A step-by-step guide for data transformation and co-variate adjustment in analysis

1. **Compensation and transformation of the raw fluorescence intensities:** The goal of compensation was to remove "spillover" noise arisen from adjacent channels. Furthermore, signals from flow cytometry usually have a large dynamic range spanning several orders of magnitude. To facilitate protein expression analysis, a popular approach is to apply the Logicle transform to the compensated data. Output from this step are values reflecting relative protein expression levels, which are then used in downstream steps in our analysis pipeline. For details, see Section 3.1 of the Supplemental Experimental Procedures.
2. **Compute heterogeneity parameters (HPs) from single cell expression levels**: To quantify the expression heterogeneity of a protein *g* among single cells in a cell population P, we calculated the standard deviation and median absolute deviation (MAD) of its expression within the cell population. In addition, to quantify overall expression heterogeneity of all the markers in the cell population, we computed the "total SD" (and "total MAD") as the weighted sum of SDs (and MADs) of the measured proteins. A weighted sum is needed because different proteins span different dynamic ranges. For details, see Supplemental Definitions.
3. **Identify temporally stable HPs by assessing multiple baseline measurements:** Our goal is to find those HPs that are not changing very much within a person over time. We first require that the HP be significantly correlated between the two pre-vaccination time-points (days 0 and -7). For each HP, we then calculated its total variance over all subjects and the three baseline time points (days -7, 0, and 70). By decomposing the total variance into components corresponding to intra-subject variations and inter-subject variations, we then identified the HPs that are stable over the three baseline time points. Note that in these tests cutoff thresholds need to be used and therefore whether an HP is deemed "temporally stable" is dependent on these thresholds.
4. **Age association analysis:** We used mixed effect models to identify HPs associated with age, because the approach is well suited to repeated measurements as we have here with data from days -7, 0 and 70 from each subject. To account for potential confounding factors, the models include gender, mean expression of the protein (the same protein that we are assessing the cell-to-cell heterogeneity of), and the relative frequency of the cell population as covariates. In addition, we include batch and subject to model experimental batch and subject effects.
5. **Visualization of age association results (Figures 4 and 5)**: To facilitate visualization while accounting for co-variates, for each significant age-associated HP, we calculated its "partial residuals" by using the fitted model to subtract out effects explained by covariates other than age. The resulting partial residual (per subject) is then plotted against age to illustrate their relationship (Figures 4A and 5A).
6. **SNP association analysis and visualization of association results (Figure 6):** We used linear regression to identify HPs that are associated with SNPs in our candidate list. Similar to age association analysis, to control for potential confounding factors, we

included age, gender, mean expression of the protein, and relative frequency of the cell population as covariates. Similarly, for visualization purpose we calculated partial residuals of a significant HP by using the fitted model to subtract out effects explained by covariates other than the SNP (Figures 6A and 6D).