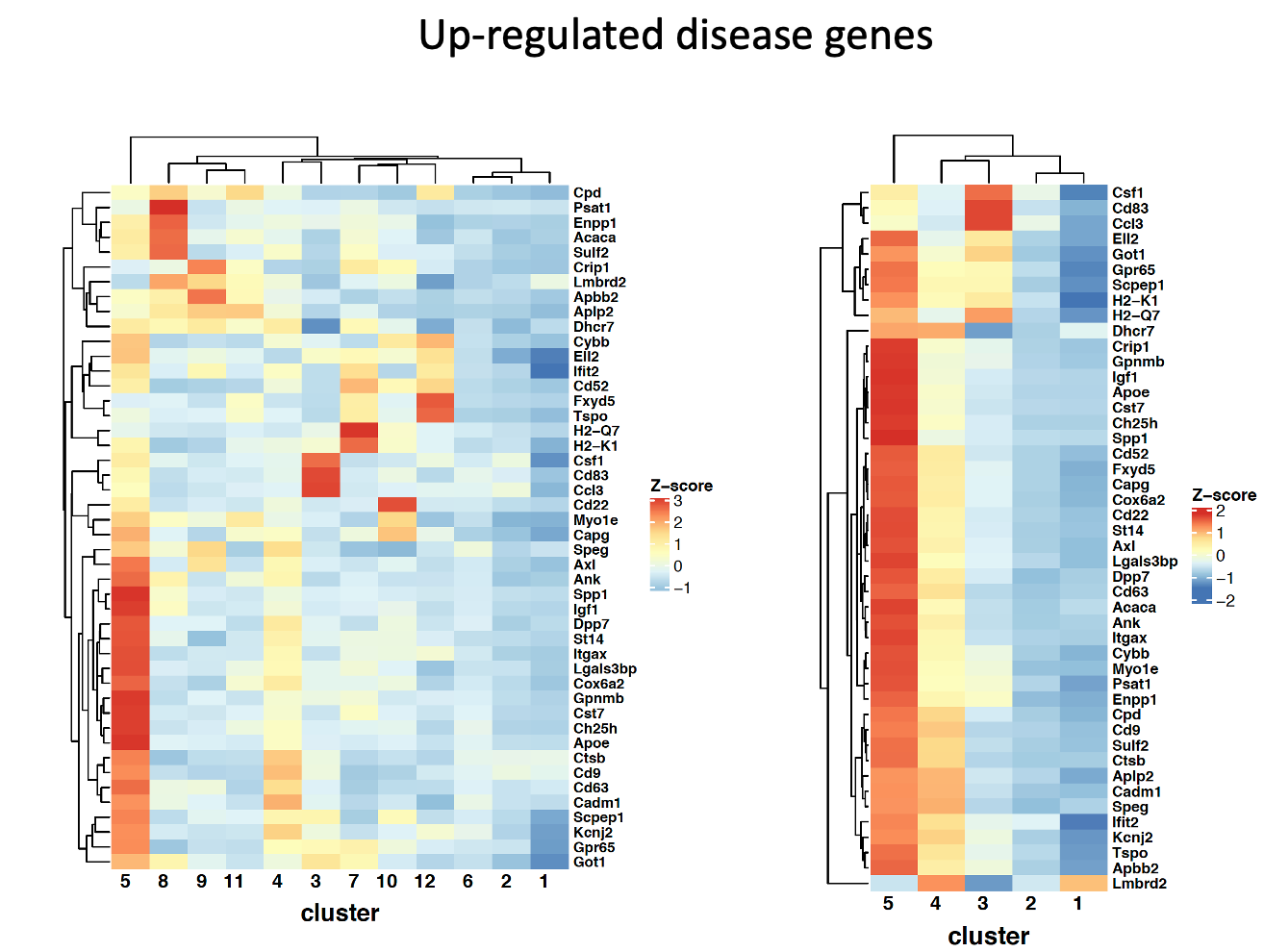
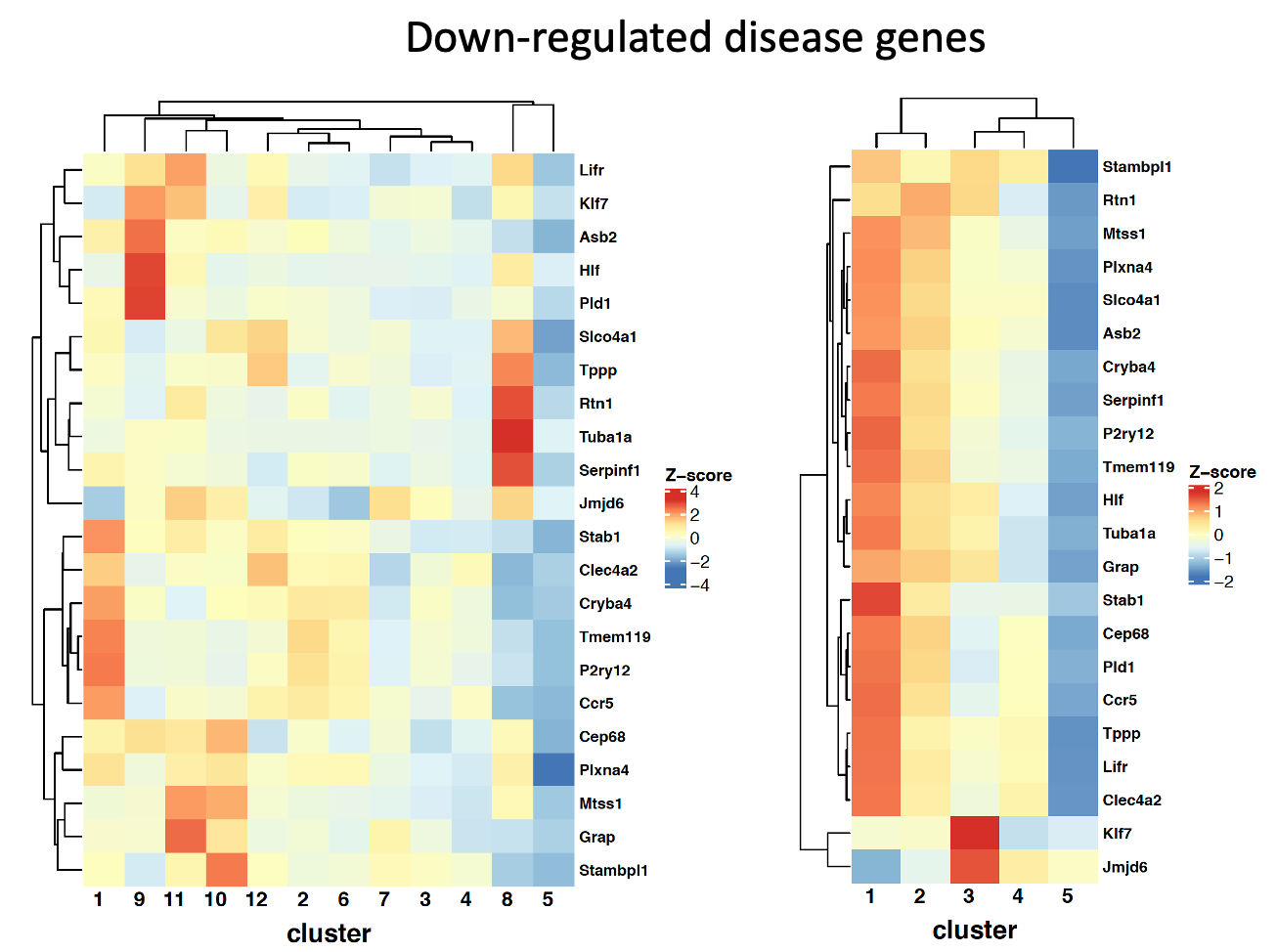
# 

# 

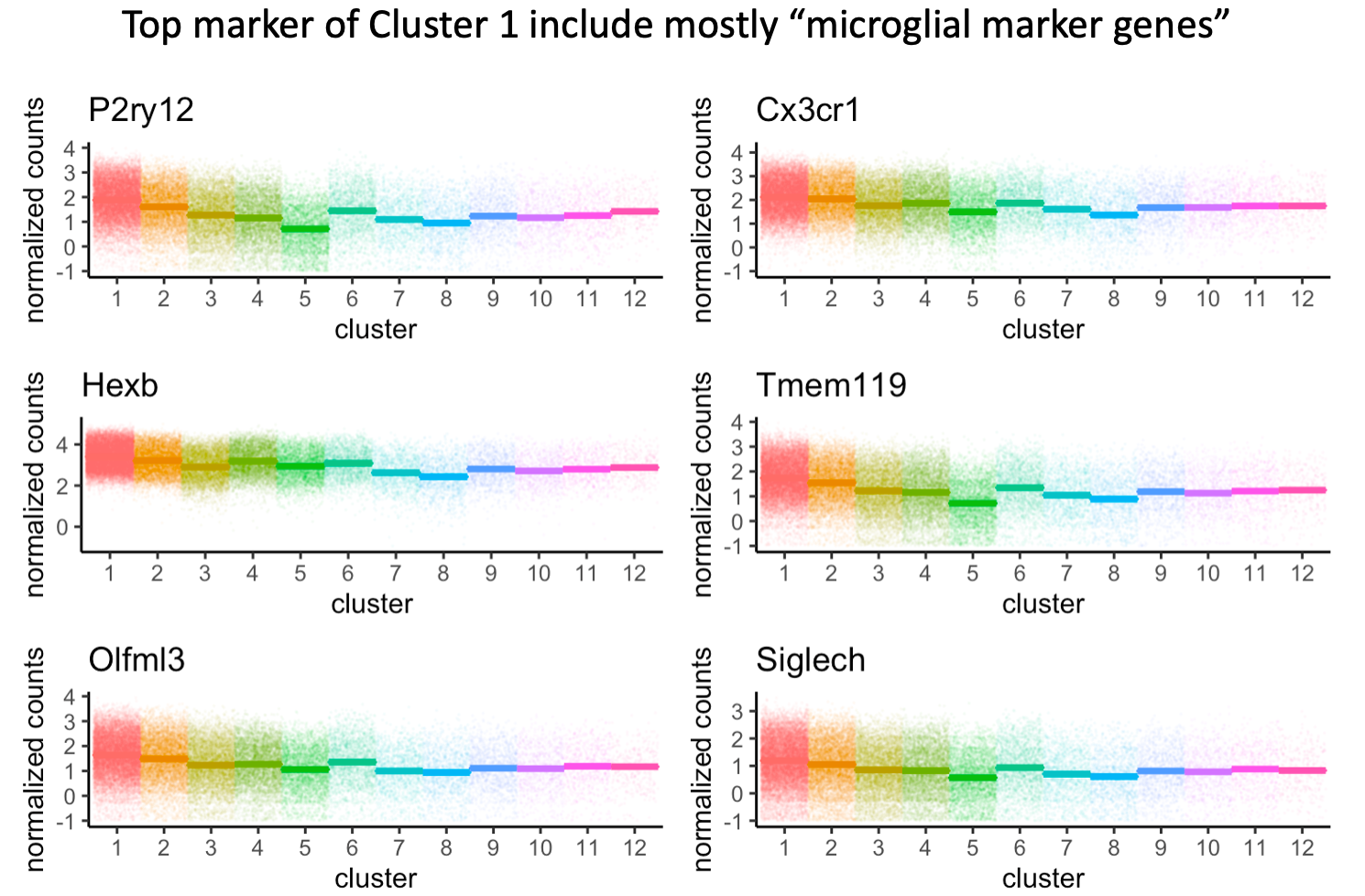
To systematically test the relative proportion of disease genes in each cluster, we obtained commonly differentially expressed dysregulated genes in the ALS and AD study. There are 46 up-regulated genes that include some well-known disease risk factor such as Apoe, Spp1, Cd63. Similarly there are 22 down-regulated genes that include gene such as P2ry12, Tmem119 etc. We asked what fraction of these up-regulated and down-regulated genes are also concordantly up-regulated and down-regulated in the differentially expressed genes for each microglia cluster compared to rest of the clusters. Fisher’s exact test showed a significant over-representation of genes associated with diseases for cluster3, 4, 5, and 7. Among these cluster 4 and 5 have very high fractions at 0.33 and 0.43.



Next we used heatmap to show cluster average normalized expression for these 46 up-regulated and 22 down-regulated disease genes. Examining the up-regulated disease genes, Cluster5 form a single group with the many of genes showing a large positive Z score which indicates also up-regulation. Similarly in down-regulated disease gene set, cluster5 is grouped with cluster8 appears to have most low Z scores indicating low expression of these genes.

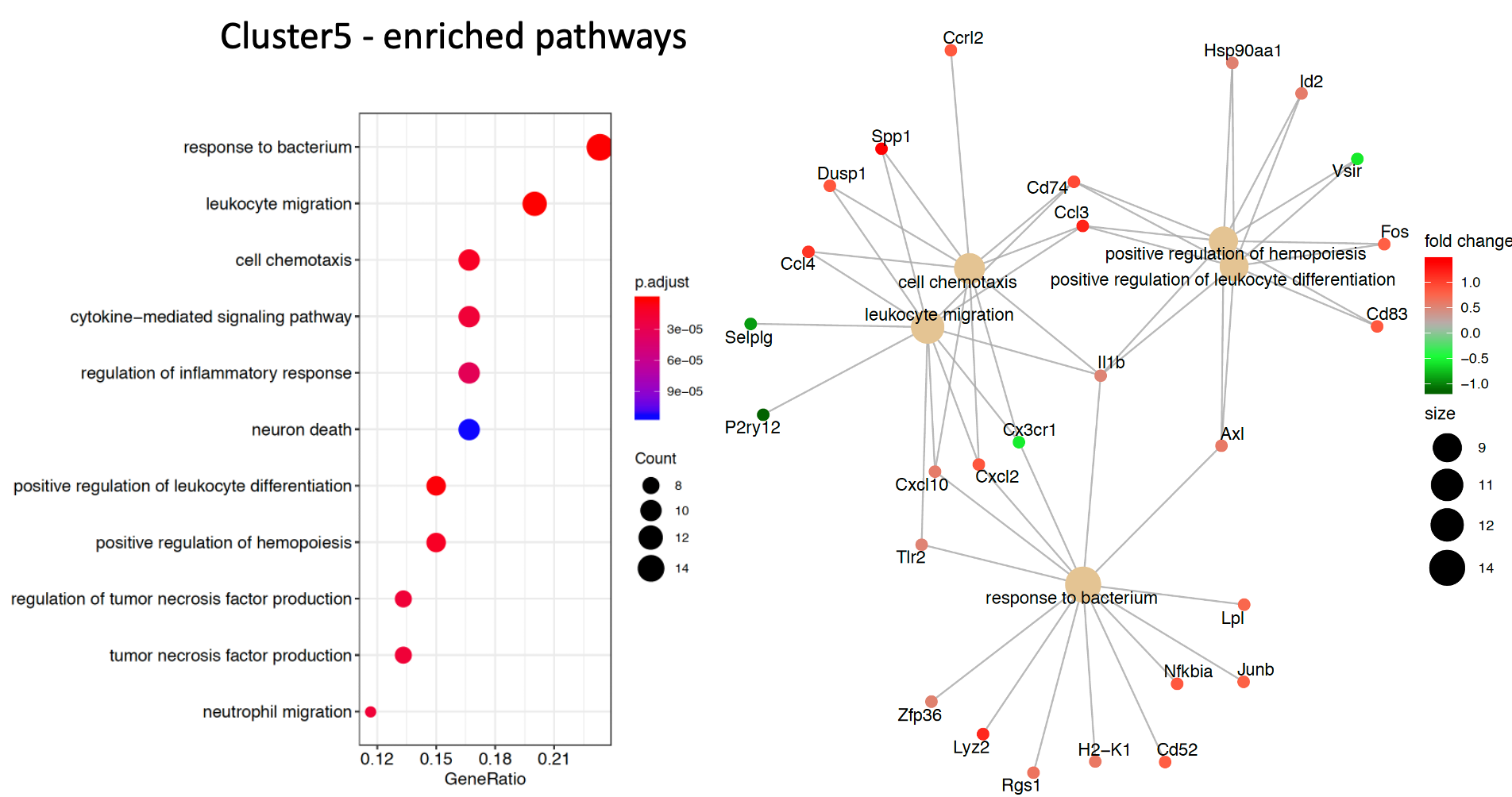
 

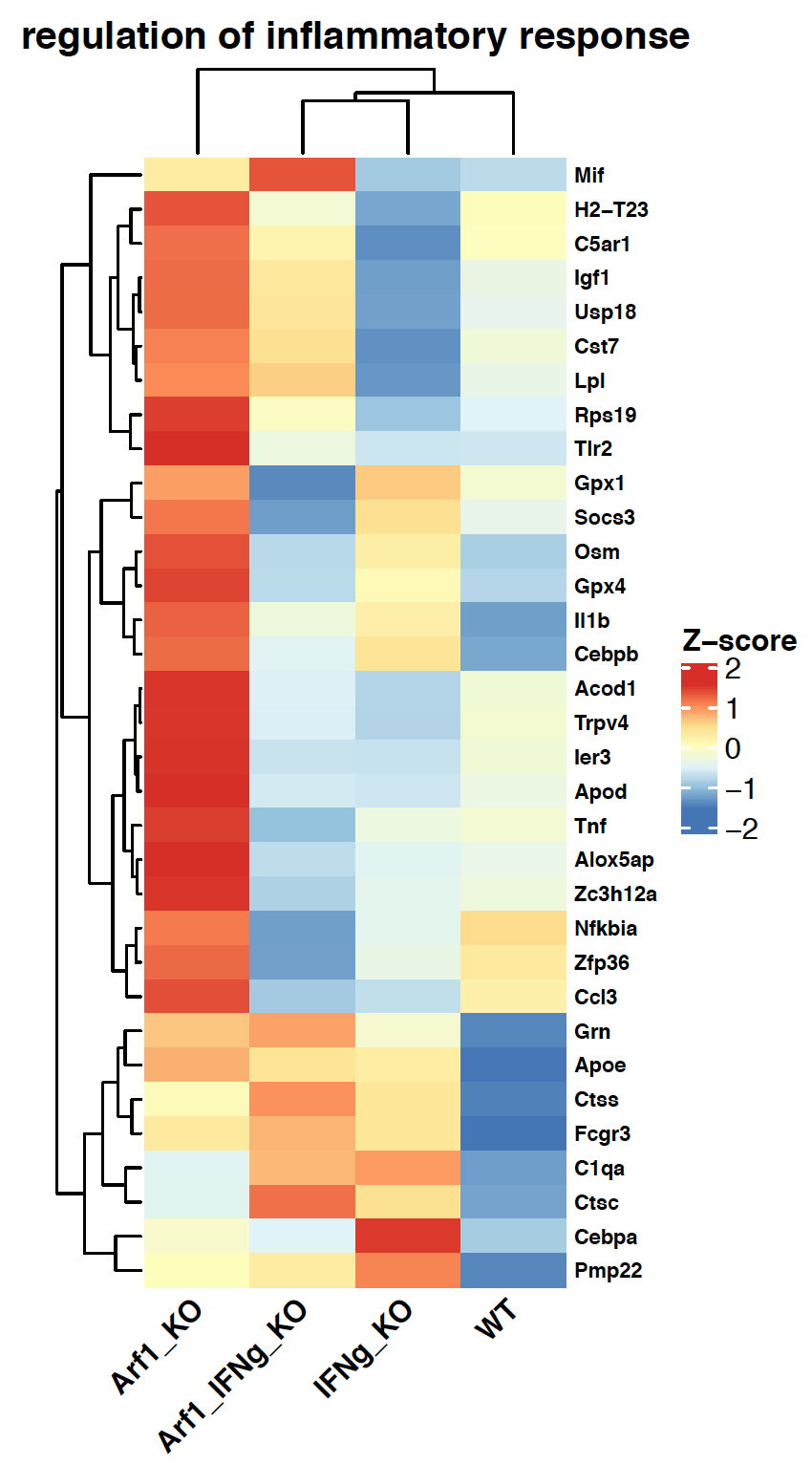
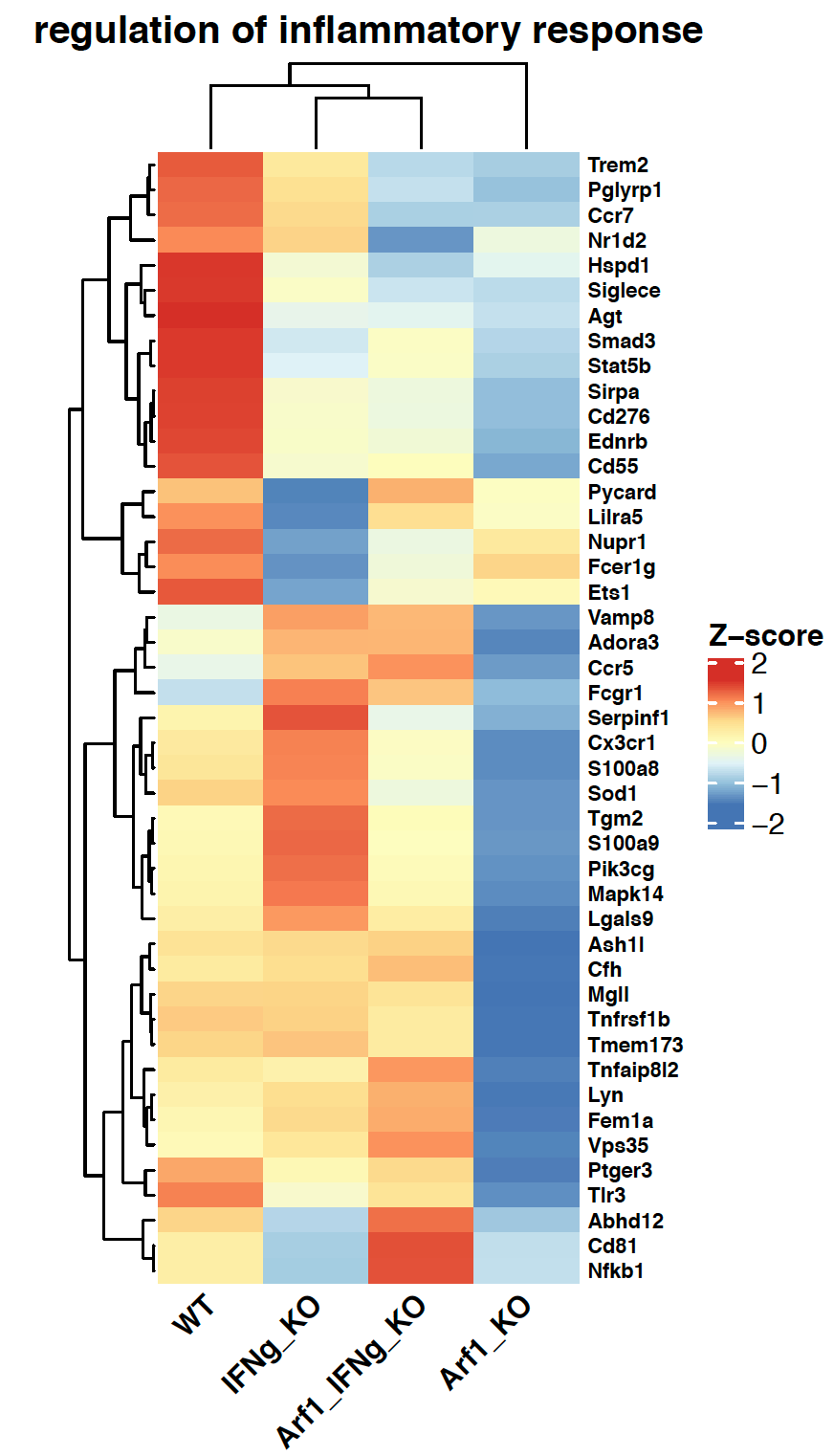
Next we examined the expression of some of well-known microglia markers for all clusters, including the [purinergic receptors](https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/purinergic-receptor) P2ry12/P2ry13, Cx3cr1, and Tmem119 (Butovsky et al., 2015, Haynes et al., 2006, Hickman et al., 2013, Merino et al., 2016, Mildner et al., 2017). Microglia cells that have stable expression of these genes are presumably at a normal and homeostatic state. Each dot on the plot represent the normalized expression level of a single cell. Cluster1 (and cluster2) have stable(high) expression of these genes compared to cluster5. These genes are also predicted as top markers for cluster1.



Taken together, Cluster1 and 2 represent showed the transcriptomic signatures of homeostatic state of microglia cells while cluster4 and cluster5 showed transcriptomic patterns of inflammatory and neurodegenerative patterns of microglia.

Next we computed Cluster5-cluster1 differentially expressed genes and performed the over-representation analysis using the Gene Ontology – Biological Process gene set. The top enriched processes include response to bacterium, leukocyte migration, chemotaxis, inflammatory response etc. Genes involved in these significant terms are plotted as networks. Similarly we also performed ORA on the Arf1-WT at the sample level and found top enriched terms include positive regulation of cell migration, chemotaxis, phagocytosis etc. The genes for involved in several of these selected processes were also plotted as heatmap to show cell average expression across samples.





As dysregulated genes in cluster5-cluster1 is very similar to Arf1-WT (volcano plot) and they have many overlapping over-represented top pathways, such as cell chemotaxis, leukocyte migration, leukocyte differentiation, these indicates cluster 5 is representative of the most dysregulated cells in response to the Arf1 KO.

