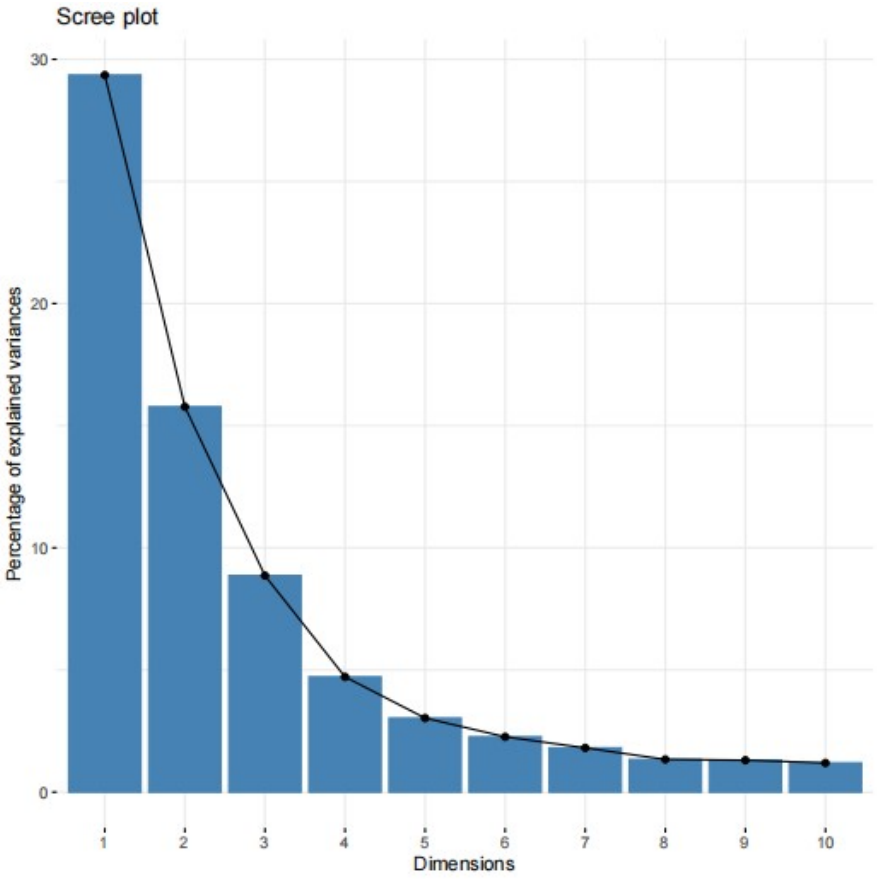
Variables Contribute to Gene Expression Variance

Conduct Principle Component Analysis on gene expression in 921 samples after normalization



Elbow plot

Explore clinical variables signigicantly related to PC1 and PC2

Principle Component ~ Clinical variable

H0: Clinical variables do not related to Principle Component

Select clinical variables with adjusted-p avalue < 0.05

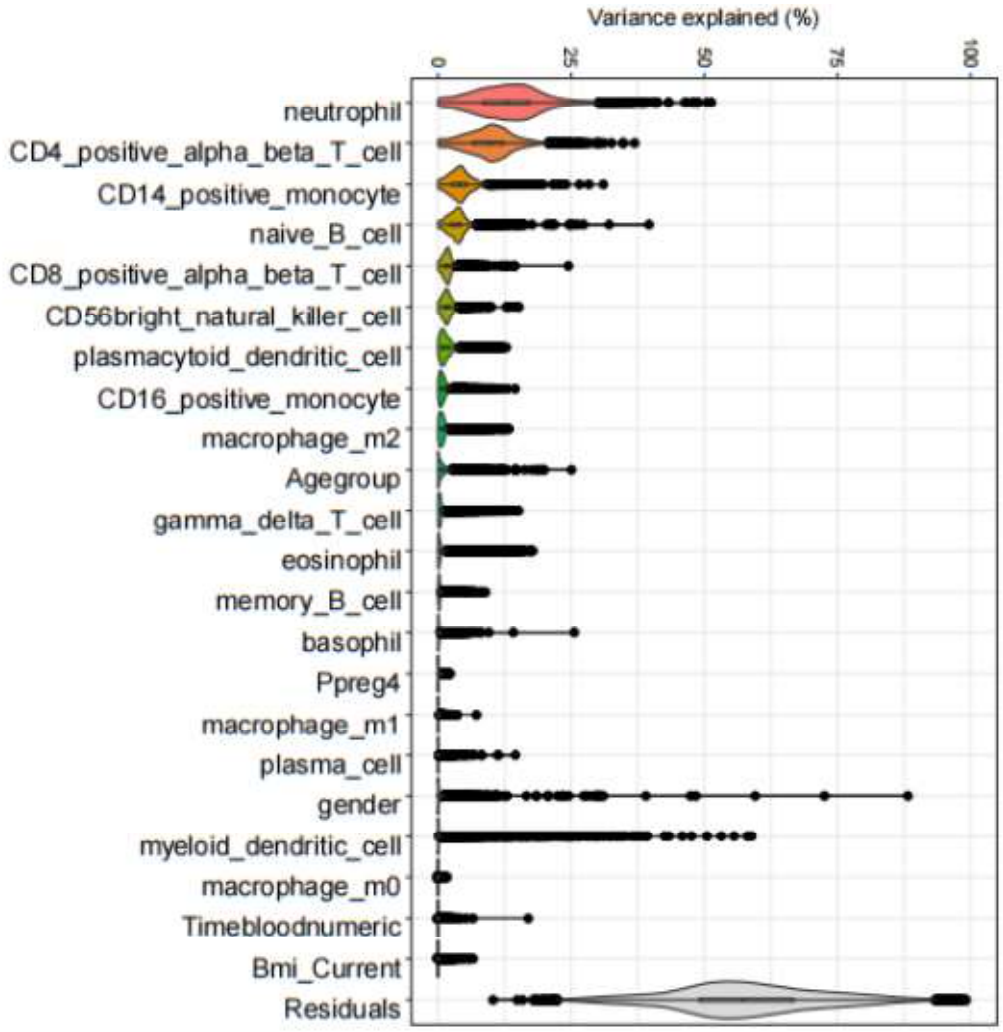
Clinical variables related to PC1: Ppreg4, Agegroup, myeloid_dendritic_cell, CD14_positive_monocyte, CD16_positive_monocyte, CD4_positive_alpha_beta_T_cell,

CD56bright_natural_killer_cell, eosinophil, gamma_delta_T_cell, macrophage_m2, memory_B_cell, naive_B_cell, neutrophil, plasma_cell, plasmacytoid_dendritic_cell

Clinical variables related to PC2: gender, Bmi_Current, Timebloodnumeric, CD8_positive_alpha_beta_T_cell, basophil, macrophage_m0, macrophage_m1

variance partitioning analyses

The violin plot shows the contribution of variables we are interested to variance for each gene



My strategy is identifying the variable that is the next strongest driver of variance, adjusted for this variable using linear modeling along with all previously selected ones and repeated this procedure iteratively until no more confounding variables were observed to be strong drivers of variance in the data

V**ariables that contribute to gene expression variance:** neutrophil, eosinophil,

CD16_positive_monocyte, plasmacytoid_dendritic_cell, macrophage_m2, gamma_delta_T_cell, CD14_positive_monocyte

For remaining variables, they are not strong drivers of variance