Figure 1 – Landscape of major circulating immune cell populations

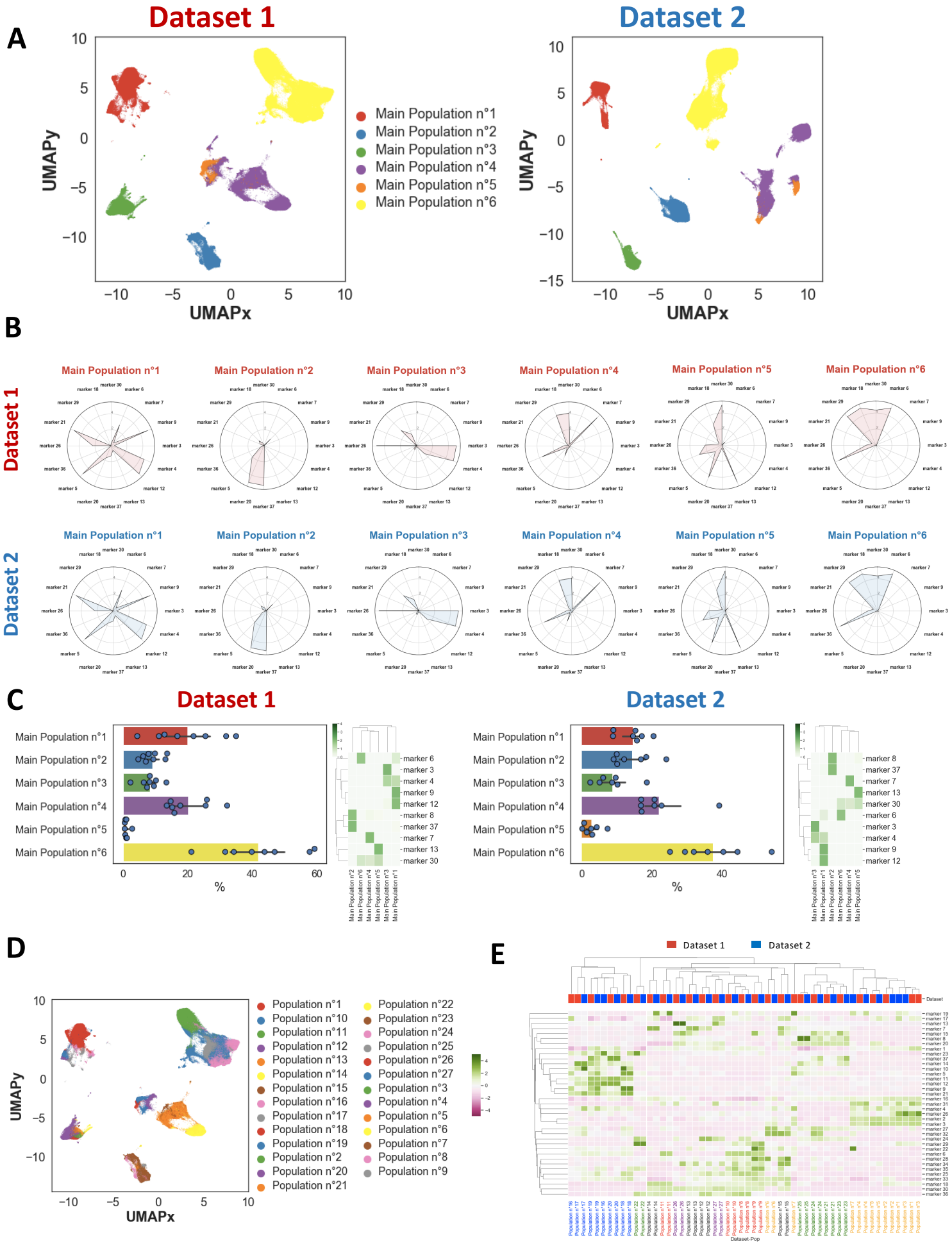
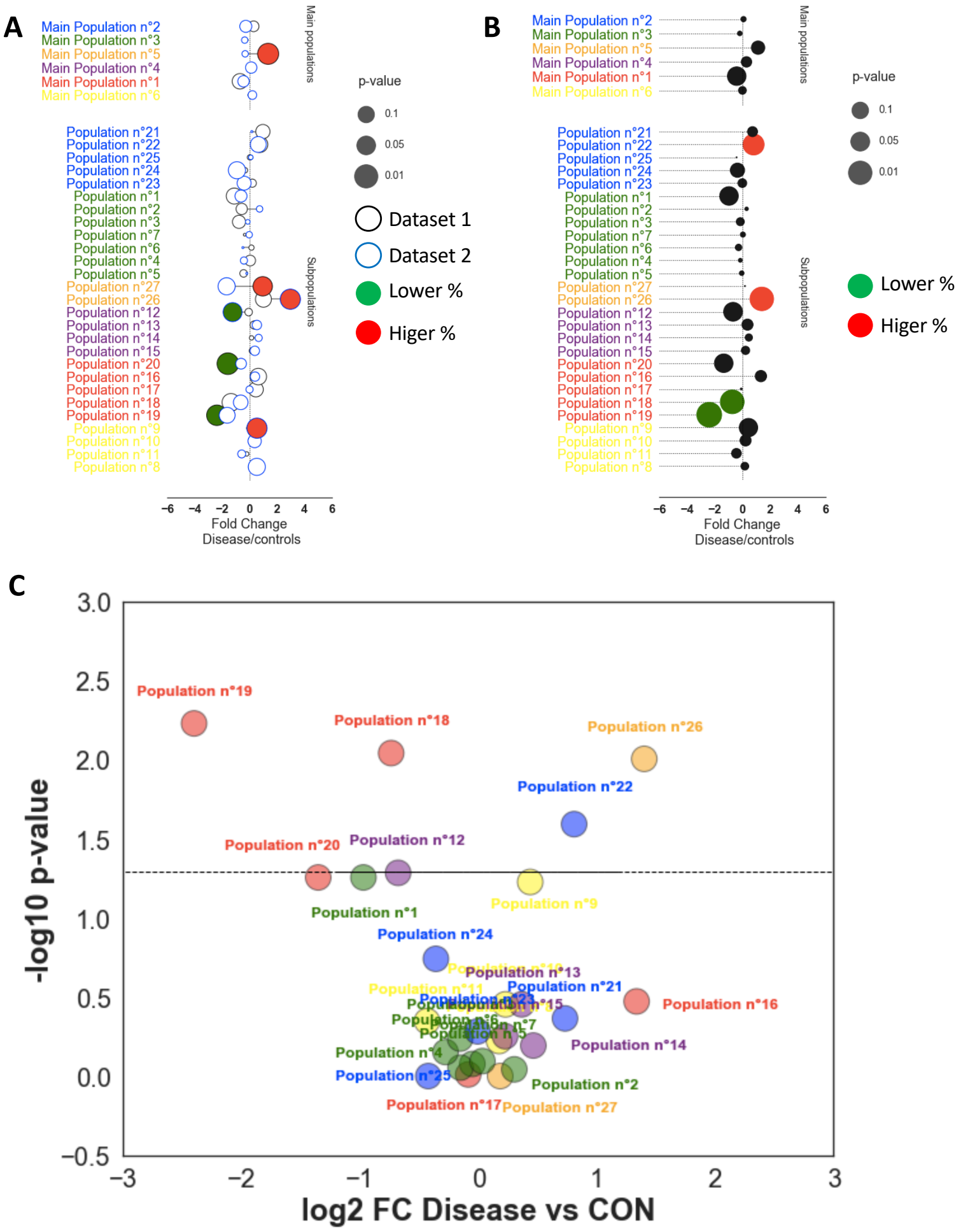


Figure 2 – Differences between controls and patients



N.B. This version is still incomplete, more figures will progressively be added

This document aims at showing the type of analyses I did and how I visualized the data. These data correspond to mass cytometry data from two independent experiments including one common donor and performed on circulating immune cells from a total of 16 patients and 24 control individuals. Importantly, here I anonymized data so you will see real figures but without names of cell populations or markers (for confidentiality reasons).

Figure 1A shows UMAP coordinates of the two datasets and the possible distinction of roughly 5 to 6 main populations of cells (people used to this method will know which ones). This is a typical step one in such experiments.

Figure 1B When performing two experiments there is almost always a batch-effect, namely some differences between them that can confound the analyses. This is also the case here of course. However, for the main immune cell populations described in Figure 1A, their phenotype or "signature" or the relative expression of some key markers are really comparable between the two datasets. I did these « polar plots » to illustrate this point (and I was totally relieved when I realized that populations where in a sense very similar between the two datasets).

Figure 1C This is just a barplot displaying percentages of each of the main immune cell populations (with roughly similar percentages between the two datasets) and a very basic heatmap showing that these populations express the expected markers (here, as they are anonymized, this is of course totally cryptic so you have to believe me).

Figure 1D shows UMAP coordinates of one of the two datasets (I planned to perform metric learning for this figure, as the two datasets are very similar, but this is not the case here) with the various subpopulations that have been identified (mostly using Phenograph).

Figure 1E This is a heat-map of the pooled results from the two datasets. Don't get me wrong, I could not just pool the two datasets and perform one big analysis because there is a batch-effect preventing this, but I can pool the results of the separate analyses performed on these two datasets. Here again, I was quite happy to see that most populations cluster together and almost all the time we have corresponding populations from each dataset next to each other.

Figure 2A This is really a figure I had a lot of fun doing. This is between a Cleveland dot plot and a Lollipop chart. Basically, here again I am focusing on differences between the two datasets, illustrating that there are differences between them. There is not a single situation where one given population is "significantly" (biological weakness here) increased or increased in both datasets. And there are even situations where there is an opposite tendency between the two datasets.

Figure 2B This is the same figure as before, but with pooled results. Sometimes, this might make sense as for Population n°22 which was increased in the disease I am studying in both datasets, albeit not significantly, and which is significantly increased after pooling results. We could argue that we did not have enough statistical power in each of the datasets taken individually and that pooling enables this.

Figure 2C This is just a volcano-plot I did. Somehow, this is redundant with Figure 2A and 2B, but here it might be easier to visualize fold changes and significances. But the real reason why I did this figure is because my supervisors wanted to visualize cell populations that were almost significantly different (those would correspond to bubbles in Figure 2A and B that are relatively big, but not big enough to be colorized in red or in green).