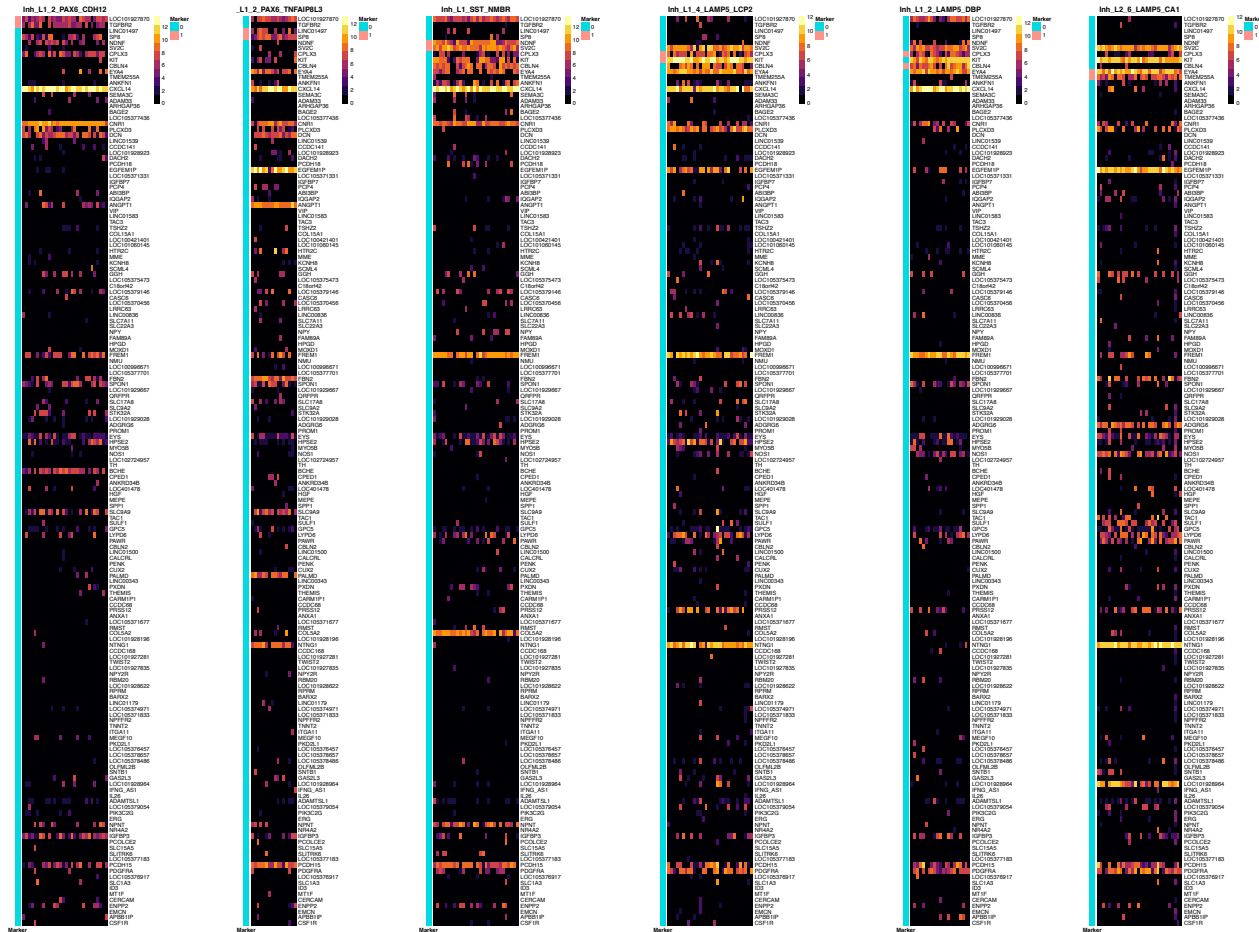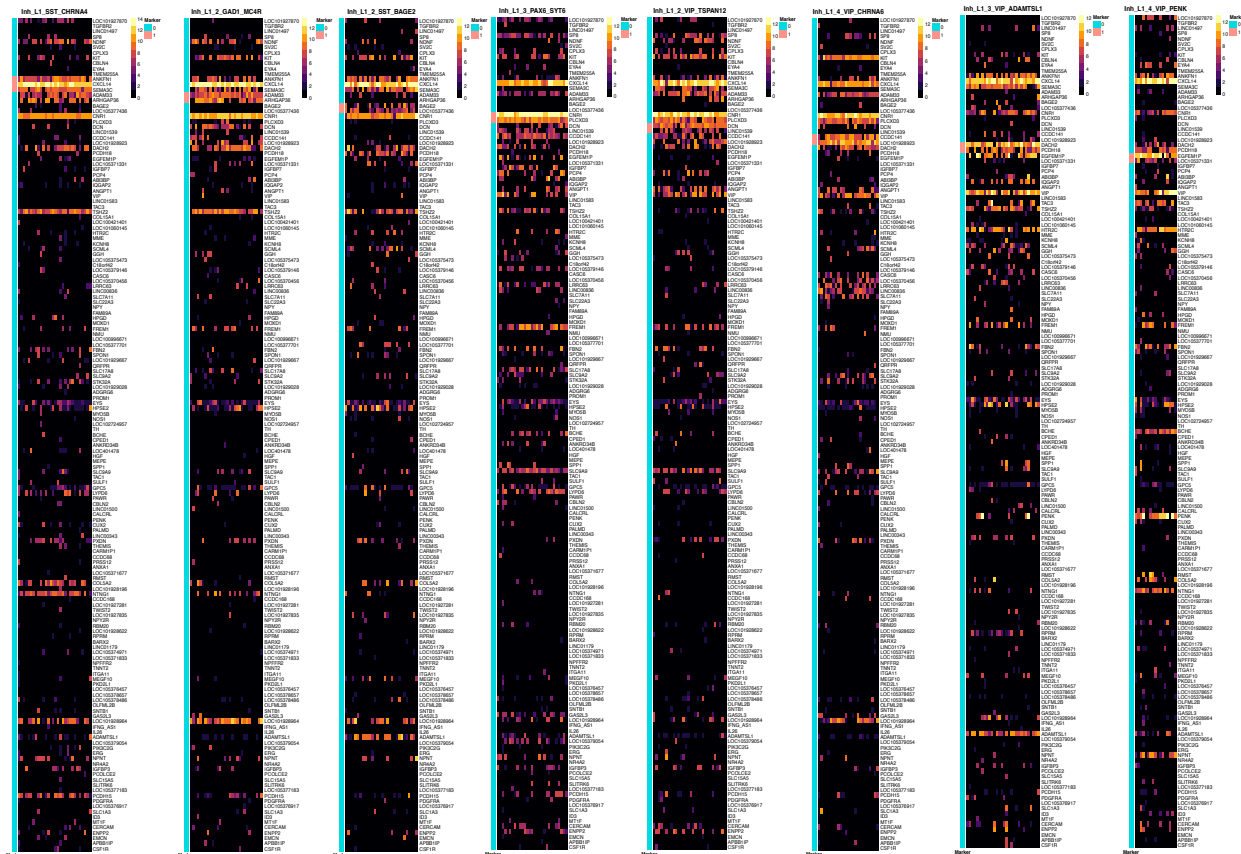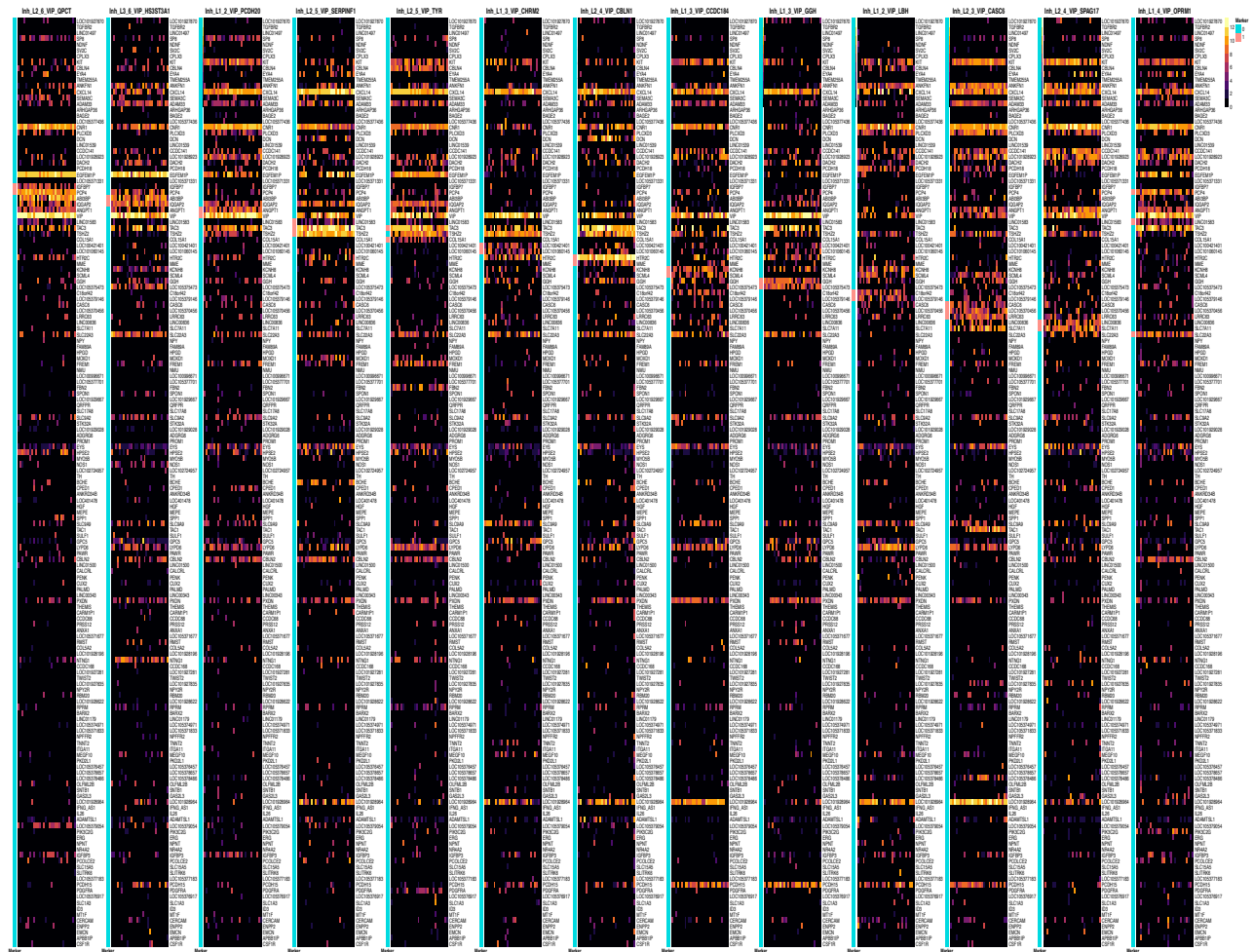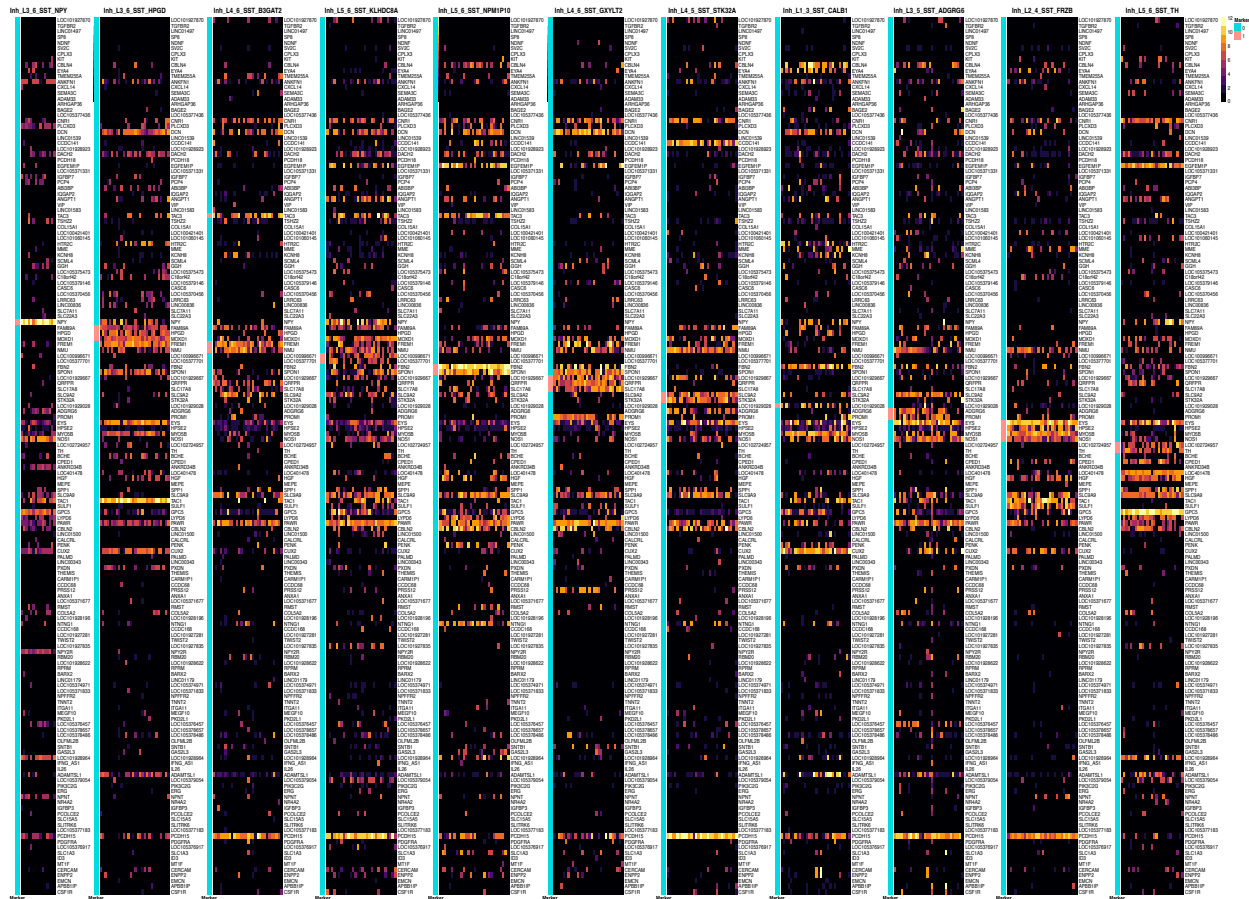# Supplemental Figures

Supplemental Fig S1: NS-Forest v2.0 barcodes for human MTG *LAMP5/PAX6* branch. Cell type clusters were grouped based on upper-level branch nodes in the hierarchical cluster dendrogram shown in Figure 3. Log2 CPMs are plotted as a heatmap with the 157 marker genes as rows and individual nuclei as columns.

Supplemental Fig S2: NS-Forest v2.0 barcodes for human MTG *VIP* left branch. Cell type clusters were grouped based on upper-level branch nodes in the hierarchical cluster dendrogram shown in Figure 3. Log2 CPMs are plotted as a heatmap with the 157 marker genes as rows and individual nuclei as columns.

Supplemental Fig S3: NS-Forest v2.0 barcodes for human MTG *VIP* right branch.
Cell type clusters were grouped based on upper-level branch nodes in the hierarchical cluster dendrogram shown in Figure 3. Log2 CPMs are plotted as a heatmap with the 157 marker genes as rows and individual nuclei as columns.
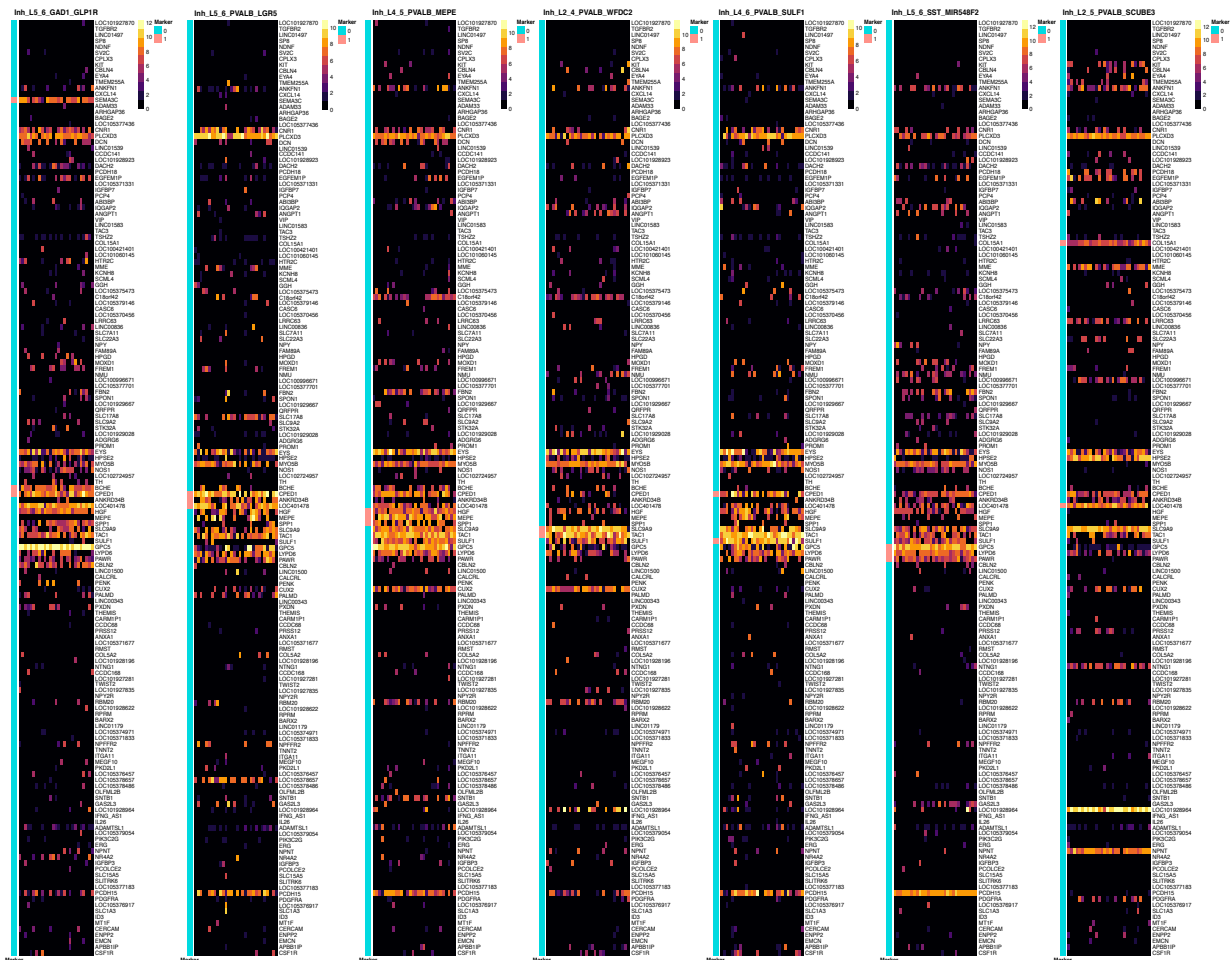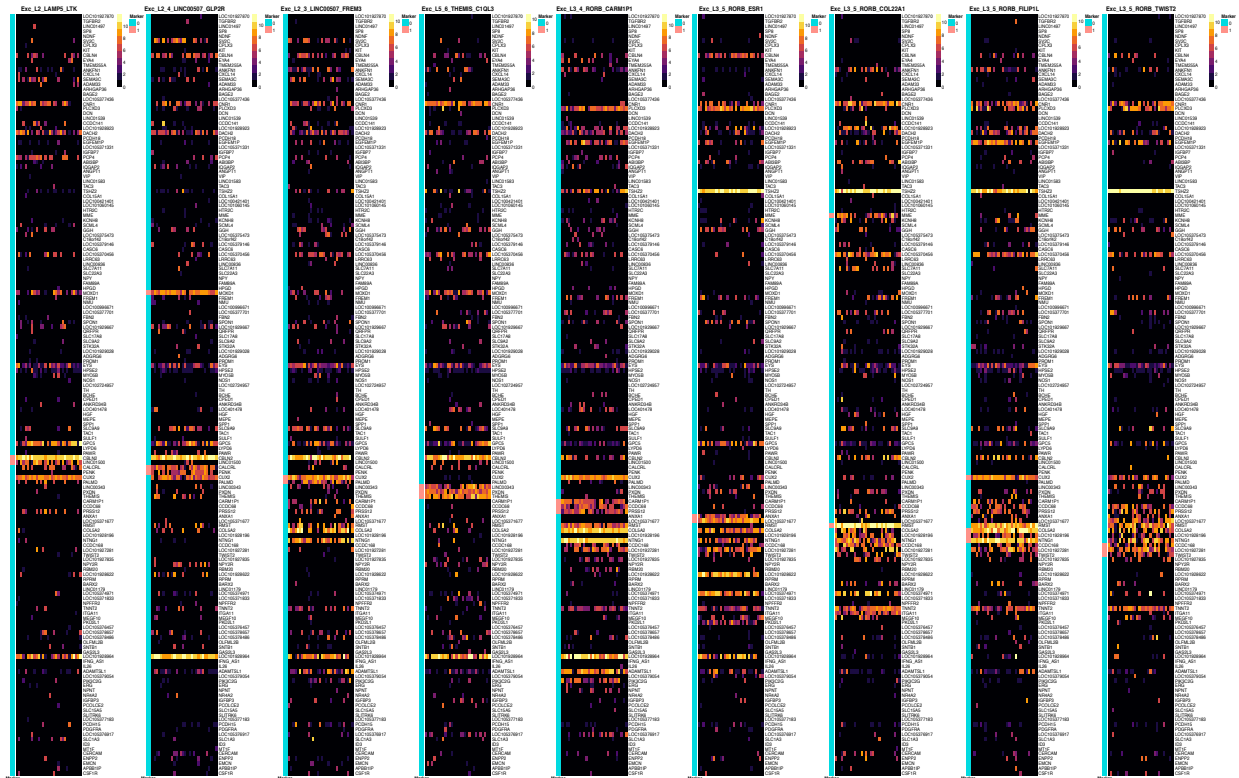
Supplemental Fig S4: NS-Forest v2.0 barcodes for human MTG *SST* branch. Cell type clusters were grouped based on upper-level branch nodes in the hierarchical cluster dendrogram shown in Figure 3. Log2 CPMs are plotted as a heatmap with the 157 marker genes as rows and individual nuclei as columns.

Supplemental Fig S5: NS-Forest v2.0 barcodes for human MTG *PVALB* branch. Cell type clusters were grouped based on upper-level branch nodes in the hierarchical cluster dendrogram shown in Figure 3. Log2 CPMs are plotted as a heatmap with the 157 marker genes as rows and individual nuclei as columns.

Supplemental Fig S6: NS-Forest v2.0 barcodes for human MTG *RORB* left branch.
Cell type clusters were grouped based on upper-level branch nodes in the hierarchical cluster dendrogram shown in Figure 3. Log2 CPMs are plotted as a heatmap with the 157 marker genes as rows and individual nuclei as columns.
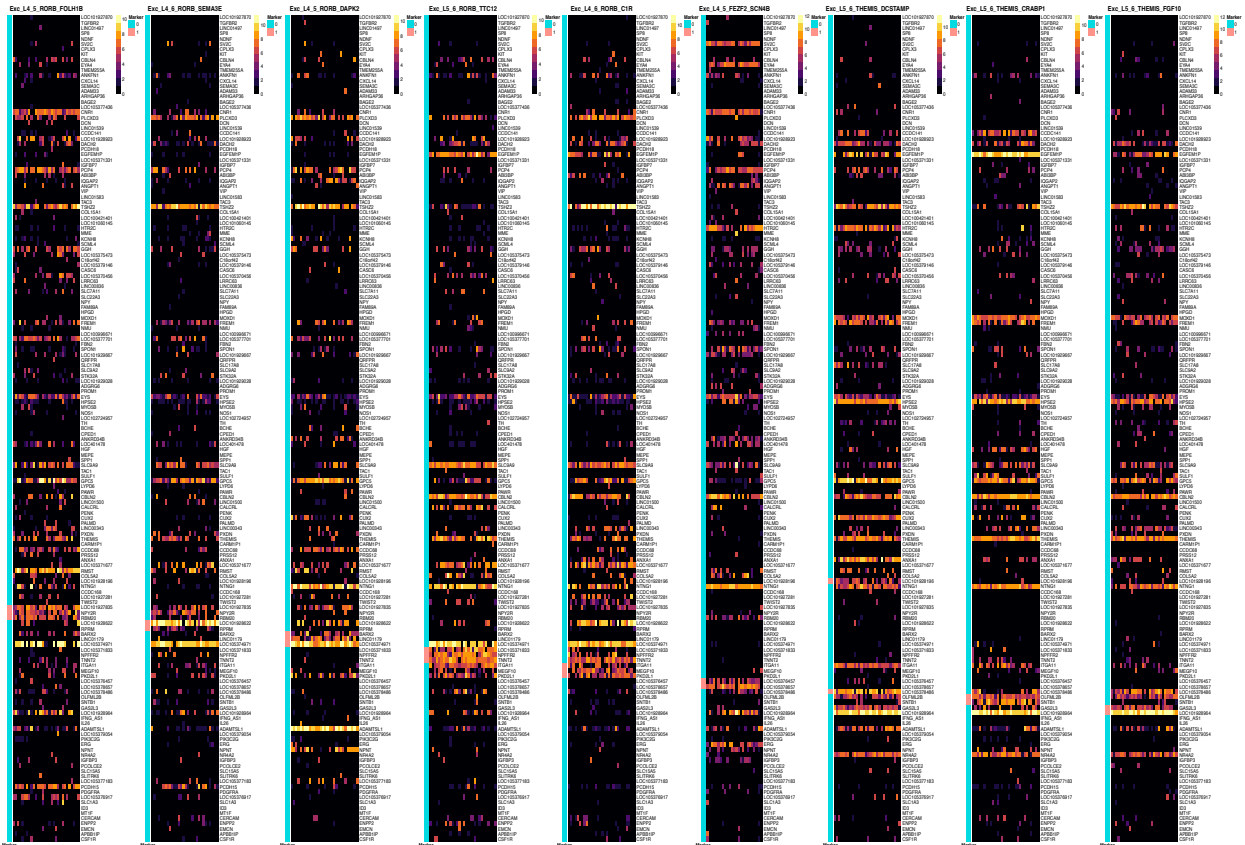
Supplemental Fig S7: NS-Forest v2.0 barcodes for human MTG *RORB* right branch. Cell type clusters were grouped based on upper-level branch nodes in the hierarchical cluster dendrogram shown in Figure 3. Log2 CPMs are plotted as a heatmap with the 157 marker genes as rows and individual nuclei as columns.

Supplemental Fig S8: NS-Forest v2.0 barcodes for human MTG *FEZF2* branch. Cell type clusters were grouped based on upper-level branch nodes in the hierarchical cluster dendrogram shown in Figure 3. Log2 CPMs are plotted as a heatmap with the 157 marker genes as rows and individual nuclei as columns.

Supplemental Fig S9: NS-Forest v2.0 barcodes for human MTG non-neuronal branch. Cell type clusters were grouped based on upper-level branch nodes in the hierarchical cluster dendrogram shown in Figure 3. Log2 CPMs are plotted as a heatmap with the 157 marker genes as rows and individual nuclei as columns.
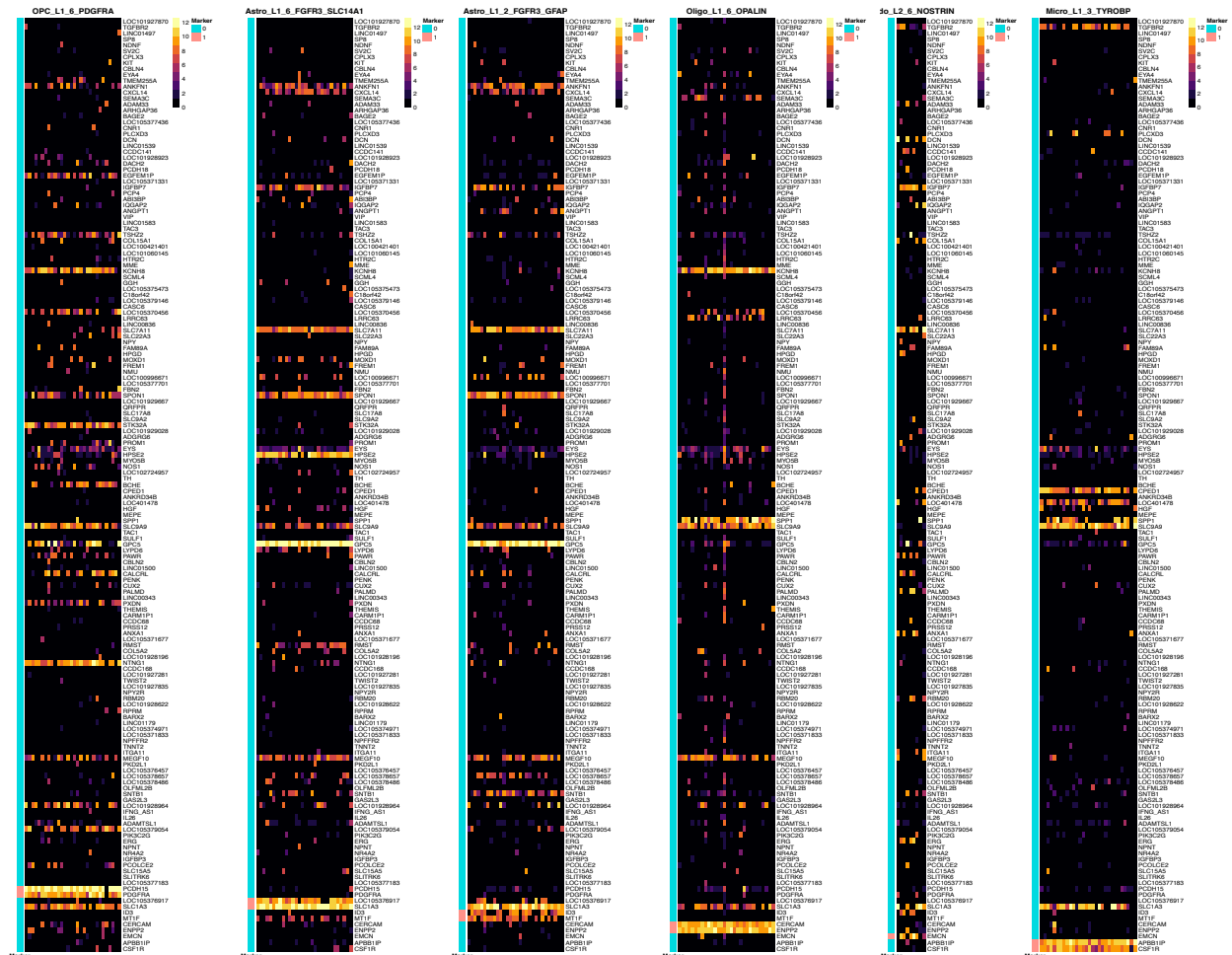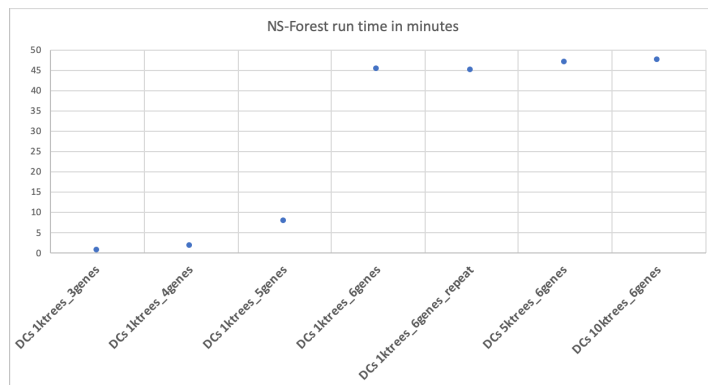
Complete Clusters	NS-Forest v2.0	COMET	RANKCORR
Cluster_0	CD14, CSF3R	ITGAM, CD44	CD14, VCAN, S100A12, S100A8
Cluster_1	CD1C, FCER1A	S100A9, CD74	CD1C, FCER1A, CLEC10A
Cluster_2	TMEM176B, FCGR3A	IFITM3, ICAM4	LOC200772, AK307192, TCF7L2, FCGR3A,
Cluster_3	GZMB	SLAMF7, CTSB	DERL3, MZB1, TSPAN13
Cluster_4	CLEC9A, C1ORF54	FLT3, ECE1	CADM1, CLEC9A, C1ORF54
Cluster_5	LAIR2, FCGR3A, IFI30, F	CSF1R, CTSB	LAIR2, IFI30, FTL, PSAP
Cluster_6	SIGLEC6, AXL	EPHB4, CXCR2	HAMP, GPR146, PPP1R14A
Cluster_7	PRF1, GNLY	IFITM1, CCND2	FGFBP2, KLRC4_KLRK1, IL2RB
Dendritic Only Clusters	NS-Forest v2.0	COMET	RANKCORR
Clusters_0	FCGR3A, SERPINA1	FCER1G, IFITM3	TCF7L2, AK307192, MTSS1
Clusters_1	GZMB	NRP1, SELL	DERL3, AK128525, MZB1
Clusters_2	CLEC9A	FLT3, IDO1	CADM1, CLEC9A, C1ORF54
Clusters_3	CD1C, FCER1A	FCER1A, CD33	CD1C, FCER1A, CLEC10A
Clusters_4	CD14, S100A9	CD14, S100A8	CD14, RNASE2, S100A8
Clusters_5	AXL	CD22, AXL	BDKRB2, HAMP, PPP1R14A, SIGLEC6

Supplemental Figure S10: Marker genes from NS-Forest, COMET, and RANKCORR. A) markers for the clustering solution for all monocyte and dendritic cells. B) markers for the clustering solution for dendritic cells only.
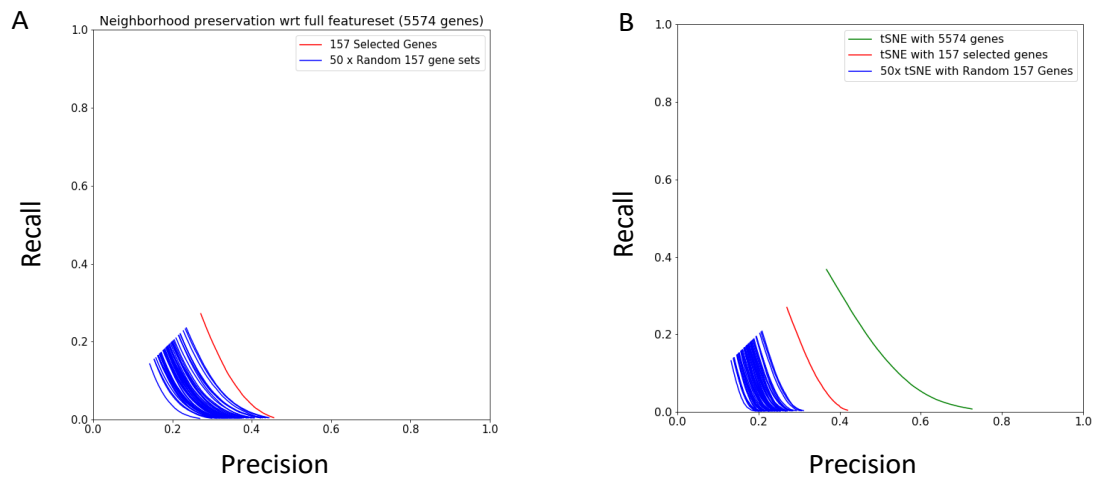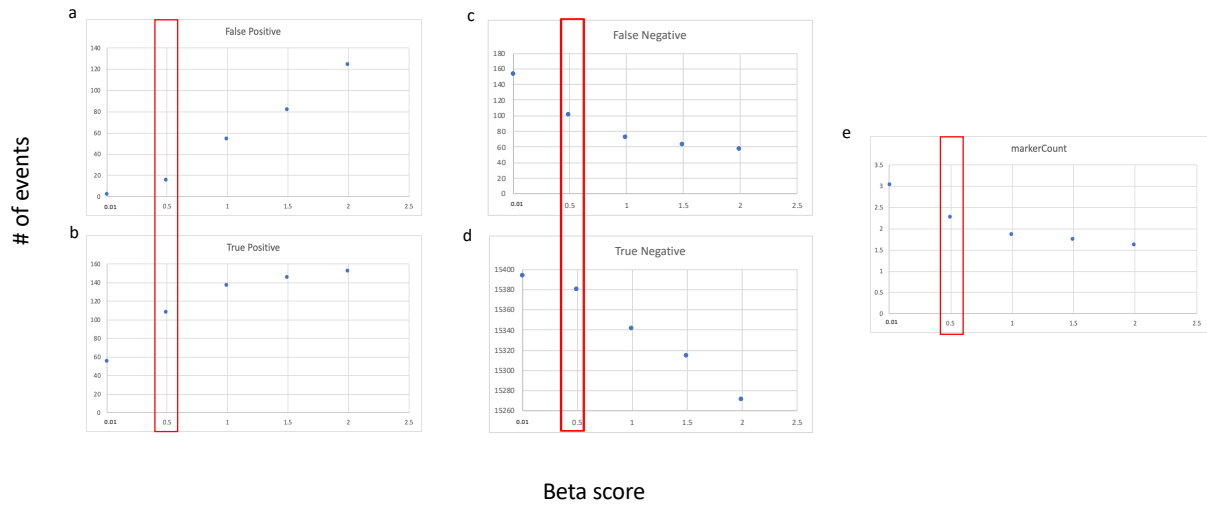
A

B

Cluster	Default 10k trees	DCs 5k trees	DCs 1k trees	DCs 1k trees 5genes	DCs 1k trees 4genes	DCs 1k trees 3genes
Clusters_0	0.99	0.99	0.99	0.99	0.99	0.99
Clusters_1	1.00	1.00	1.00	1.00	1.00	1.00
Clusters_2	0.98	0.98	0.98	0.98	0.98	0.98
Clusters_3	0.76	0.76	0.76	0.76	0.75	0.75
Clusters_4	0.87	0.87	0.87	0.86	0.85	0.84
Clusters_5	0.92	0.92	0.92	0.92	0.77	0.76

Supplemental Figure S11: Bench marking results for NS-Forest v2.0 using the dendritic cell subset from Villani et al, 2017. A) Varying the number of trees or the number of top genes for permutation. B) Resulting F-beta scores from varied parameters in panel A.

Supplemental Fig S12: Quantitative assessment of Nearest Neighbor Preservation metric - A) Nearest Neighbor Preservation (NNP) curves for the random-forest selected 157 genes (red) for 50 random gene sets of the same size (selected from the full 5574 high-variance gene set)(blue) are shown. B) NNP curves for t-SNEs generated from the complete 5574 variable genes (green), the random-forest selected 157 genes (red), or for 50 random gene sets of the same size (selected from the full 5574 high-variance gene set) (blue) are shown.

Supplemental Fig S13: Empirical determination of the beta weighting parameter for the F-beta score – For beta values of 0.01, 0.5, 1, 1.5 and 2 (x-axis), the average true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values, and number of markers were calculated for all 75 clusters from the MTG data.