Annotation of cells to broad cell type categories:

Each cell was annotated as a cell type using the final "usage" matrix output of cNMF or the calculated usage matrices as discussed above. The usage scores were normalized to 100% for each cell. Programs were then put into broad cell type categories as follows:

a. Myeloid - Microglia, CD14_mono, DC, Neutrophils
b. T cells
c. Vascular - Pericytes, Endothelial
d. Oligo
e. Malignant - AC, NPC1/OPC, MES2, MES1, Vascular-MES1, NPC2
i. Giant Cell Malignant was only added for MGH720
f. Cycling
g. Other_Immune - Mast, Plasma

For each cell, the usage scores for all programs in each category were summed to create a usage score for the cell type category. For example, the usage scores for 4 myeloid programs were summed to create the "myeloid usage" per cell. Cells were then annotated as one of the cell types using the top-scoring usage for the cell type category.

cNMF determined doublets:

1. Summed usage scores per category were used with the annotations of each cell.
    a. Cells were considered Doublets by cNMF if two programs scored above a specific usage score. This threshold was cell type x cell type-specific. Thresholds were created by comparing background usage between cell types and identifying clear thresholds:

	Secondary programs					
Annotation	Malignant	Myeloid	Stromal	T cells	Oligo	Other_Immune
Malignant		30%	25%	25%	20%	20%
Myeloid	20%		20%	30%	20%	20%
Stromal	20%	20%		20%	20%	20%
T cells	20%	30%	20%		20%	20%
Oligo	40%	30%	20%	20%		20%
Other_Immune	30%	30%	30%	40%	30%	

b. The Cycling programs were not considered in doublet analysis as Malignant and Non-Malignant cells in a tumor can use this program.
c. The Giant Cell Program was only used when considering GBM Giant Cell tumors.

inferCNV determined doublets:

1. Cells that were annotated as Malignant, Non-Malignant, or Undetermined by inferCNV.
    a. inferCNV annotations were determined as described above, using specific thresholds of CNV detection scores for tumor-specific CNVs. Cells from tumors with no detectable CNVs were annotated as Undetermined.
2. cNMF was used to annotate cells as “Malignant” if the cNMF annotation was “Malignant”. It was also labeled “Malignant” if it was annotated as “Cycling” AND had a combined malignant usage score of greater than 25% or if it was annotated as “Cycling” AND was considered “Malignant” by inferCNV. Cells that did not meet one of these criteria were annotated as “Non-Malignant”
    a. Cycling usage programs can dominate a cell transcriptome, but different cell types use the same cell cycle program (e.g. Myeloid, T cell, Malignant cell types). Therefore, special care needs to be taken with cells annotated as Cycling.
3. Cells were considered Doublets by inferCNV if their annotation by inferCNV did not match that of cNMF. Undetermined cells by inferCNV were not considered for doublets by this method.

Integration of cNMF and inferCNV data to create final Doublet list:

Cells considered “Doublets” by cNMF and/or inferCNV were used to create a Doublet list for each tumor cohort dataset.