

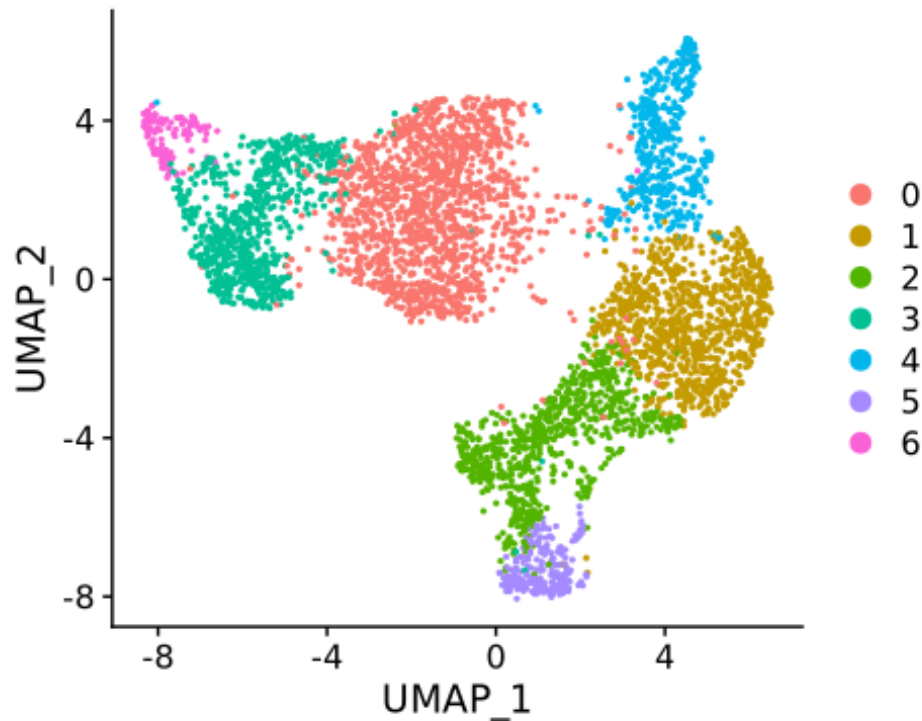
Reference mapping analysis, Supplementary Figures 2E,F Gilioli, Fusco, et al.

```
library(Seurat)
library(SeuratWrappers)
library(ggplot2)
library(dplyr)
library(ggrepel)
library(openxlsx)
library(circlize)
library(RColorBrewer)
library(ComplexHeatmap)
library(reshape2)
library(stringr)
library(scales)
library(readxl)
library(tidyr)
library(gtools)
library(grid)
library(gridExtra)
library(clusterProfiler)

set.seed(42)

obj <- readRDS(paste(folder, 'Full_cluster_recalc.rds', sep = '/'))

Idents(obj) <- 'RNA_snn_res.0.4' #seurat_clusters
DimPlot(obj)
```



```
t <- list() # List results from reference mapping analysis
```

Mapping data from healthy bone marrow samples Human BoneMarrowMap Zheng et al 2025

```
andyBM <- readRDS('BoneMarrowMap_Annotated_Dataset_expandedFeatures.rds')

# Predict Hematopoietic Cell Types
Exvivo.anchors <- FindTransferAnchors(reference = andyBM, query = obj, dims = 1:50,
, reference.reduction = "pca" )

predictionsandy <- TransferData(anchorset = Exvivo.anchors, refdata = andyBM$CellType,
dims = 1:50)
andy.query <- AddMetaData(obj, metadata = predictionsandy)

t[['andy']] <- table(andy.query$predicted.id, andy.query$RNA_snn_res.0.4)

Idents(andy.query) <- 'predicted.id'
pandy <- DimPlot(andy.query)

pandyx <- pandy +
  ggtitle("Healthy Bone Marrow Map \n (Zheng et al 2025)") +
  theme_classic(base_size = 10) +
  theme(legend.text = element_text(size = 10),
        plot.title = element_text(hjust = 0.5))

andyxfeature <- FeaturePlot(andy.query,
```

```

        features = "prediction.score.max",
        min.cutoff = 0.3,
        max.cutoff = 1) +
scale_color_gradient2(low = "grey90",
                      mid = "orange",
                      high = "darkred",
                      midpoint = 0.6) +
ggtitle("Healthy Bone Marrow Map \n (Zheng et al 2025)") +
theme_classic()

```

Analysis with healthy BM dataset from van Galen et al., 2019

```

load('VanGalen.Rdata')
updated_obj <- UpdateSeuratObject(merge.object)

Exvivo.anchors <- FindTransferAnchors(reference = updated_obj, query = obj, dims =
1:50, reference.reduction = "pca" )

predictionsvangalen <- TransferData(anchorset = Exvivo.anchors, refdata = updated_
obj$predictionRF, dims = 1:50)
VanGalen_healthy.query <- AddMetaData(obj, metadata = predictionsvangalen)

t[['vangalen']] <- table(VanGalen_healthy.query$predicted.id, VanGalen_healthy.que
ry$RNA_snn_res.0.4)

Idents(VanGalen_healthy.query) <- 'predicted.id'
VanGalen_healthy <- DimPlot(VanGalen_healthy.query) +
  ggtitle("Healthy Bone Marrow \n (van Galen et al 2019)") +
  theme_classic(base_size = 10) +
  theme(legend.text = element_text(size = 10),
        plot.title = element_text(hjust = 0.5))

VanGalen_healthy_score <- FeaturePlot(VanGalen_healthy.query,
                                     features = "prediction.score.max",
                                     min.cutoff = 0.3,
                                     max.cutoff = 1) +
  scale_color_gradient2(low = "grey90",
                        mid = "orange",
                        high = "darkred",
                        midpoint = 0.6) +
  ggtitle("Healthy Bone Marrow \n (van Galen et al 2019)") +
  theme_classic()

```

Mapping data from Naldini et al. 2023, NPM1-mutant dataset

```
data <- readRDS('NPM1_AML.rds') # already with cluster group

Idents(data) <- 'Clusters_0.6'
p <- DimPlot(data, label=T, label.size = 12)

Exvivo.anchors <- FindTransferAnchors(reference = data, query = obj, dims = 1:50,
                                     reference.reduction = "pca" )

predictions_nmp1 <- TransferData(anchorset = Exvivo.anchors, refdata = data$Clusters_0.6, dims = 1:50)
nald.query <- AddMetaData(obj, metadata = predictions_nmp1)

t[['naldini']] <- table(nald.query$predicted.id, nald.query$RNA_snn_res.0.4)

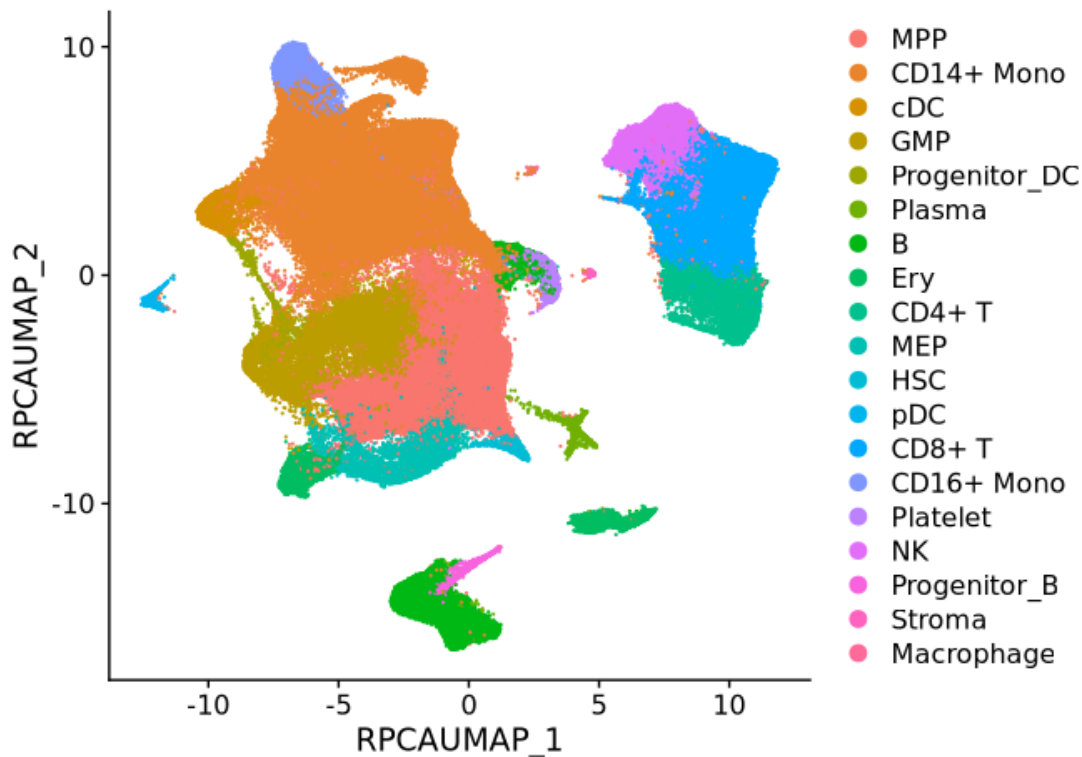
Idents(nald.query) <- 'predicted.id'
p0 <- DimPlot(nald.query)
p0nald <- p0 +
  ggtitle("NPM1-mut AML \n (Naldini et al 2023)") +
  theme_classic(base_size = 10) +
  theme(legend.text = element_text(size = 10),
        plot.title = element_text(hjust = 0.5))

Naldfeature <- FeaturePlot(nald.query,
                          features = "prediction.score.max",
                          min.cutoff = 0.3,
                          max.cutoff = 1) +
  scale_color_gradient2(low = "grey90",
                        mid = "orange",
                        high = "darkred",
                        midpoint = 0.6) +
  ggtitle("NPM1-mut AML \n (Naldini et al 2023)") +
  theme_classic()
```

Analysis with reference from Sirenko et al 2025, IDH-mutant dataset

```
idhdataset <- readRDS('IDH_AML_integrated_seurat_object_normandscaled.rds')
Idents(idhdataset) <- 'manual_annotation_v3_l3_final'

DimPlot(idhdataset, raster = FALSE)
```



```
# perform standard preprocessing on each object <- was done before
# idhdataset <- NormalizeData(idhdataset)
# idhdataset <- FindVariableFeatures(idhdataset)
# idhdataset <- ScaleData(idhdataset)
```

```
Exvivo.anchors <- FindTransferAnchors(reference = idhdataset, query = obj, dims =
1:50, reference.reduction = "integrated_pca" )
```

```
predictionsidh <- TransferData(anchorset = Exvivo.anchors, refdata = idhdataset$ma
nual_annotation_v3_l3_final, dims = 1:50)
idh.query <- AddMetaData(obj, metadata = predictionsidh)
```

```
t[['sirenko']] <- table(idh.query$predicted.id,idh.query$RNA_snn_res.0.4)
```

```
Idents(idh.query) <- 'predicted.id'
```

```
p_IDH <- DimPlot(idh.query, #group.by = "annotation1",
  label = FALSE, pt.size = 0.5) +
  ggtitle("IDH-mutant AML \n (Sirenko et al 2025)") +
  theme_classic(base_size = 10) +
  theme(legend.text = element_text(size = 10),
    plot.title = element_text(hjust = 0.5))
```

[illegible]

```

        min.cutoff = 0.3,
        max.cutoff = 1) +
scale_color_gradient2(low = "grey90",
                      mid = "orange",
                      high = "darkred",
                      midpoint = 0.6) +
ggtitle("IDH-mutant AML \n (Sirenko et al 2025)") +
theme_classic()

```

Number of cells assigned per cluster for each reference dataset

```

namex <- list('sirenko' = 'Sirenko \n et al. 2025', 'naldini' = 'Naldini \n et al.
2023', 'andy' = 'Zheng \n et al. 2025', 'vangalen' = 'van Galen \n at al. 2019')
df <- data.frame()
for ( n in names(t)){n
  x <- as.data.frame.matrix(t[[n]])
  x$Dataset <- namex[[n]]
  if (nrow(df) > 0) {
    df <- rbind(df, x)
  } else {
    df <- x
  }
}

df$Celltype <- rownames(df)
df$Name <- paste0(df$Celltype, ' ', df$Dataset)
df_melt <- melt(df)

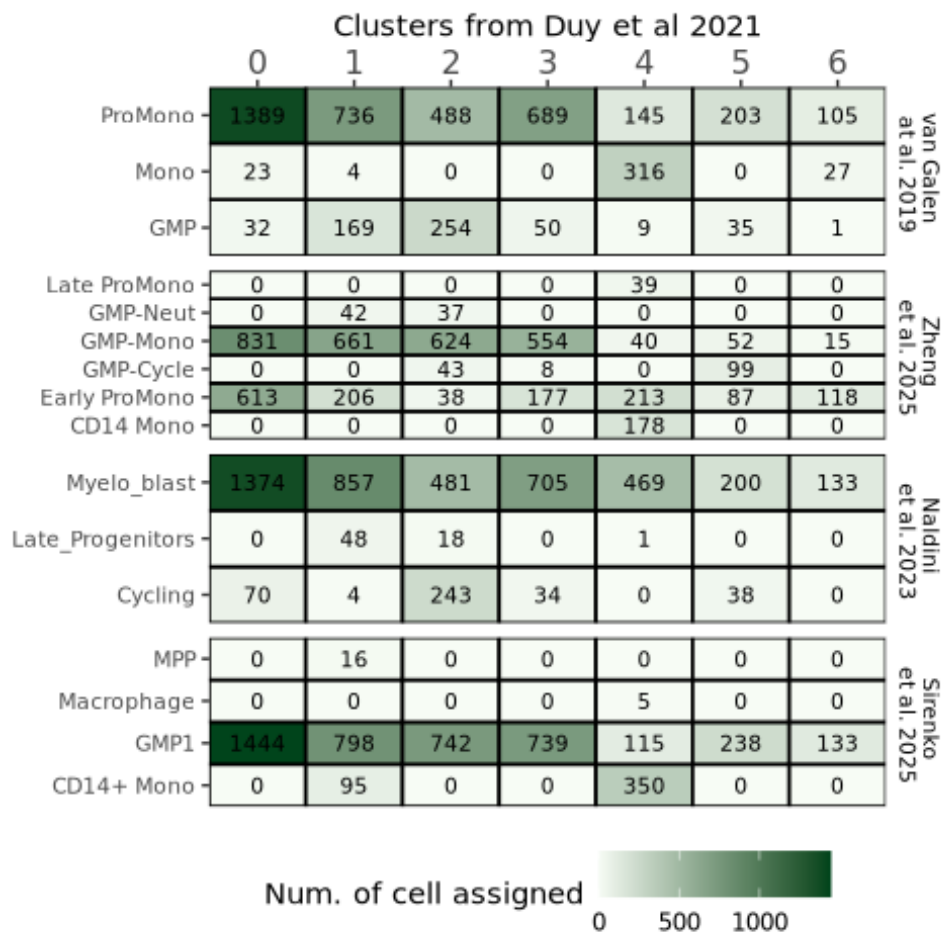
df_melt$Dataset <- factor(df_melt$Dataset, levels = c("van Galen \n at al. 2019",
"Zheng \n et al. 2025", "Naldini \n et al. 2023", "Sirenko \n et al. 2025"))

tableplot <- df_melt %>%
  group_by(Dataset) %>%
  ggplot( aes(variable, Celltype, fill = value)) +
  geom_tile(color = "black",
            lwd = 0.6,
            linetype = 1) +
  geom_text(aes(label = value), size = 8 / .pt) +
  scale_fill_gradient(low = "#F7FCF5", high = "#00441B") +
  labs(x = 'Clusters from Duy et al 2021', y = NULL, fill = 'Num. of cell assigned
')+
  facet_wrap(Dataset~., ncol=1, scales = 'free_y', strip.position = "right")+
  theme(axis.text.x = element_text(size = 12),
        axis.text.y = element_text(size = 8),
        legend.position = "bottom",
        strip.background = element_rect(fill = "white", colour = NA),
        strip.text = element_text(colour = "black", size =8))+
  scale_x_discrete(expand=c(0,0), position = "top" )+
  scale_y_discrete(expand=c(0,0))
tableplot

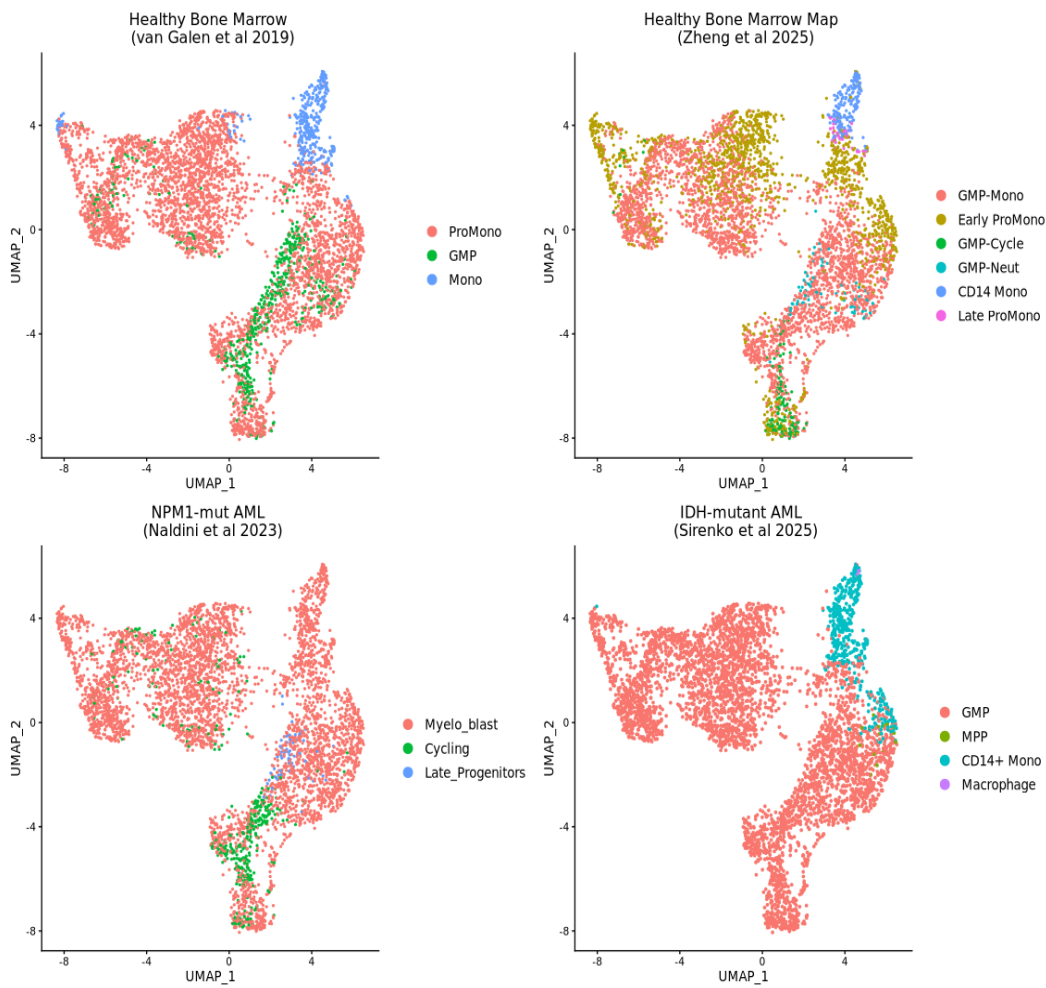
```

```
dp <- DotPlot(obj,
  features = #c(cd14_markers) ,#negative_markers),
  c("CD14", "CD68", "CD36", "VCAN",
    "S100A8", "S100A9", "MNDA", "CST3"),
  group.by = "RNA_snn_res.0.4",
  cols = c("grey90", "darkred"),
  dot.scale = 8) +
  coord_flip() +
  theme_classic() +
  ggtitle("Monocyte markers per cluster")
```

dp



Reference mapping results for all reference datasets



Prediction score results from the reference mapping analysis

