# Cell QC and Broad Cell Types

## inputs

In [1]:
```
source('jupyterFunctions_broadCellType.R')
source('jupyterFunctions_perCellType.R')
```

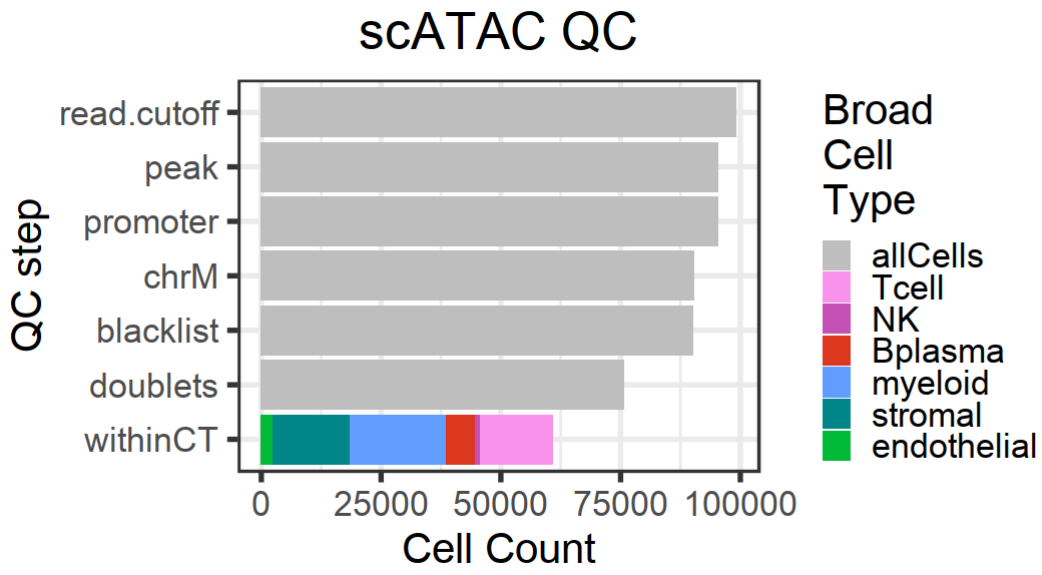
In [2]:
```
CT <- 'broadCT'
CT_label <- 'Broad Cell Type'
data_prefix <- paste(sep='','../data/',CT,'/')
scATAC_meta <- readRDS(paste(sep='',data_prefix,'scATAC_meta.rds'))
multiome_meta <- readRDS(paste(sep='',data_prefix,'multiome_bothQC_meta.
rds'))
scATAC_cellCount <- readRDS(paste(sep='',data_prefix,'scATAC_cellQC_cell
Counts.rds'))
snATAC_cellCount <- readRDS(paste(sep='',data_prefix,'snATAC_cellQC_cell
Counts.rds'))
snRNA_cellCount <- readRDS(paste(sep='',data_prefix,'snRNA_cellQC_cellCo
unts.rds'))
peakComparison_df <- readRDS(paste(sep='',data_prefix,'peakComparison_d
f.rds'))
chosenPeaks <- readRDS(paste(sep='',data_prefix,'broadCT_chosenPeaks.rd
s'))
scATAC_pxc_norm <- readRDS(paste(sep='',data_prefix,'scATAC_pxc_norm.rd
s'))
snATAC_pxc_norm <- readRDS(paste(sep='',data_prefix,'snATAC_pxc_norm.rd
s'))
snRNA_gxc_norm <- readRDS(paste(sep='',data_prefix,'snRNA_gxc_norm.rd
s'))
scATAC_pxCT_norm <- readRDS(paste(sep='',data_prefix,'scATAC_pxCT_norm.r
ds'))
snATAC_pxCT_norm <- readRDS(paste(sep='',data_prefix,'snATAC_pxCT_norm.r
ds'))
snRNA_gxCT_norm <- readRDS(paste(sep='',data_prefix,'snRNA_gxCT_norm.rd
s'))
```

In [3]:
```
broadCT_colors <- readRDS('../data/misc/broadCT_colors.rds')
broadCT_order <- c('Tcell','NK','Bplasma','myeloid','stromal','endotheli
al')
```

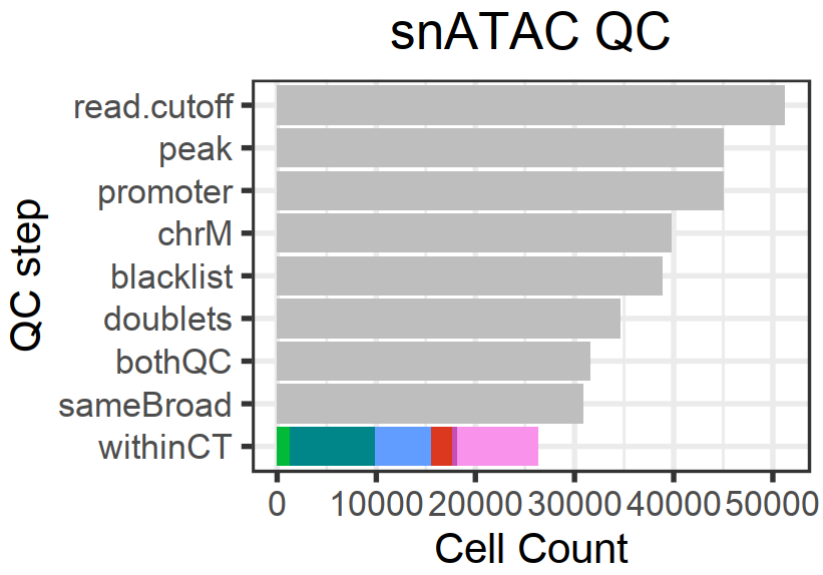
In [22]:
```
save_dir <- NA #'../output/' #or NA if don't want to save
```

## QC steps

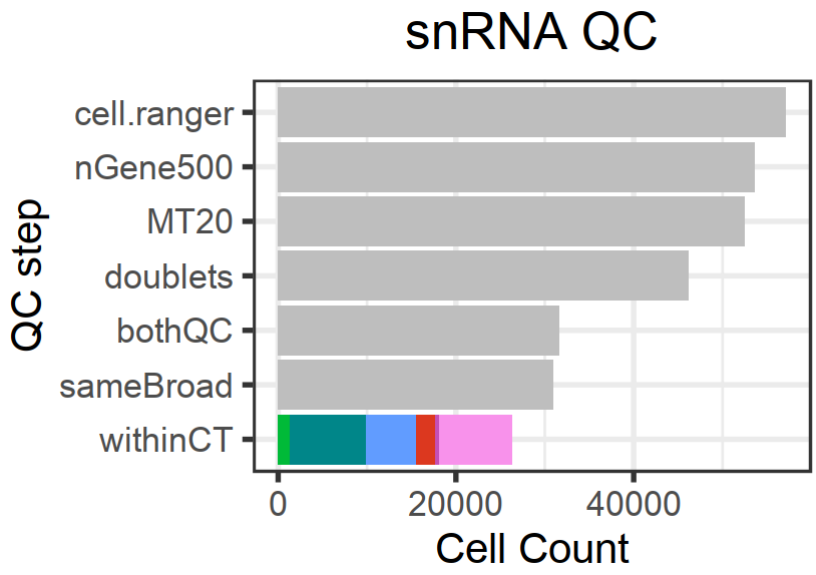
In [5]:
```
# Fig S1a

options(repr.plot.height=5,repr.plot.width=9)
g <- QC_steps_barplot(scATAC_cellCount,scATAC_meta[which(scATAC_meta$wit
hinCT_passQC_flag==TRUE),],
                      'cellType',broadCT_colors,broadCT_order,tLab='scAT
AC QC')
print(g)

if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'scATAC_QCsteps.p
ng'),
                            plot=g,units='in',height=5,width=9,dpi=600)
```



In [6]:
```
# Fig S1b

options(repr.plot.height=5,repr.plot.width=7)
g <- QC_steps_barplot(snATAC_cellCount,multiome_meta[which(multiome_meta
$withinCT_passQC_flag==TRUE),],
                      'snATAC_cellType',broadCT_colors,broadCT_order,tLa
b='snATAC QC')
g <- g + theme(legend.position="none")
print(g)

if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'snATAC_QCsteps.p
ng'),
                            plot=g,units='in',height=5,width=7,dpi=600)
```

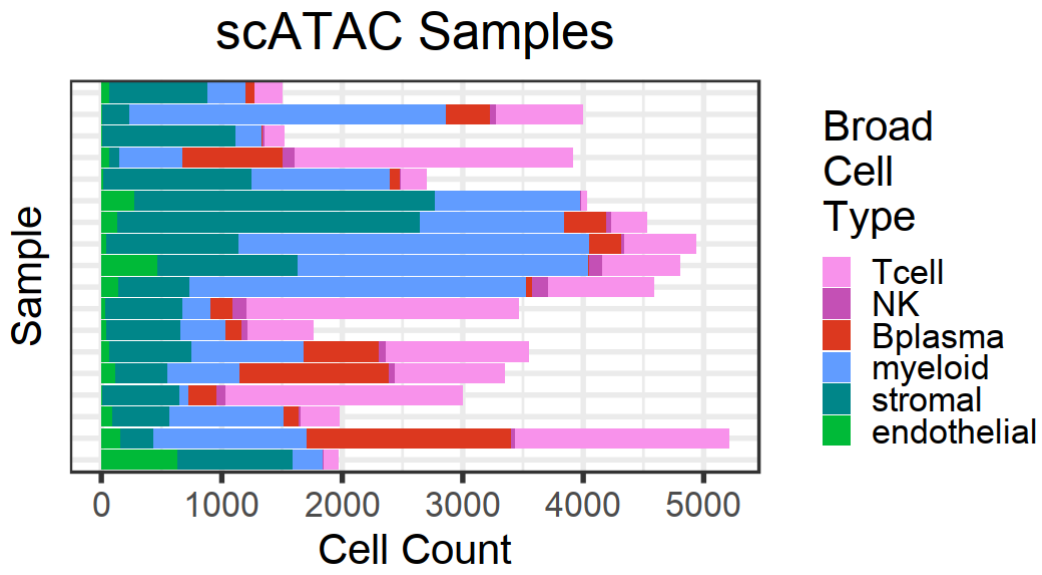


In [7]:
```
# Fig S1c

options(repr.plot.height=5,repr.plot.width=7)
g <- QC_steps_barplot(snRNA_cellCount,multiome_meta[which(multiome_meta
$withinCT_passQC_flag==TRUE),],
                      'snRNA_cellType',broadCT_colors,broadCT_order,tLab
='snRNA QC')
g <- g + theme(legend.position="none")
print(g)

if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'snRNA_QCsteps.pn
g'),
                            plot=g,units='in',height=5,width=7,dpi=600)
```



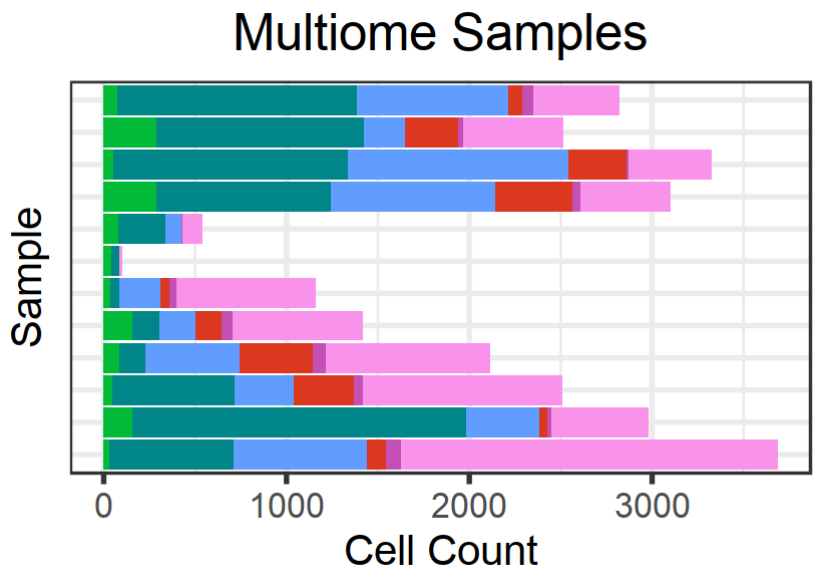
## Sample cell counts

In [8]:
```
#Fig S1d

toPlot <- as.data.frame(table(scATAC_meta[which(scATAC_meta$withinCT_pas
sQC_flag==TRUE),
                                              c('sample','cellType')]),strin
gsAsFactors=FALSE)
toPlot$sample <- factor(toPlot$sample,levels=rev(sort(unique(toPlot$samp
le))))
toPlot$cellType <- factor(toPlot$cellType,levels=c(broadCT_order))

options(repr.plot.height=5,repr.plot.width=9)
g <- ggplot(toPlot,aes_string(x='Freq',y='sample',fill='cellType')) + ge
om_bar(stat='identity',position='stack') +
        theme_bw(base_size=25) + labs(x='Cell Count',y='Sample',fill='Br
oad\nCell\nType',title='scATAC Samples') +
        theme(plot.title = element_text(hjust = 0.5)) + scale_fill_manua
l(values=broadCT_colors) +
        theme(axis.text.y=element_blank(),axis.ticks.y=element_blank())
print(g)

if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'scATAC_sample_br
oadCT_cellCounts.png'),
                            plot=g,units='in',height=5,width=9,dpi=600)
```



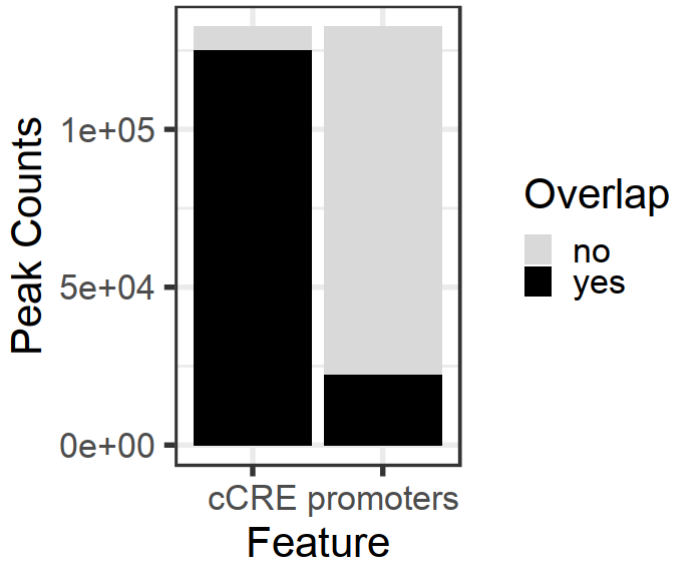
In [9]:

```
#Fig S1e

toPlot <- as.data.frame(table(multiome_meta[which(multiome_meta$withinCT
_passQC_flag==TRUE),
                                                c('sample','snATAC_cellTyp
e')]),stringsAsFactors=FALSE)
toPlot$sample <- factor(toPlot$sample,levels=rev(sort(unique(toPlot$samp
le))))
toPlot$snATAC_cellType <- factor(toPlot$snATAC_cellType,levels=c(broadCT
_order))

options(repr.plot.height=5,repr.plot.width=7)
g <- ggplot(toPlot,aes_string(x='Freq',y='sample',fill='snATAC_cellTyp
e')) + geom_bar(stat='identity',position='stack') +
        theme_bw(base_size=25) + labs(x='Cell Count',y='Sample',fill='Br
oad\nCell\nType',title='Multiome Samples') +
        theme(plot.title = element_text(hjust = 0.5)) + scale_fill_manua
l(values=broadCT_colors) +
        theme(axis.text.y=element_blank(),axis.ticks.y=element_blank())
+
        theme(legend.position="none")
print(g)

if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'multiome_sample_
broadCT_cellCounts.png'),
                            plot=g,units='in',height=5,width=7,dpi=600)
```



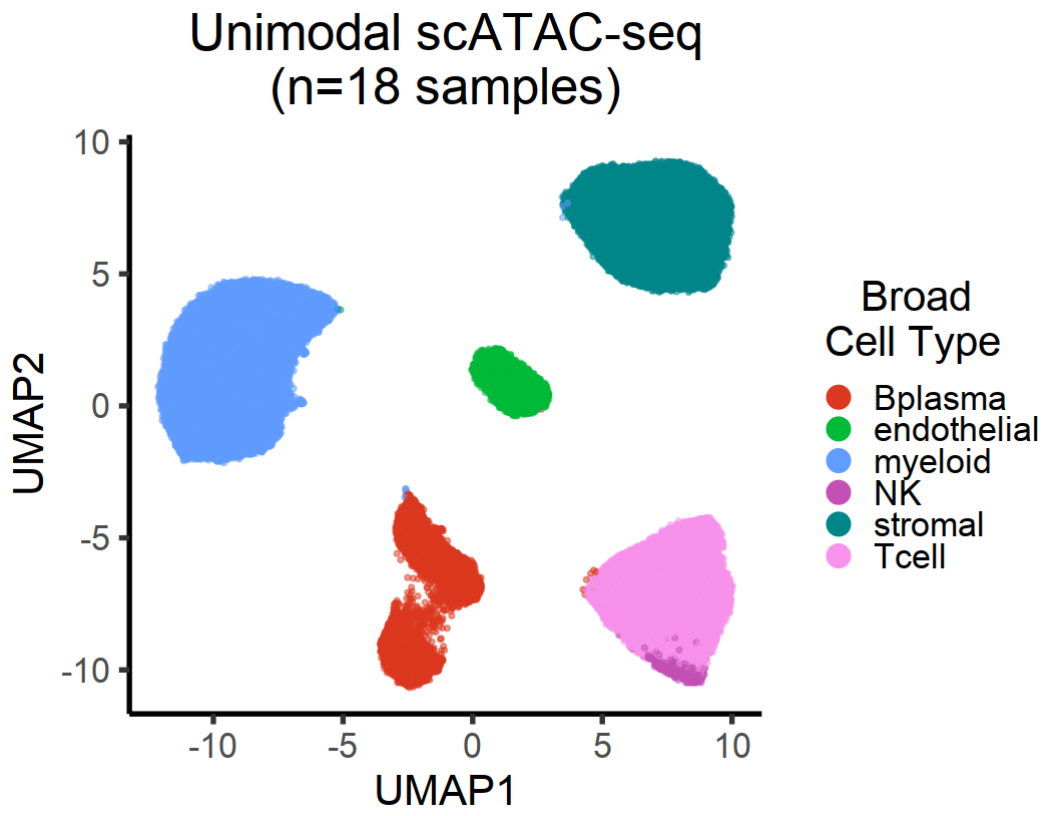
# Peak Overlaps

In [10]:
```
#Fig S1f

options(repr.plot.height=5,repr.plot.width=6)
g <- ggplot(peakComparison_df,aes_string(x='feature',y='peakCt',fill='ov
erlap')) +
        geom_bar(stat='identity',position='stack') +
        theme_bw(base_size=25) + labs(x='Feature',y='Peak Counts',fill
='Overlap') +
        scale_fill_manual(values=c('grey85','black'))
print(g)

if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'peakComparison.p
ng'),
                            plot=g,units='in',height=5,width=6,dpi=600)
```



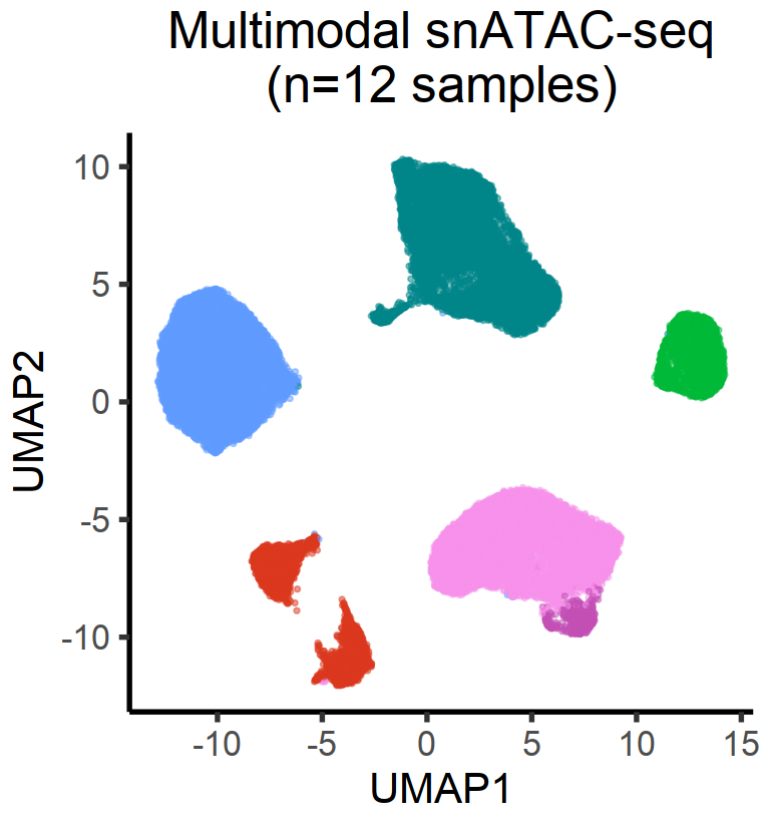
## Cell Type UMAPs

In [11]:
```
#Fig 1b left

options(repr.plot.height=7,repr.plot.width=9)
g <- ggplot(scATAC_meta,aes_string(x='UMAP1',y='UMAP2',color='cellTyp
e')) + geom_point(size=1,alpha=0.5) +
        theme_classic(base_size=25) + scale_color_manual(values=broadCT_
colors) +
        labs(color='   Broad\nCell Type',title='Unimodal scATAC-seq\n(n=
18 samples)') +
        theme(plot.title = element_text(hjust = 0.5)) +
        guides(colour = guide_legend(override.aes = list(size=6,alpha=
1)))
print(g)

if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'scATAC_broadCT_U
MAP.png'),
                            plot=g,units='in',height=7,width=9,dpi=600)
```

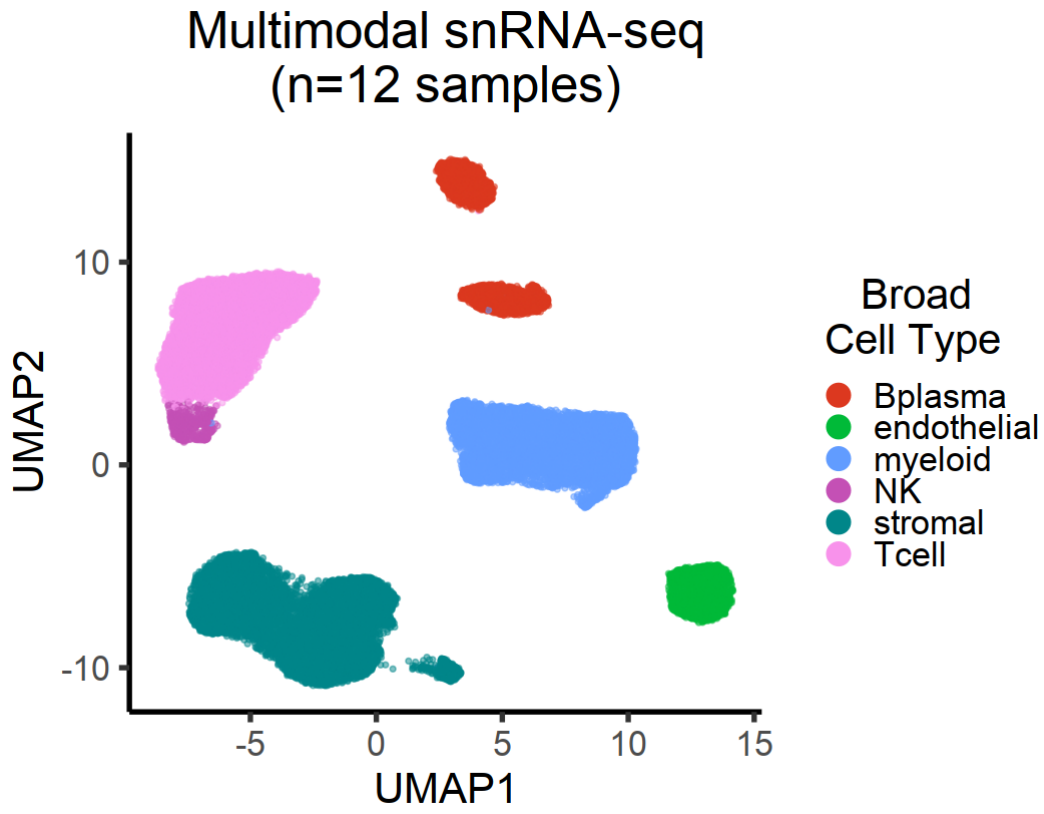


In [12]:

```
#Fig 1b left

options(repr.plot.height=7,repr.plot.width=6.5)
g <- ggplot(multiome_meta,aes_string(x='snATAC_UMAP1',y='snATAC_UMAP2',c
olor='snATAC_cellType')) +
        geom_point(size=1,alpha=0.5) +
        theme_classic(base_size=25) + scale_color_manual(values=broadCT_
colors) +
        labs(color='   Broad\nCell Type',title='Multimodal snATAC-seq\n
(n=12 samples)',x='UMAP1',y='UMAP2') +
        theme(plot.title = element_text(hjust = 0.5)) +
        theme(legend.position="none")
print(g)

if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'snATAC_broadCT_U
MAP.png'),
                            plot=g,units='in',height=7,width=6.5,dpi=60
0)
```



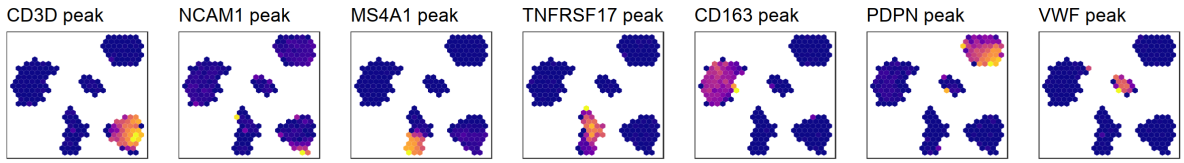
In [13]:
```
#Fig S1k

options(repr.plot.height=7,repr.plot.width=9)
g <- ggplot(multiome_meta,aes_string(x='snRNA_UMAP1',y='snRNA_UMAP2',col
or='snRNA_cellType')) +
        geom_point(size=1,alpha=0.5) +
        theme_classic(base_size=25) + scale_color_manual(values=broadCT_
colors) +
        labs(color='   Broad\nCell Type',title='Multimodal snRNA-seq\n(n
=12 samples)',x='UMAP1',y='UMAP2') +
        theme(plot.title = element_text(hjust = 0.5)) +
        guides(colour = guide_legend(override.aes = list(size=6,alpha=
1)))
print(g)

if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'snRNA_broadCT_UM
AP.png'),
                            plot=g,units='in',height=7,width=9,dpi=600)
```



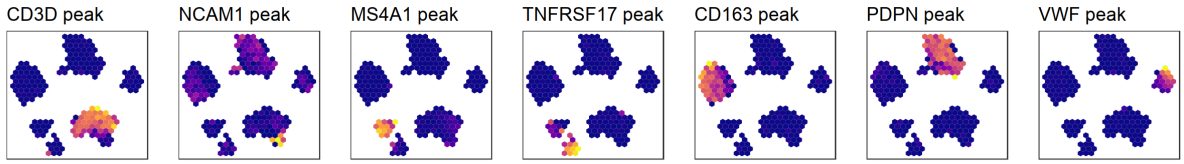

## Marker UMAPs

In [14]:
```
genes_forUMAPs <- c('CD3D','NCAM1','MS4A1','TNFRSF17','CD163','PDPN','VW
F')
if(!all(genes_forUMAPs %in% names(chosenPeaks))) stop('Genes for UMAP no
t in chosen genes')
peaks_forUMAPs <- chosenPeaks[genes_forUMAPs]
```

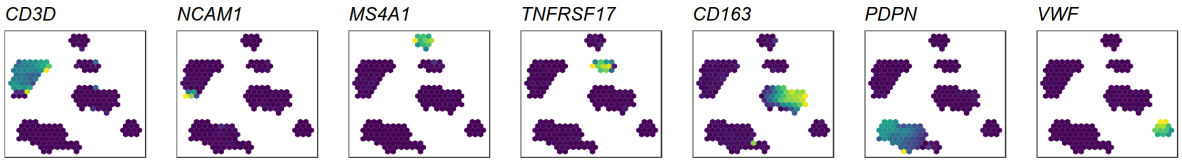
In [15]:
```
#Fig S1g
toPlot <- scATAC_pxc_norm[c(unname(peaks_forUMAPs)),]
rownames(toPlot) <- paste(sep='',names(peaks_forUMAPs),' peak')

options(repr.plot.height=3,repr.plot.width=20)
g <- plot_markerPeaks_norm_hex_v2(scATAC_meta,toPlot,'UMAP1','UMAP2',
                                  plot_genes=rownames(toPlot),plotCol=nr
ow(toPlot),
                                  titleSize=22,hex_bins=23,cutCap=0,colo
rOpt='plasma',titleFace='plain')
grid.draw(g)
if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'scATAC_markerPea
k_UMAP.png'),
                            plot=g,units='in',height=3,width=20,dpi=600)
```




In [16]:
```
#Fig S1h
toPlot <- snATAC_pxc_norm[c(unname(peaks_forUMAPs)),]
rownames(toPlot) <- paste(sep='',names(peaks_forUMAPs),' peak')

options(repr.plot.height=3,repr.plot.width=20)
g <- plot_markerPeaks_norm_hex_v2(multiome_meta,toPlot,'snATAC_UMAP1','s
nATAC_UMAP2',
                                  plot_genes=rownames(toPlot),plotCol=nr
ow(toPlot),
                                  titleSize=22,hex_bins=23,cutCap=0,colo
rOpt='plasma',titleFace='plain')
grid.draw(g)
if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'snATAC_markerPea
k_UMAP.png'),
                            plot=g,units='in',height=3,width=20,dpi=600)
```



In [17]:
```
#Fig S11

options(repr.plot.height=3,repr.plot.width=20)
g <- plot_markerPeaks_norm_hex_v2(multiome_meta,snRNA_gxc_norm,'snRNA_UMAP1','snRNA_UMAP2',
                                  plot_genes=genes_forUMAPs,plotCol=length(genes_forUMAPs),
                                  titleSize=22,hex_bins=23,cutCap=0)
grid.draw(g)
if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'snRNA_markerGene_UMAP.png'),
                            plot=g,units='in',height=3,width=20,dpi=600)
```

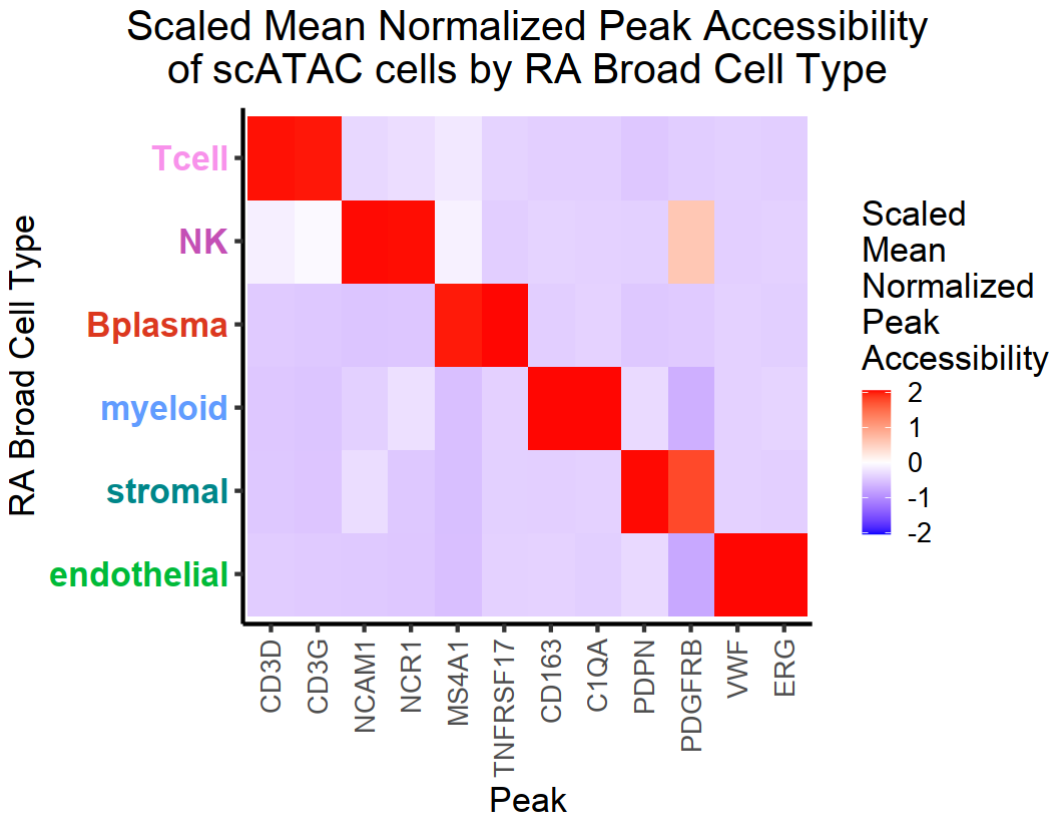



## Marker Heatmaps

In [18]:
```
scale_lim <- 2.05
```

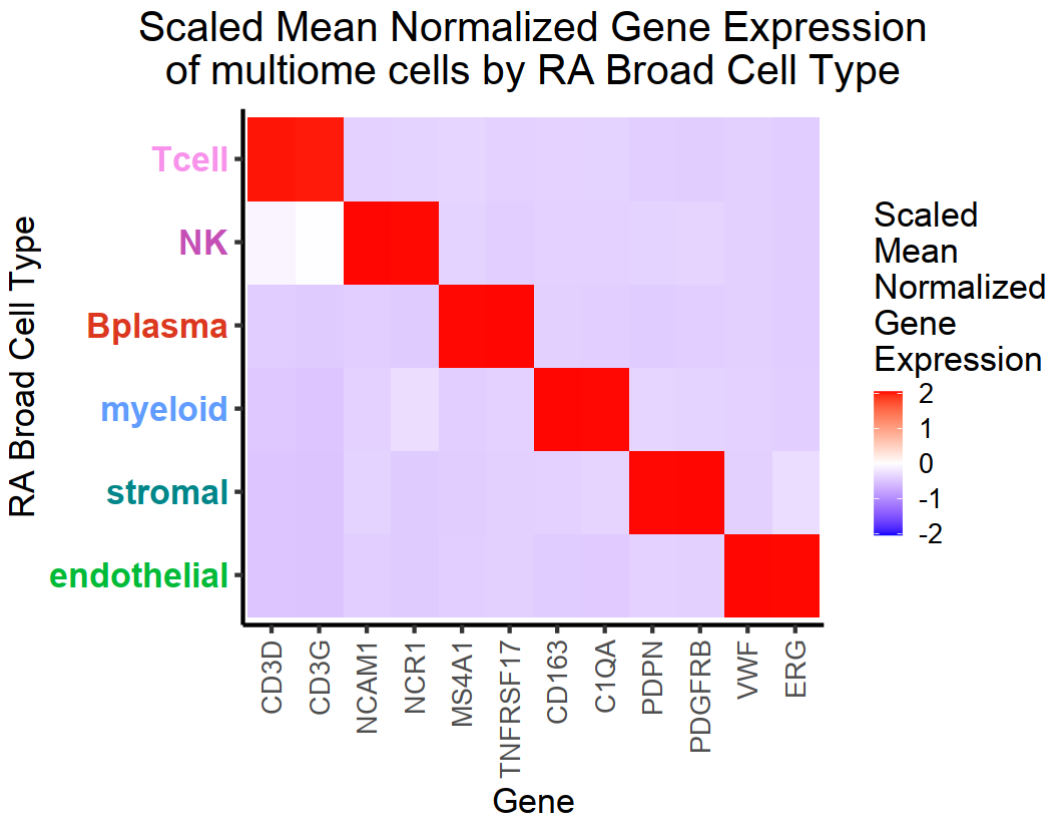
In [19]:

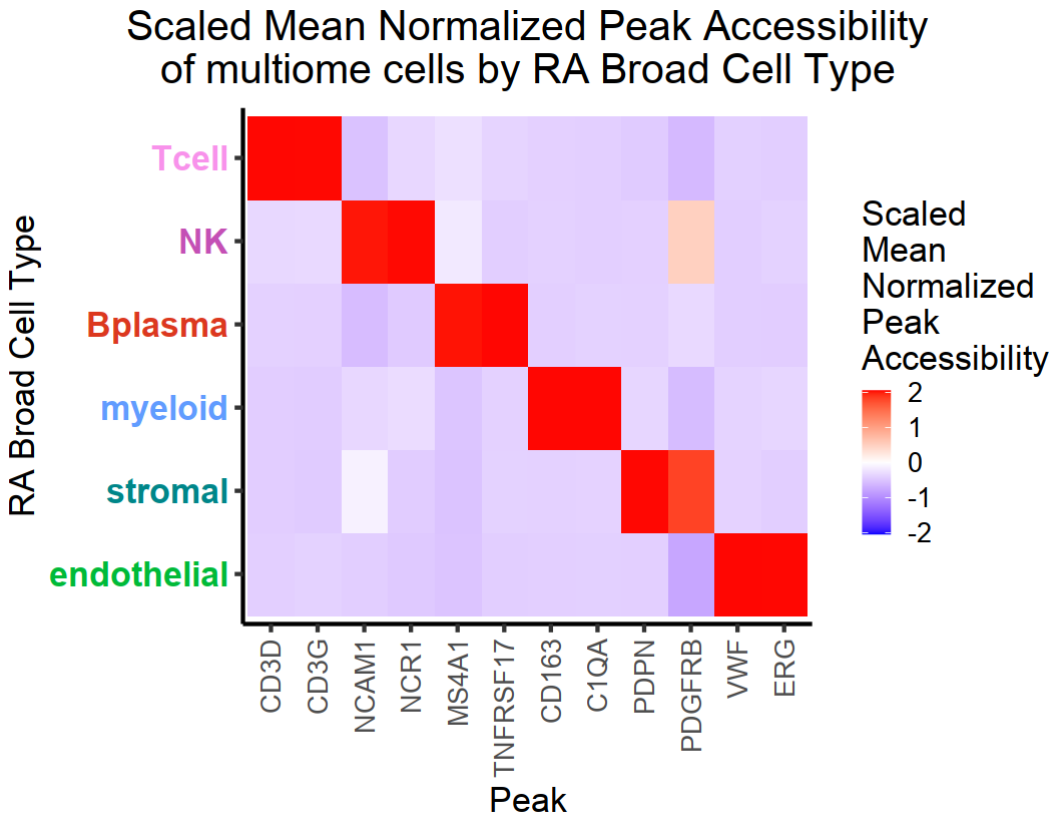
```
#Fig S1i

scATAC_pxCT_norm_subset_scaled <- scalePeak_forHeatmap(names(chosenPeak
s),broadCT_order,chosenPeaks,scATAC_pxCT_norm)

if(max(abs(scATAC_pxCT_norm_subset_scaled))>=scale_lim) stop('scale limi
t too low')

options(repr.plot.height=7,repr.plot.width=9)
g <- pseudobulk_scaled_heatmap(scATAC_pxCT_norm_subset_scaled,'Peak',pas
te('RA',CT_label),
                               'Scaled\nMean\nNormalized\nPeak\nAccessib
ility',
                               plotTit=paste('Scaled Mean Normalized Pea
k Accessibility\nof scATAC cells by RA',CT_label),
                               scale_lim=scale_lim,clustColors=broadCT_c
olors)
print(g)
if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'scATAC_markerPea
k_heatmap.png'),
                            plot=g,units='in',height=7,width=9,dpi=600)
```



In [20]:

```
#Fig S1j,m

res <- scaleFeat_forHeatmap(names(chosenPeaks),broadCT_order,chosenPeak
s,snRNA_gxCT_norm,snATAC_pxCT_norm)
snRNA_gxCT_norm_subset_scaled <- res$gxCT_norm_subset_scaled
snATAC_pxCT_norm_subset_scaled <- res$pxCT_norm_subset_scaled

if(max(abs(snRNA_gxCT_norm_subset_scaled),abs(snATAC_pxCT_norm_subset_sc
aled),
       na.rm=TRUE)>=scale_lim) stop('scale limit too low')


options(repr.plot.height=7,repr.plot.width=9)
g <- pseudobulk_scaled_heatmap(snRNA_gxCT_norm_subset_scaled,'Gene',past
e('RA',CT_label),
                               'Scaled\nMean\nNormalized\nGene\nExpressi
on',
                               plotTit=paste('Scaled Mean Normalized Gen
e Expression\nof multiome cells by RA',CT_label),
                               scale_lim=scale_lim,clustColors=broadCT_c
olors)
print(g)
if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'snRNA_markerGene
_heatmap.png'),
                            plot=g,units='in',height=7,width=9,dpi=600)

g <- pseudobulk_scaled_heatmap(snATAC_pxCT_norm_subset_scaled,'Peak',pas
te('RA',CT_label),
                               'Scaled\nMean\nNormalized\nPeak\nAccessib
ility',
                               plotTit=paste('Scaled Mean Normalized Pea
k Accessibility\nof multiome cells by RA',CT_label),
                               scale_lim=scale_lim,clustColors=broadCT_c
olors)
print(g)
if(!is.na(save_dir)) ggsave(file=paste(sep='',save_dir,'snATAC_markerPea
k_heatmap.png'),
                            plot=g,units='in',height=7,width=9,dpi=600)
```


Scaled Mean Normalized Gene Expression of multiome cells by RA Broad Cell Type
RA Broad Cell Type
Tcell
NK
Bplasma
myeloid
stromal
endothelial
CD3D
CD3G
NCAM1
NCR1
MS4A1
TNFRSF17
CD163
C1QA
PDPN
PDGFRB
VWF
ERG
Gene
Scaled Mean Normalized Gene Expression
2
1
0
-1
-2


Scaled Mean Normalized Peak Accessibility of multiome cells by RA Broad Cell Type
RA Broad Cell Type
Tcell
NK
Bplasma
myeloid
stromal
endothelial
CD3D
CD3G
NCAM1
NCR1
MS4A1
TNFRSF17
CD163
C1QA
PDPN
PDGFRB
VWF
ERG
Peak
Scaled Mean Normalized Peak Accessibility
2
1
0
-1
-2

# Session Info

In [21]:
```
sessionInfo()
```

```
R version 3.6.1 (2019-07-05)
Platform: x86_64-conda_cos6-linux-gnu (64-bit)
Running under: Red Hat Enterprise Linux Server release 6.5 (Santiago)

Matrix products: default
BLAS/LAPACK: /PHShome/kew47/miniconda3/lib/R/lib/libRblas.so

locale:
[1] en_US.UTF-8

attached base packages:
[1] grid      stats     graphics  grDevices utils     datasets  methods
[8] base

other attached packages:
 [1] repr_1.0.1        gridExtra_2.3     scales_1.1.1      viridis_0.5.
1
 [5] viridisLite_0.3.0 ggrepel_0.8.2     ggrastr_0.2.3     stringr_1.4.
0
 [9] ROCR_1.0-7        gplots_3.0.1.1    Rmisc_1.5.1       plyr_1.8.6
[13] lattice_0.20-41   gtools_3.8.2      tidyr_1.0.3       Matrix_1.2-1
8
[17] ggplot2_3.3.0

loaded via a namespace (and not attached):
 [1] pbdZMQ_0.3-3        beeswarm_0.2.3      tidyselect_1.1.0
 [4] purrr_0.3.4         colorspace_1.4-1    vctrs_0.3.5
 [7] generics_0.0.2      htmltools_0.4.0     base64enc_0.1-3
[10] rlang_0.4.8         hexbin_1.28.1       pillar_1.4.4
[13] glue_1.4.0          withr_2.2.0         uuid_0.1-2
[16] lifecycle_0.2.0     munsell_0.5.0       gtable_0.3.0
[19] caTools_1.18.0      evaluate_0.14       labeling_0.3
[22] Cairo_1.5-10        vipor_0.4.5         IRdisplay_0.7.0
[25] Rcpp_1.0.4.6        KernSmooth_2.23-15  gdata_2.18.0
[28] IRkernel_1.0.2.9000 jsonlite_1.7.1      farver_2.0.3
[31] digest_0.6.25       stringi_1.4.6       dplyr_1.0.2
[34] tools_3.6.1         bitops_1.0-6        magrittr_1.5
[37] tibble_3.0.1        crayon_1.3.4        pkgconfig_2.0.3
[40] ellipsis_0.3.1      ggbeeswarm_0.6.0    R6_2.4.1
[43] compiler_3.6.1
```

In [ ]:
```

```