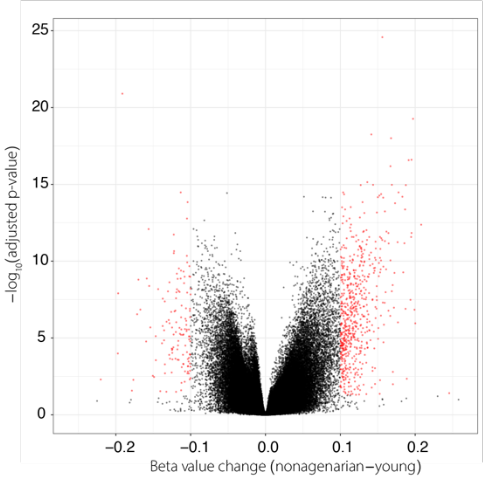
Surrogate variable analysis (SVA)

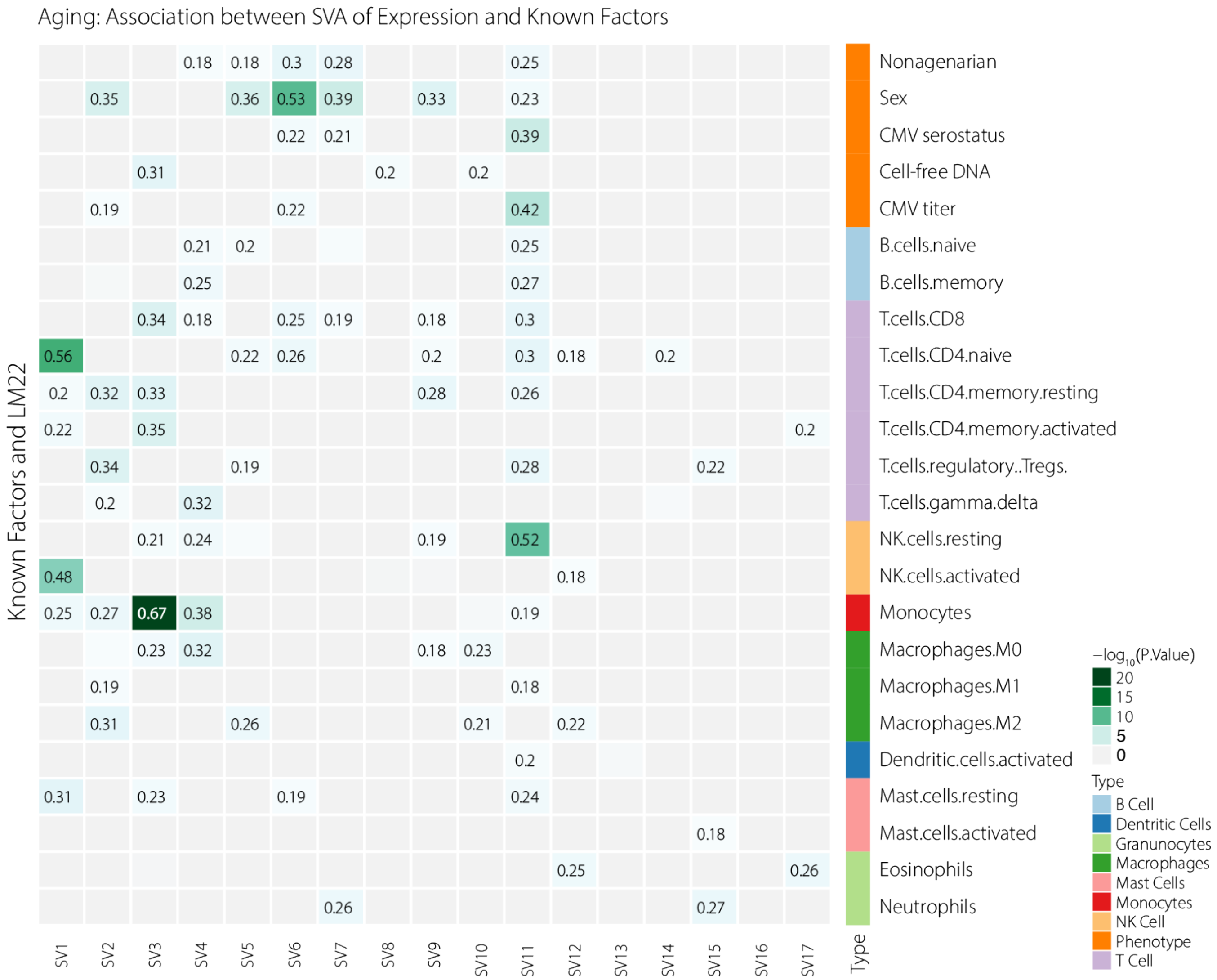
We also tested a different adjustment method, surrogate variable analysis (SVA), which adjusts for unknown covariates as a reference-free approach. Surrogate variables (SVs) are covariates constructed directly from high-dimensional data. SVA seeks to estimate the remaining unobserved factors including cell subtype proportion variations.

We used *lmFit* function of R package *limma* to identify DMPs after adjusting for SVs. We only identified 1778 DMPs. This was a much smaller number than from using the other reference-based approaches.



We tested the correlations of each SV to known covariates as well as cell subtype proportions. Although SVs were correlated with cell subtype proportions, SVs were also strongly and significantly associated with the phenotypic status (young/nonagenarians) in unadjusted data (**Figure 3a**), after removal of batch effect (**Figure 3b**) as well as after removal of batch and cell subtype proportion effects (**Figure 3c**). These results suggest that the SVA approach has the potential to mask some of the genuinely phenotype-associated effects.

We also performed similar analysis on gene expression analysis and observed similar trends to DNA methylation analysis.

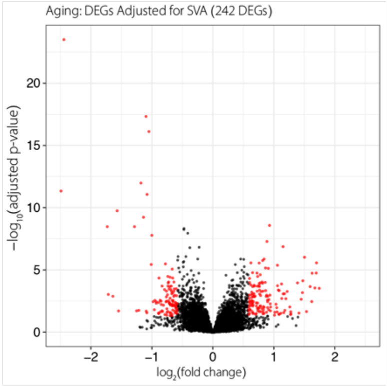


Aging: Association between SVA of Expression and Known Factors

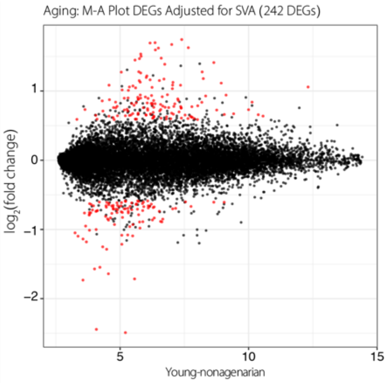
	SV1	SV2	SV3	SV4	SV5	SV6	SV7	SV8	SV9	SV10	SV11	SV12	SV13	SV14	SV15	SV16	SV17
Nonagenarian				0.18	0.18	0.3	0.28				0.25						
Sex		0.35			0.36	0.53	0.39		0.33		0.23						
CMV serostatus						0.22	0.21				0.39						
Cell-free DNA			0.31					0.2		0.2							
CMV titer		0.19				0.22					0.42						
B.cells.naive				0.21	0.2						0.25						
B.cells.memory				0.25							0.27						
T.cells.CD8			0.34	0.18		0.25	0.19		0.18		0.3						
T.cells.CD4.naive	0.56				0.22	0.26			0.2		0.3	0.18		0.2			
T.cells.CD4.memory.resting	0.2	0.32	0.33						0.28		0.26						
T.cells.CD4.memory.activated	0.22		0.35														0.2
T.cells.regulatory..Tregs.		0.34			0.19						0.28				0.22		
T.cells.gamma.delta		0.2		0.32													
NK.cells.resting			0.21	0.24					0.19		0.52						
NK.cells.activated	0.48											0.18					
Monocytes	0.25	0.27	0.67	0.38							0.19						
Macrophages.M0			0.23	0.32					0.18	0.23							
Macrophages.M1		0.19									0.18						
Macrophages.M2		0.31			0.26					0.21		0.22					
Dendritic.cells.activated											0.2						
Mast.cells.resting	0.31		0.23			0.19					0.24						
Mast.cells.activated															0.18		
Eosinophils												0.25					0.26
Neutrophils							0.26								0.27		

The rows show the known covariates and cell subtype proportions (LM22), and each column represents the SV of gene expression.

We also performed *lmFit()* to identify DEGs after adjustment for SVs.



Aging: DEGs Adjusted for SVA (242 DEGs)



Aging: M-A Plot DEGs Adjusted for SVA (242 DEGs)

The left volcano plot showed $\log_2$ fold changes of the DEGs (red dots) with significance and with signal intensities in the right MA-plot. After adjusting for PCs from cell subtype proportion estimates, we found 233 DEGs. After adjustment, we identified 242 DEGs.