FACS_analysis

Fay

2022-05-18

Janeway's immunobiology:

An adaptive immune repsone is generated when naive T cells contact activated antigen-presenting cells in the secondary lymphoid organs.

T cell responses are initiated when a mature naive CD4 or CD8 T cell encounters an activated antigen-presenting cell displaying the appropriate peptide: MHC ligand.

Priming of naive CD8 T cells generates cytotoxic T cells capable of directly killing pathogen infected cells. CD4 cells develop into a diverse arrray of effector cell types depending on the nature of the signals they receive during priming.

CD4 effector can also include cytotoxicity, but more frequently it involves the secretions of a set of cytokines.

CD4: In molecular biology, CD4 (cluster of differentiation 4) is a glycoprotein that serves as a co-receptor for the T-cell receptor (TCR). CD4 is found on the surface of immune cells such as T helper cells, monocytes, macrophages, and dendritic cells.

CD4+ T helper cells are white blood cells that are an essential part of the human immune system. They are often referred to as CD4 cells, T-helper cells or T4 cells. They are called helper cells because one of their main roles is to send signals to other types of immune cells, including CD8 killer cells, which then destroy the infectious particle. If CD4 cells become depleted, for example in untreated HIV infection, or following immune suppression prior to a transplant, the body is left vulnerable to a wide range of infections that it would otherwise have been able to fight. CD4 is a co-receptor of the T cell receptor (TCR) and assists the latter in communicating with antigen-presenting cells. The TCR complex and CD4 bind to distinct regions of the antigen-presenting MHC class II molecule. The extracellular D1 domain of CD4 binds to the 2 region of MHC class II.

Treg_prop: T regulatory cells are a component of the immune system that suppress immune responses of other cells. This is an important "self-check" built into the immune system to prevent excessive reactions. Regulatory T cells come in many forms with the most well-understood being those that express CD4, CD25, and FOXP3 (CD4+CD25+ regulatory T cells). These Treg cells are different from helper T cells.[8] Another regulatory T cell subset is Treg17 cells.[9] Regulatory T cells are involved in shutting down immune responses after they have successfully eliminated invading organisms, and also in preventing autoimmunity.[10]

CD4+ FOXP3+ CD25(high) regulatory T cells have been called "naturally occurring" regulatory T cells[11] to distinguish them from "suppressor" T cell populations that are generated in vitro. Additional regulatory T cell populations include Tr1, Th3, CD8+CD28−, and Qa-1 restricted T cells. The contribution of these populations to self-tolerance and immune homeostasis is less well defined. FOXP3 can be used as a good marker for mouse CD4+CD25+ T cells, although recent studies have also shown evidence for FOXP3 expression in CD4+CD25− T cells. In humans, FOXP3 is also expressed by recently activated conventional T cells and thus does not specifically identify human Tregs.[12]

All T cells derive from progenitor cells in the bone marrow, which become committed to their lineage in the thymus. All T cells begin as CD4-CD8-TCR- cells at the DN (double-negative) stage, where an individual cell will rearrange its T cell receptor genes to form a unique, functional molecule, which they, in turn, test against cells in the thymic cortex for a minimal level of interaction with self-MHC. If they receive these

signals, they proliferate and express both CD4 and CD8, becoming double-positive cells. The selection of Tregs occurs on radio-resistant hematopoietically-derived MHC class II-expressing cells in the medulla or Hassall's corpuscles in the thymus.

CD8: CD8 (cluster of differentiation 8) is a transmembrane glycoprotein that serves as a co-receptor for the T-cell receptor (TCR). Along with the TCR, the CD8 co-receptor plays a role in T cell signaling and aiding with cytotoxic T cell-antigen interactions.

Like the TCR, CD8 binds to a major histocompatibility complex (MHC) molecule, but is specific for the MHC class I protein.[1]

The CD8 co-receptor is predominantly expressed on the surface of cytotoxic T cells, but can also be found on natural killer cells, cortical thymocytes, and dendritic cells. The CD8 molecule is a marker for cytotoxic T cell population. It is expressed in T cell lymphoblastic lymphoma and hypo-pigmented mycosis fungoides.[2]

The extracellular IgV-like domain of CD8- interacts with the 3 portion of the Class I MHC molecule.[5] This affinity keeps the T cell receptor of the cytotoxic T cell and the target cell bound closely together during antigen-specific activation. Cytotoxic T cells with CD8 surface protein are called CD8+ T cells. The main recognition site is a flexible loop at the 3 domain of an MHC molecule. This was discovered by doing mutational analyses. The flexible 3 domain is located between residues 223 and 229 in the genome.[4] In addition to aiding with cytotoxic T cell antigen interactions the CD8 co-receptor also plays a role in T cell signaling. The cytoplasmic tails of the CD8 co-receptor interact with Lck (lymphocyte-specific protein tyrosine kinase). Once the T cell receptor binds its specific antigen Lck phosphorylates the cytoplasmic CD3 and -chains of the TCR complex which initiates a cascade of phosphorylation eventually leading to activation of transcription factors like NFAT, NF- B, and AP-1 which affect the expression of certain genes.[6]

Treg17: The regulatory T cells (Tregs / ti r / or Treg cells), formerly known as suppressor T cells, are a subpopulation of T cells that modulate the immune system, maintain tolerance to self-antigens, and prevent autoimmune disease. Treg cells are immunosuppressive and generally suppress or downregulate induction and proliferation of effector T cells.[1] Treg cells express the biomarkers CD4, FOXP3, and CD25 and are thought to be derived from the same lineage as naïve CD4+ cells.[2] Because effector T cells also express CD4 and CD25, Treg cells are very difficult to effectively discern from effector CD4+, making them difficult to study. Research has found that the cytokine transforming growth factor beta (TGF-) is essential for Treg cells to differentiate from naïve CD4+ cells and is important in maintaining Treg cell homeostasis.[3]

Mouse models have suggested that modulation of Treg cells can treat autoimmune disease and cancer and can facilitate organ transplantation[4] and wound healing.[5] Their implications for cancer are complicated. Treg cells tend to be upregulated in individuals with cancer, and they seem to be recruited to the site of many tumors. Studies in both humans and animal models have implicated that high numbers of Treg cells in the tumor microenvironment is indicative of a poor prognosis, and Treg cells are thought to suppress tumor immunity, thus hindering the body's innate ability to control the growth of cancerous cells.[6] Immunotherapy research is studying how regulation of T cells could possibly be utilized in the treatment of cancer.[7]

Another regulatory T cell subset is Treg17 cells Induced regulatory T cells Induced regulatory T (iTreg) cells (CD4+ CD25+ FOXP3+) are suppressive cells involved in tolerance. iTreg cells have been shown to suppress T cell proliferation and experimental autoimmune diseases. These cells include Treg17 cells.

IFNy_CD8: Interferon gamma (IFN-) is a dimerized soluble cytokine that is the only member of the type II class of interferons.[5] The existence of this interferon, which early in its history was known as immune interferon, was described by E. F. Wheelock as a product of human leukocytes stimulated with phytohemagglutinin, and by others as a product of antigen-stimulated lymphocytes.[6

IFN- , or type II interferon, is a cytokine that is critical for innate and adaptive immunity against viral, some bacterial and protozoan infections. IFN- is an important activator of macrophages and inducer of major histocompatibility complex class II molecule expression. Aberrant IFN- expression is associated with a number of autoinflammatory and autoimmune diseases. The importance of IFN- in the immune system stems in part from its ability to inhibit viral replication directly, and most importantly from its immunostimulatory and immunomodulatory effects. IFN- is produced predominantly by natural killer cells

(NK) and natural killer T cells (NKT) as part of the innate immune response, and by CD4 Th1 and CD8 cytotoxic T lymphocyte (CTL) effector T cells once antigen-specific immunity develops[11][12] as part of the adaptive immune response. IFN- is also produced by non-cytotoxic innate lymphoid cells (ILC), a family of immune cells first discovered in the early 2010s.[13]

IFN- is secreted by T helper cells (specifically, Th1 cells), cytotoxic T cells (TC cells), macrophages, mucosal epithelial cells and NK cells. IFN- is both an important autocrine signal for professional APCs in early innate immune response, and an important paracrine signal in adaptive immune response. The expression of IFN- is induced by the cytokines IL-12, IL-15, IL-18, and type I IFN.[20] IFN- is the only Type II interferon and it is serologically distinct from Type I interferons; it is acid-labile, while the type I variants are acid-stable.

IFN- has antiviral, immunoregulatory, and anti-tumor properties.[21] It alters transcription in up to 30 genes producing a variety of physiological and cellular responses. Among the effects are:

Promotes NK cell activity[22] Increases antigen presentation and lysosome activity of macrophages. Activates inducible nitric oxide synthase (iNOS) Induces the production of IgG2a and IgG3 from activated plasma B cells Causes normal cells to increase expression of class I MHC molecules as well as class II MHC on antigen-presenting cells—to be specific, through induction of antigen processing genes, including subunits of the immunoproteasome (MECL1, LMP2, LMP7), as well as TAP and ERAAP in addition possibly to the direct upregulation of MHC heavy chains and B2-microglobulin itself Promotes adhesion and binding required for leukocyte migration Induces the expression of intrinsic defense factors—for example, with respect to retroviruses, relevant genes include TRIM5alpha, APOBEC, and Tetherin, representing directly antiviral effects Primes alveolar macrophages against secondary bacterial infections.[23][24] IFN- is the primary cytokine that defines Th1 cells: Th1 cells secrete IFN- , which in turn causes more undifferentiated CD4+ cells (Th0 cells) to differentiate into Th1 cells ,[25] representing a positive feedback loop—while suppressing Th2 cell differentiation. (Equivalent defining cytokines for other cells include IL-4 for Th2 cells and IL-17 for Th17 cells.)

NK cells and CD8+ cytotoxic T cells also produce IFN- . IFN- suppresses osteoclast formation by rapidly degrading the RANK adaptor protein TRAF6 in the RANK-RANKL signaling pathway, which otherwise stimulates the production of NF- B.

Th1: They help to eradicate infections by microbes hat can survive or replicate within macrophages. Examples include certain viruses, protozoans and intracellular bacteria. If a th1 cell recognizes microbes displayed on the surface of an infected macrophagem it will activate the macrophage further through the release of IFN-g which enhances the macrophages microbicidal activity to kill ingested microbes. Type 1 responses also promote B cell class switching that favors production of opsonizing IgG antibodies.

IFNy_CD4: IFNy producing cd4 cells (Look at the above)

Treg: Look at treg

Th17: T helper 17 cells (Th17) are a subset of pro-inflammatory T helper cells defined by their production of interleukin 17 (IL-17). They are related to T regulatory cells and the signals that cause Th17s to differentiate actually inhibit Treg differentiation.[1] However, Th17s are developmentally distinct from Th1 and Th2 lineages. Th17 cells play an important role in maintaining mucosal barriers and contributing to pathogen clearance at mucosal surfaces; such protective and non-pathogenic Th17 cells have been termed as Treg17 cells.[2] Treg17 (Regulatory Th17) cells are generated from CD4+ T cells.

Transforming growth factor beta (TGF-), interleukin 6 (IL-6), interleukin 21 (IL-21) and interleukin 23 (IL-23) contribute to Th17 formation in mice and humans. Key factors in the differentiation of Th17 cells are signal transducer and the activator of transcription 3 (Stat3) and retinoic acid receptor-related orphan receptors gamma (ROR) and alpha (ROR).[3] Th17 cells are differentiated when naive T cells are exposed to the cytokines mentioned above. These cytokines are produced by activated antigen presenting cells (APCs) after contact with pathogens.[4] The Th17 cells can alter their differentiation program ultimately giving rise to either protective or pro-inflammatory pathogenic cells. The protective and non-pathogenic Th17 cells induced by IL-6 and TGF- are termed as Treg17 cells. The pathogenic Th17 cells are induced by IL-23 and IL-1 .[5] IL-21, produced by Th17 cells themselves, has also been shown to initiate an alternative route for

the activation of Th17 populations.[6] Both interferon gamma (IFN) and IL-4, the main stimulators of Th1 and Th2 differentiation, respectively, have been shown to inhibit Th17 differentiation

Dividing cells: The acute adaptive immune response is complex, proceeding through phases of activation of quiescent lymphocytes, rapid expansion by cell division and cell differentiation, cessation of division and eventual death of greater than 95 % of the newly generated population. Control of the response is not central but appears to operate as a distributed process where global patterns reliably emerge as a result of collective behaviour of a large number of autonomous cells.

The immune system of vertebrates is markedly complex. Broadly, the system is organised into innate and adaptive components with the former characterized by rapid responses triggered by common features of infectious organisms (Medzhitov and Janeway 1997), while the latter is responsible for slower but highly efficient pathogen-specific responses and the generation of long-lived protection and memory. The strength and speed of the adaptive immune response is in large part determined by the sheer number of different pathogen-specific B and T lymphocytes that must be made and maintained. To pre-empt the possibility of infection by millions of possible pathogens it is impossible to code for a target-specific clone for each possible pathogen in the germ-line. Instead, DNA coding for receptors is shuffled uniquely in each cell to create a broad range of specificities combinatorially as predicted theoretically by Burnet (1957). Then, during the acute immune response, the small number of specific pathogen-specific lymphocyte clones are selected, activated and expanded many fold by cell division resulting in cell numbers sufficiently large to clear the pathogen (Fig. 1) (De Boer and Perelson 2013). As such, cell division is one of the key mechanisms behind a successful immune response via an on-demand expansion of a pathogen-specific clone.

Div_treg:

Il17A_CD4: Il17A_CD4 Interleukin-17A is a protein that in humans is encoded by the IL17A gene. In rodents, IL-17A used to be referred to as CTLA8, after the similarity with a viral gene (O40633). Function The protein encoded by this gene is a proinflammatory cytokine produced by activated T cells. This cytokine regulates the activities of NF-kappaB and mitogen-activated protein kinases. This cytokine can stimulate the expression of IL6 and cyclooxygenase-2 (PTGS2/COX-2), as well as enhance the production of nitric oxide (NO).
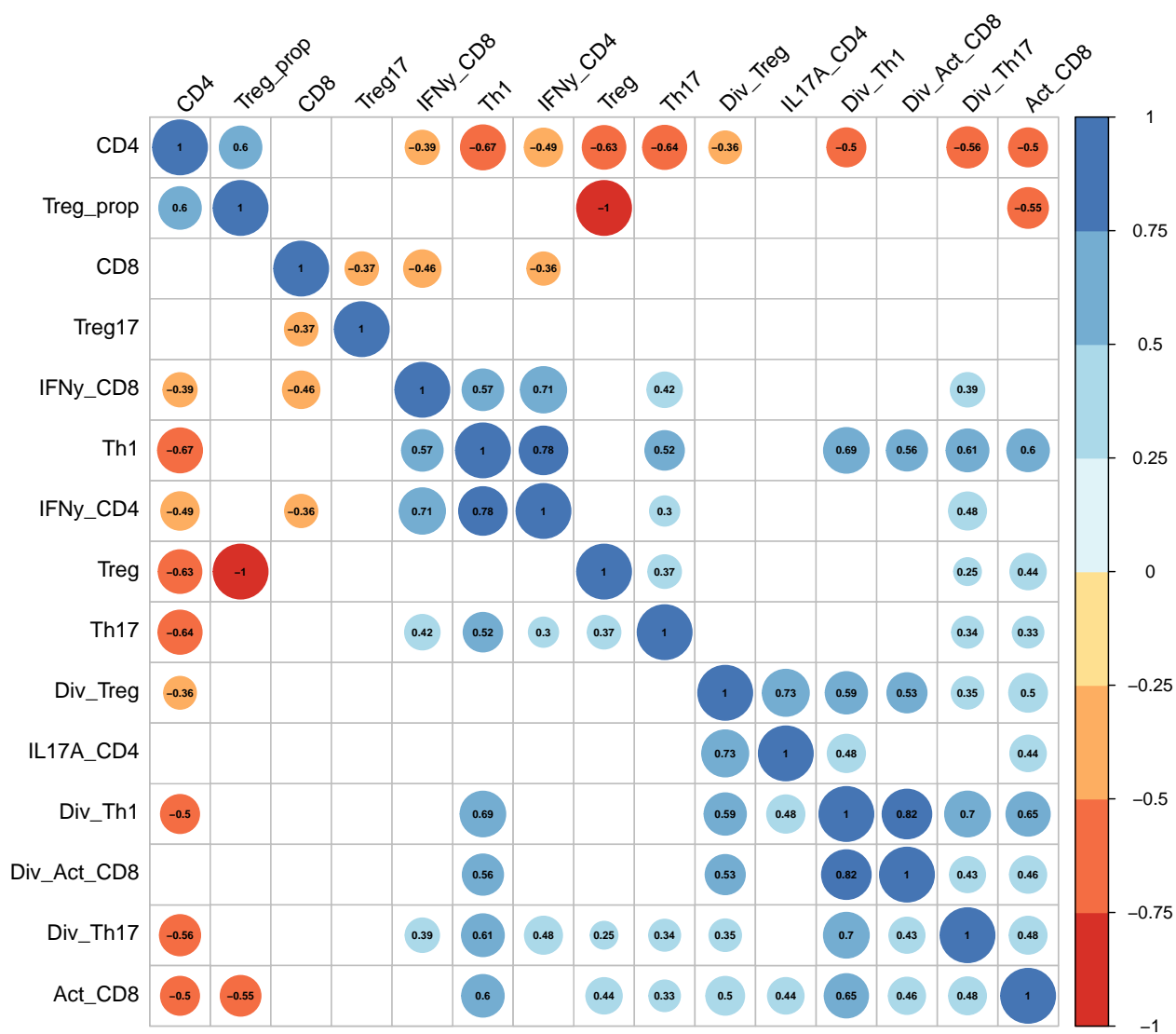
Div_Th1: Div_Act_CD8: Div_Th17: Act_CD8:

1. facs expression in the laboratory infections - Heatmap
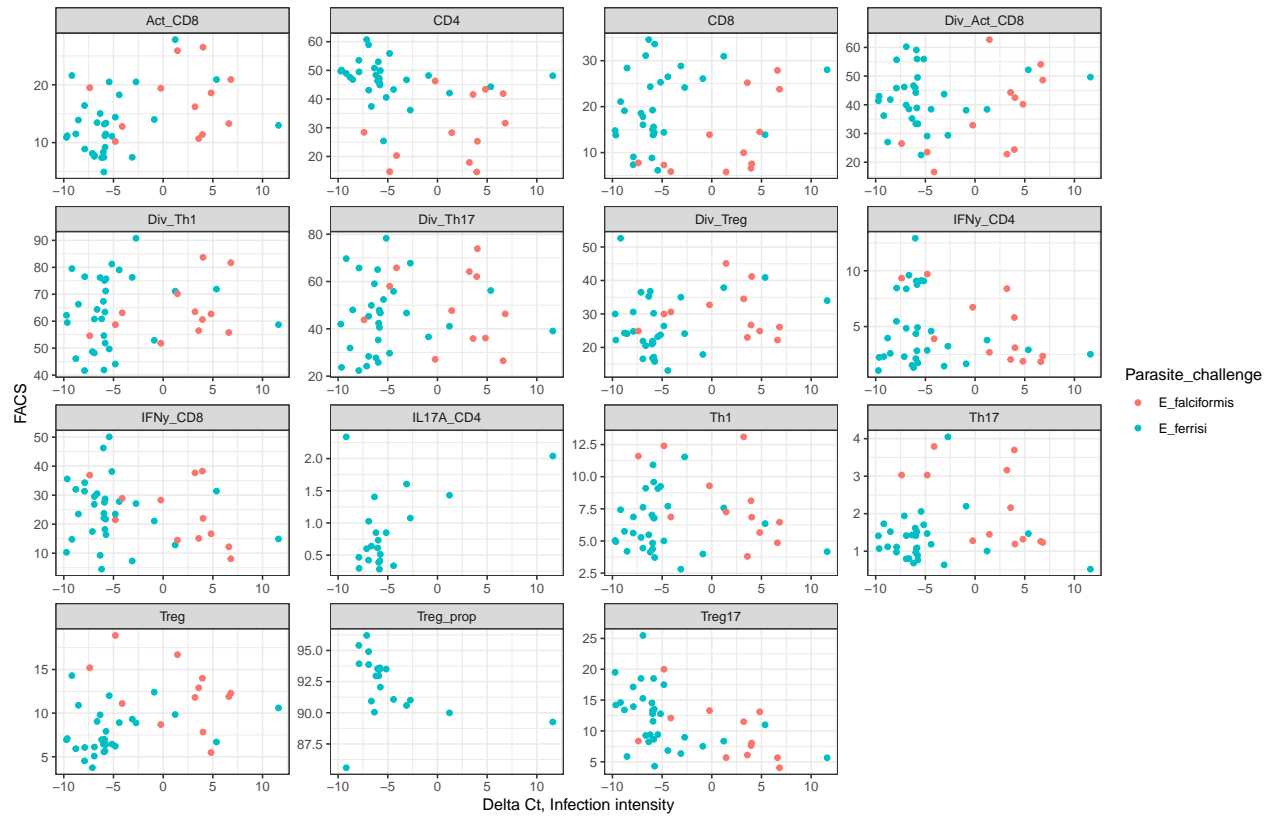
FACS

```
heatmap_data %>%
  pheatmap(annotation_col = annotation_df, scale = "row")
```
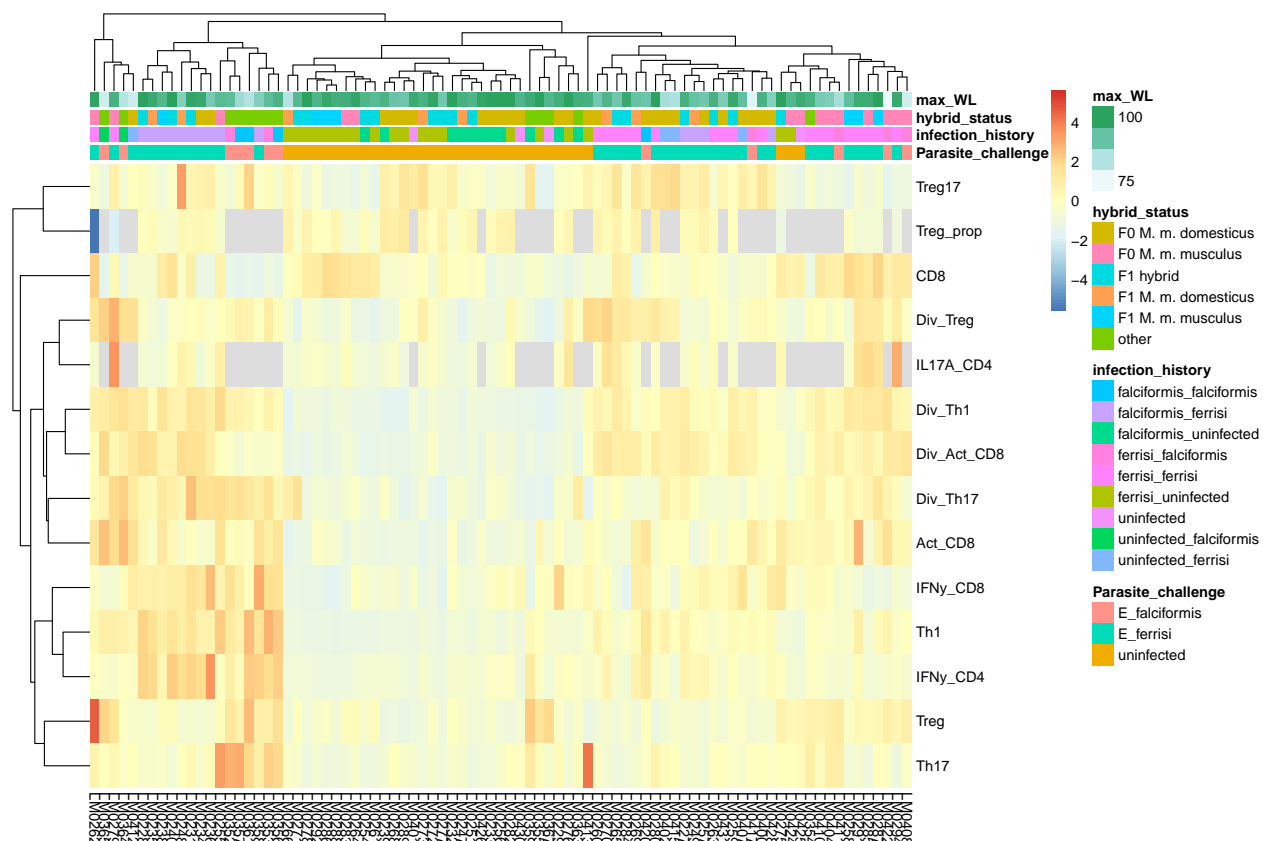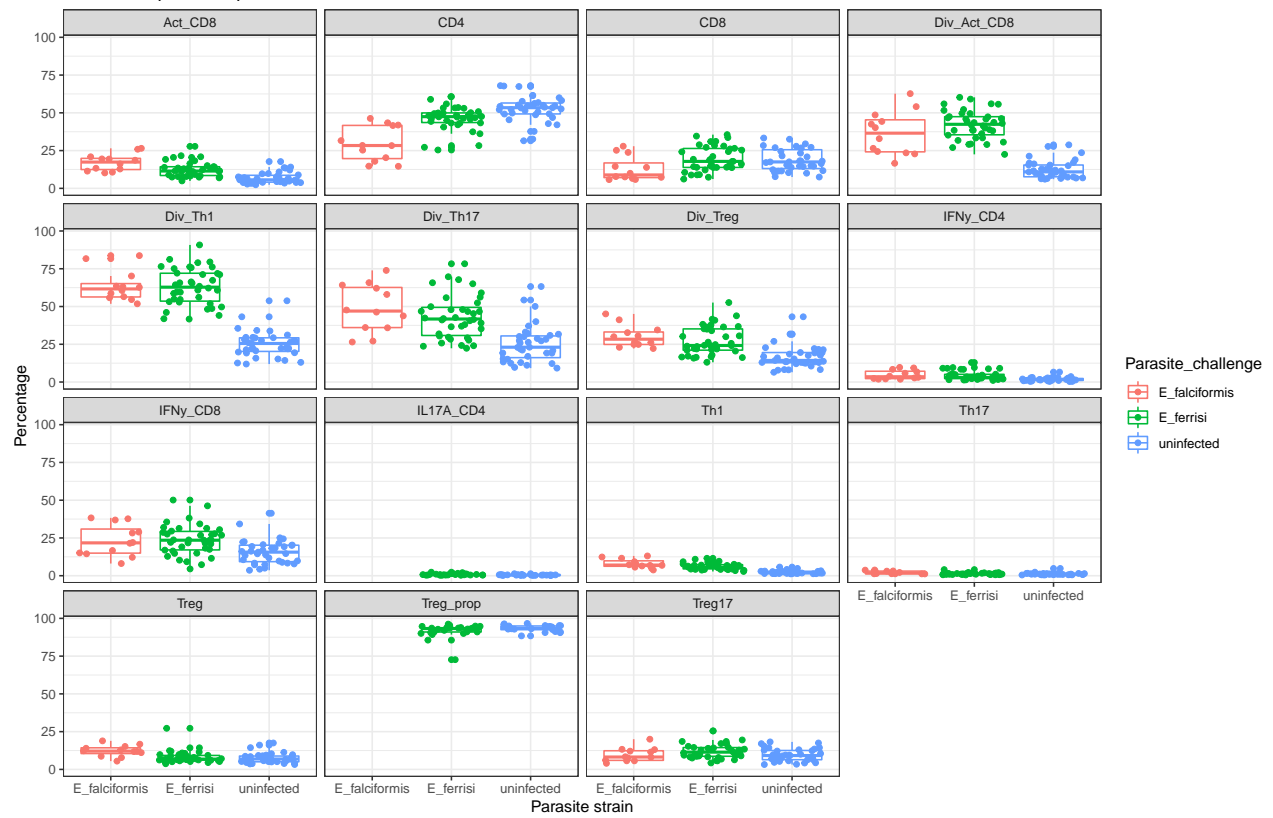
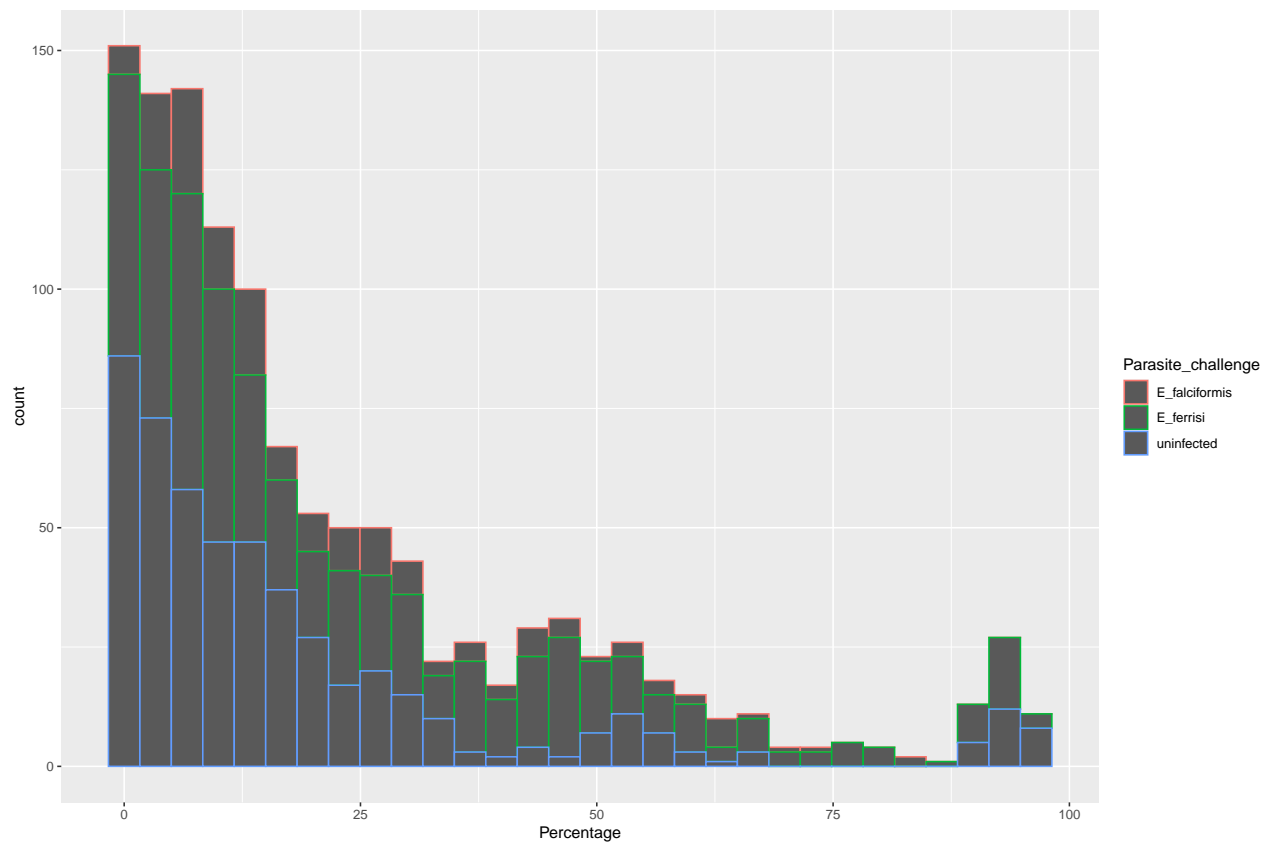## 2. Correlations between the cells
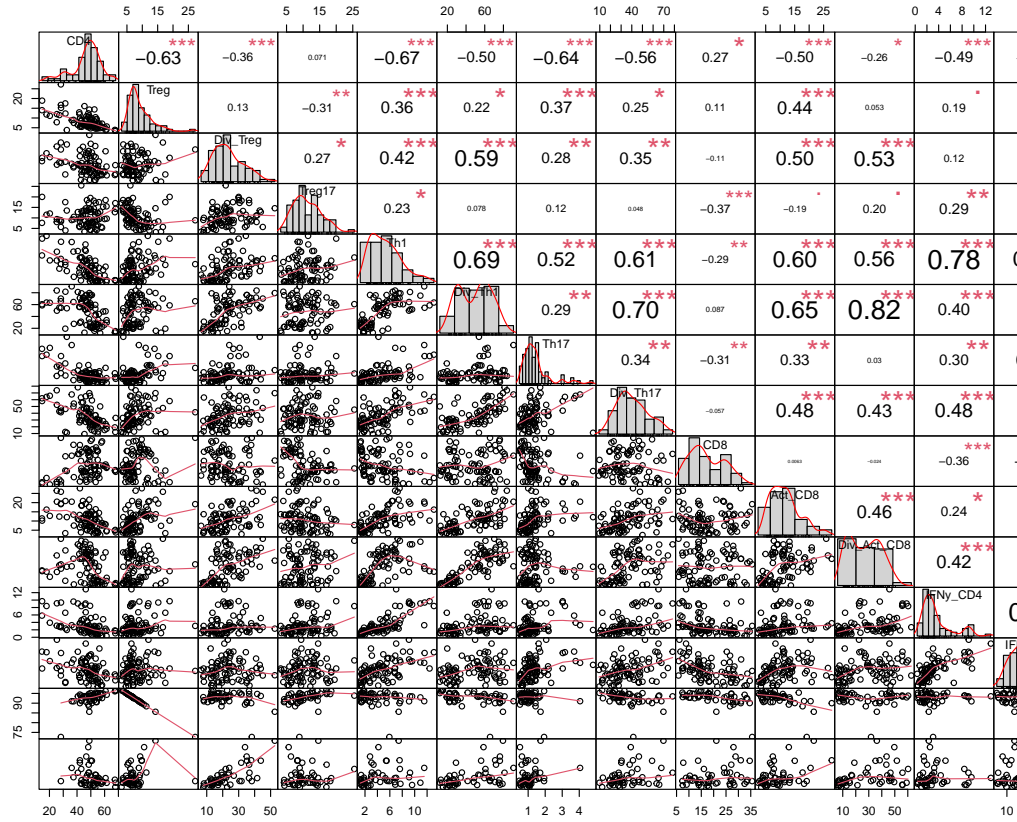
Cell counts in response to infection intensity

Cells in response to parasite strain

3. PCA

Handling missing data in a pca:

We will now continue by using an iterative pca to impute missing data A. Initialization: impute using the mean B. Step lampda: # a. do pca on imputed data table S dimensions retained # b. missing data imputed using pca # c. means (and standard deviations) updated C. Iterate the estimation and imputation steps (until convergence) (convergence: the act of converging and especially moving toward union or uniformity)

Overfitting is a common problem due to believing too much in links between variables. –> regularized iterative PCA (This version is what is being implented in missMDA) This is a way of taking less risk when imputing the missing data. The algorithm estimates the missing data values with values that have no influence on the PCA results, i.e., no influence on the coordinates of the individals or variables.

PCA graph of individuals
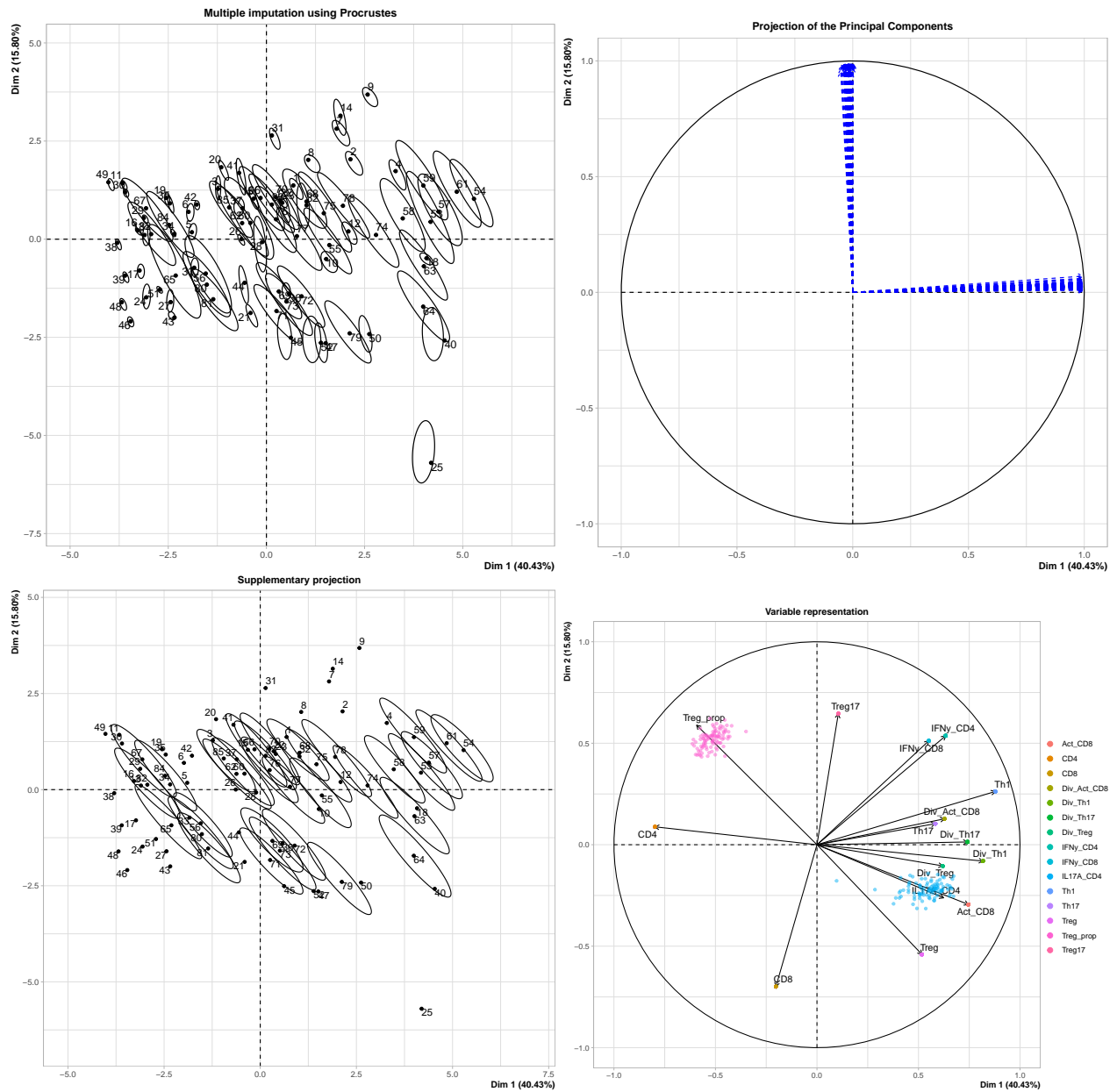
PCA graph of variables
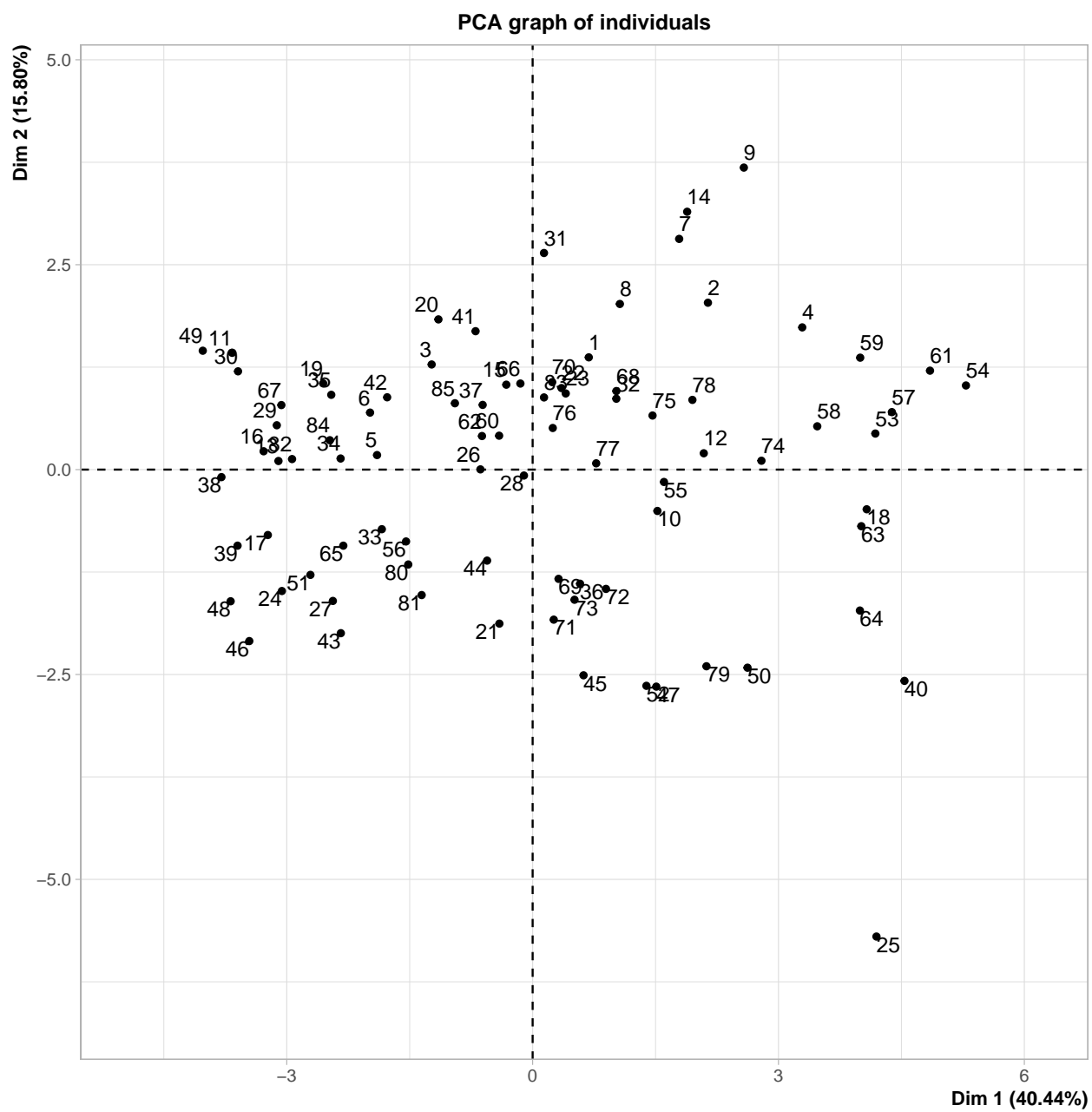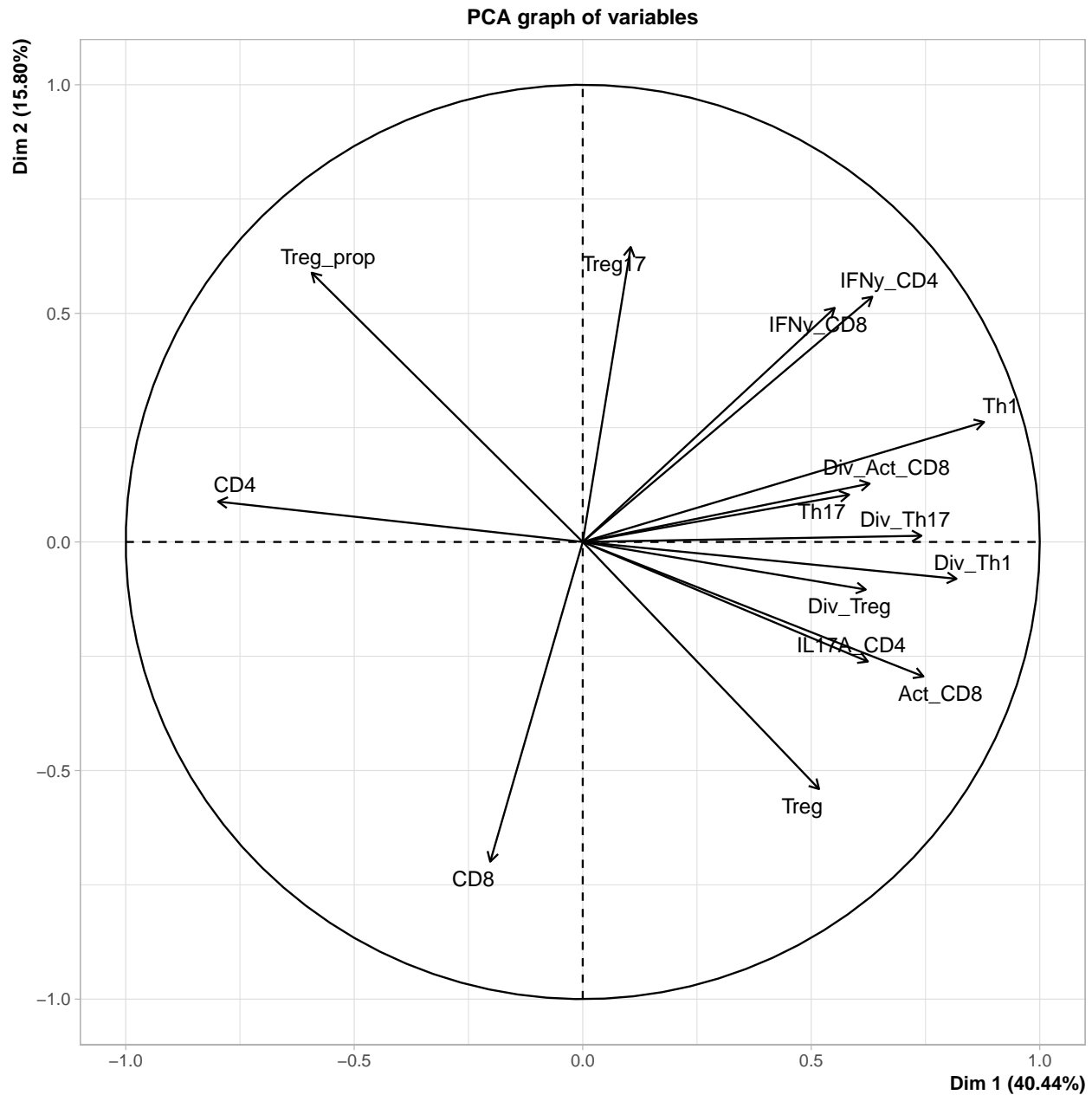
Caution: When imputing data, the percentages of inertia associated with the first dimensions will be over-estimated.

Another problem: the imputed data are, when the pca is performed considered like real observations. But they are estimations!!
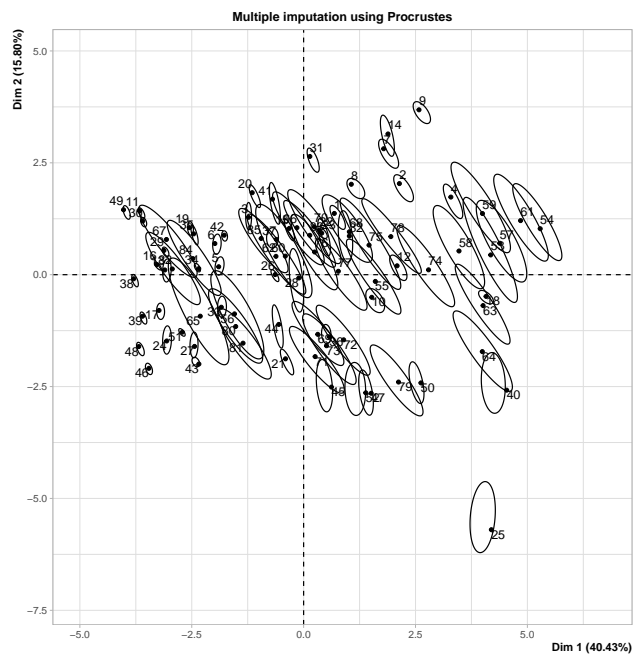
Visualizing uncertainty due to issing data:

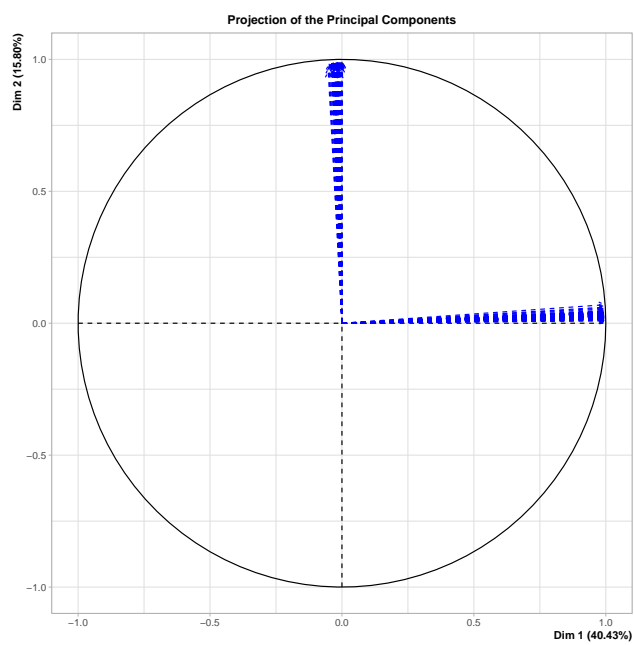–> mulrimple imputation: facsrate several plausible values for each missing data point

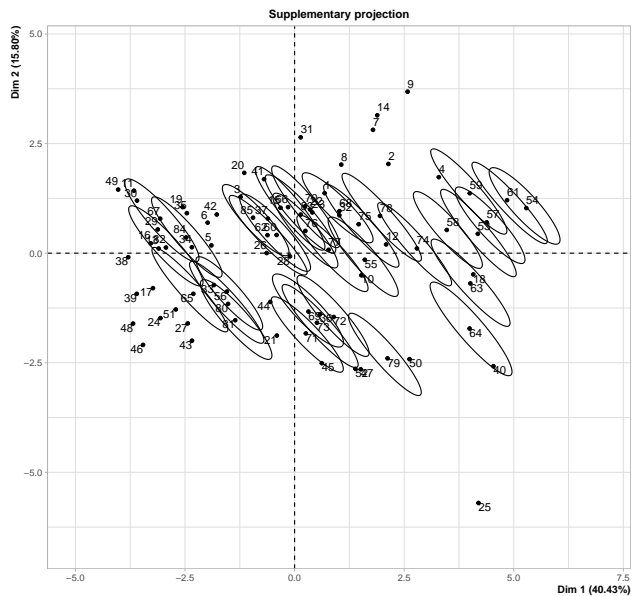We here visualize the variability, that is uncertainty on the plane defined by two pca axes.
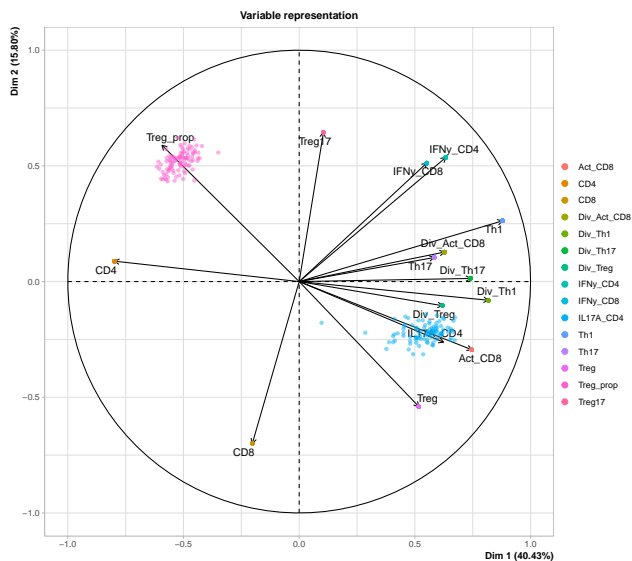
$PlotIndProc

Multiple imputation using Procrustes

##
$PlotDim

Projection of the Principal Components

##
$PlotIndSupp

Supplementary projection

$PlotVar

Variable representation

Individuals lying on the axis have no missing data, but individuals that far away have many missing data. big ellipse = big uncertainty tight elipse (line) = low uncertainty

Variable representation: Poins tight together)look like one) - have no missing variables –> low uncertainty Points spread – > higher variability – > higher uncertainty

High uncertainty–> we should interpret the result with care

The individuals with many missing data values make the axes move, and thus the positions of all individuals

Therefore in the last plots every individual is getting an eclipse as they are as well influenced by the missing data of the others.

THe plot with the dimensions shows the projections of the pca dimensions of each imputed table on the pca plane obtained using the original imputed data table

As all of the arrows are close to either the first or second axes, this means that the axes are stable with

respect to the set of imputed tables –> we don't have evidence of instability here.

```r
#Now we can make our initial plot of the PCA.


imputed_facs %>%
  pivot_longer(cols = all_of(CellCount.cols
                             ), names_to = "Cells", values_to = "Proportion") %>%
  ggplot(aes(x = pc1, y = pc2, color = Parasite_challenge, shape = Parasite_challenge)) +
  geom_hline(yintercept = 0, lty = 2) +
  geom_vline(xintercept = 0, lty = 2) +
  geom_point(alpha = 0.8) +
  stat_ellipse(geom="polygon", aes(fill = Parasite_challenge), alpha = 0.2, show.legend = FALSE,
               level = 0.95) +
  theme_minimal() +
  theme(panel.grid = element_blank(), panel.border = element_rect(fill= "transparent"))
```
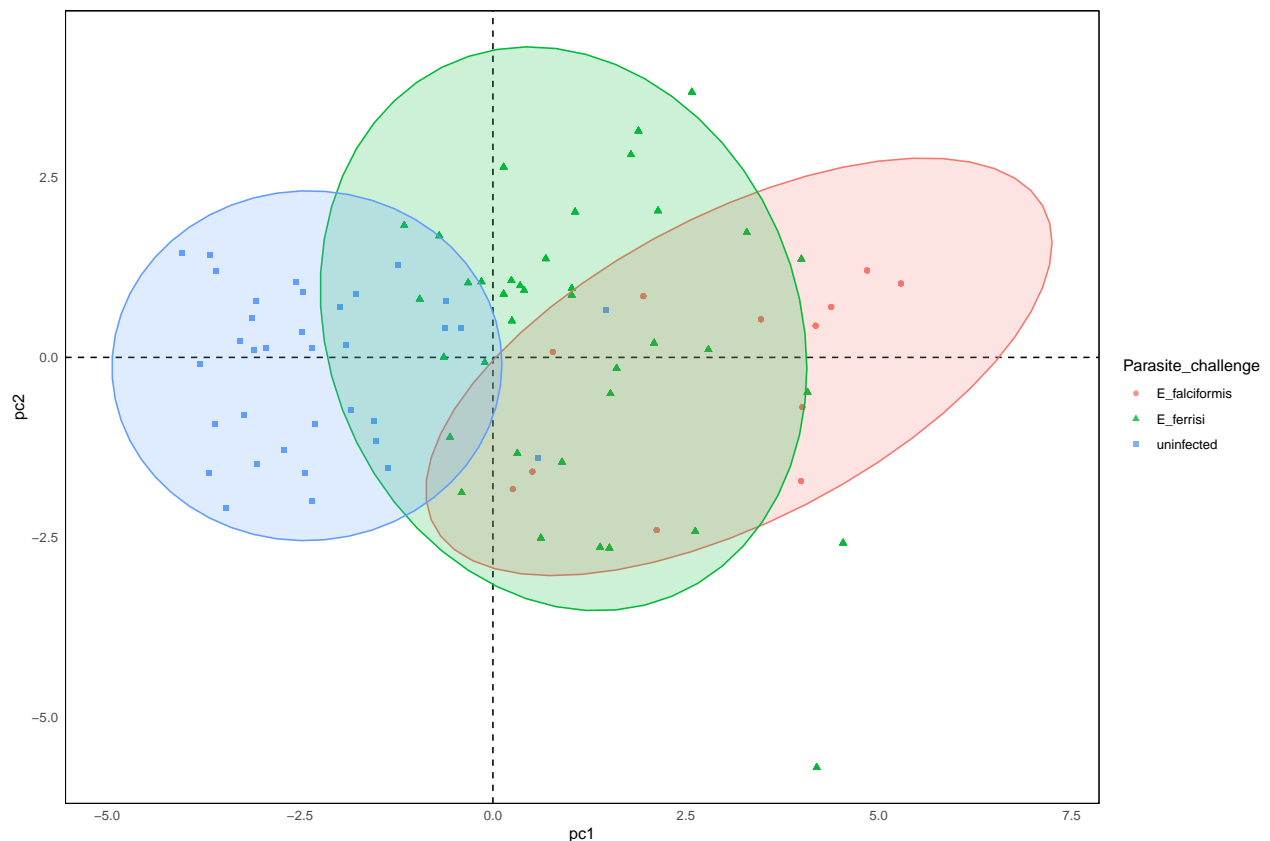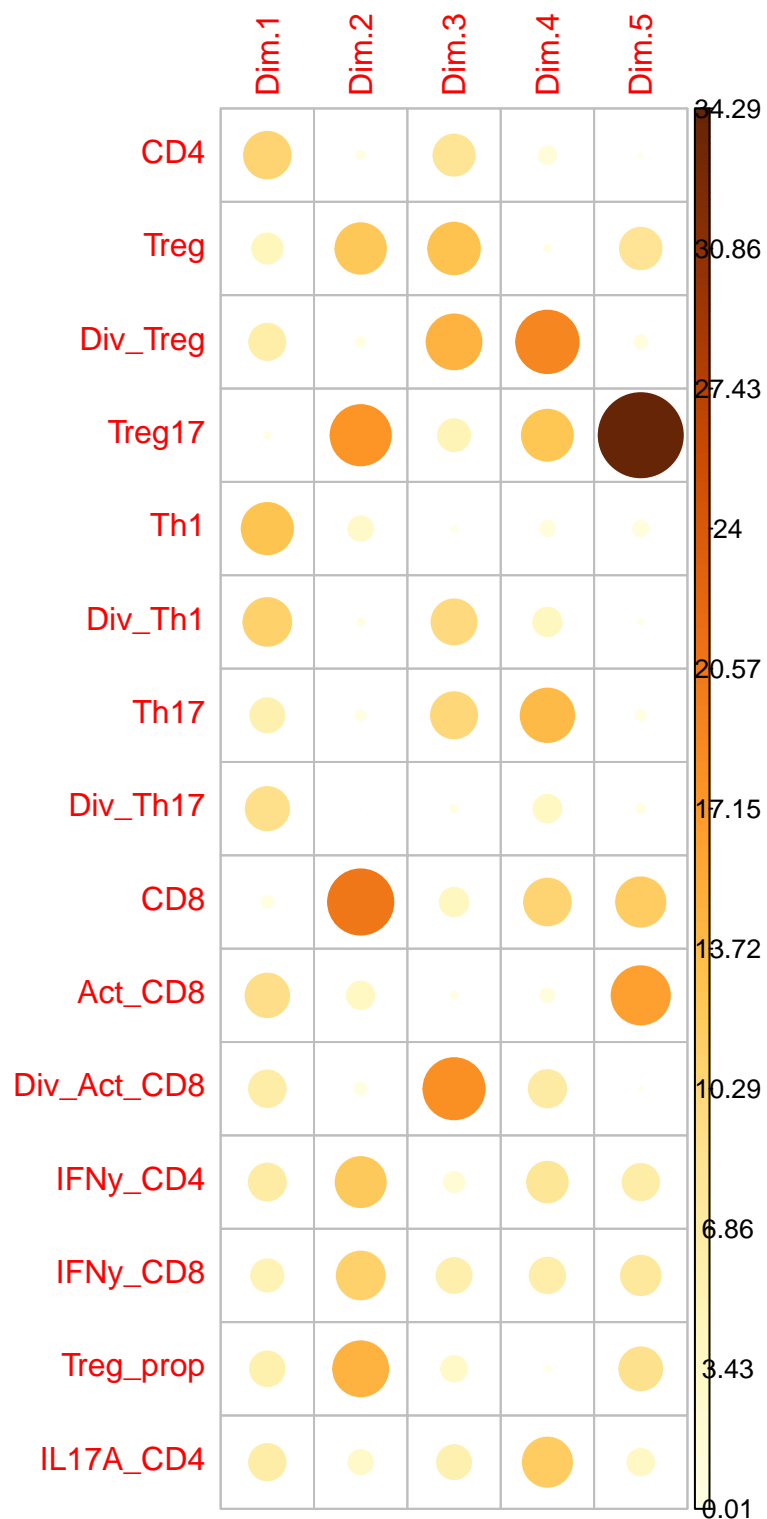
The function fviz_contrib() [factoextra package] can be used to draw a bar plot of variable contributions. If your data contains many variables, you can decide to show only the top contributing variables. The R code below shows the top 10 variables contributing to the principal components:

Contribution of variables to Dim−1

Contribution of variables to Dim−2

Contribution of variables to Dim−1−2
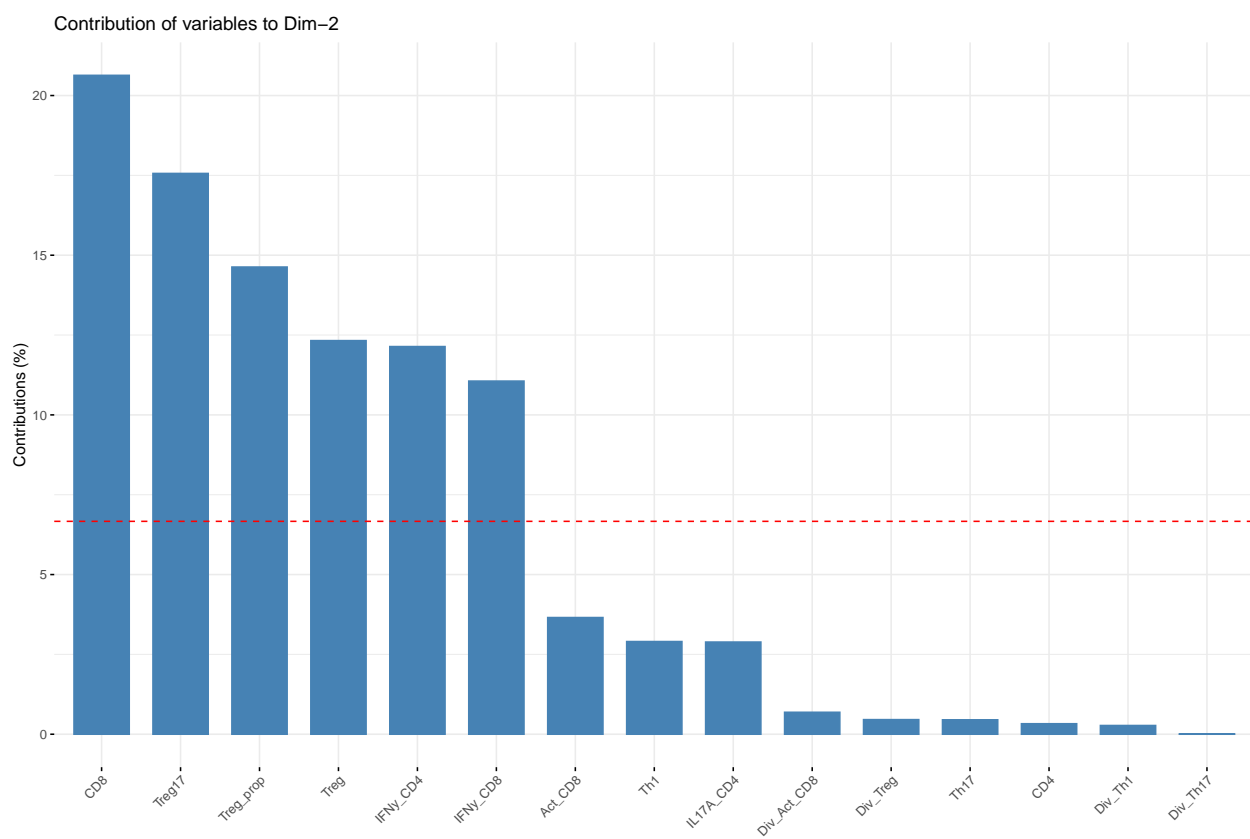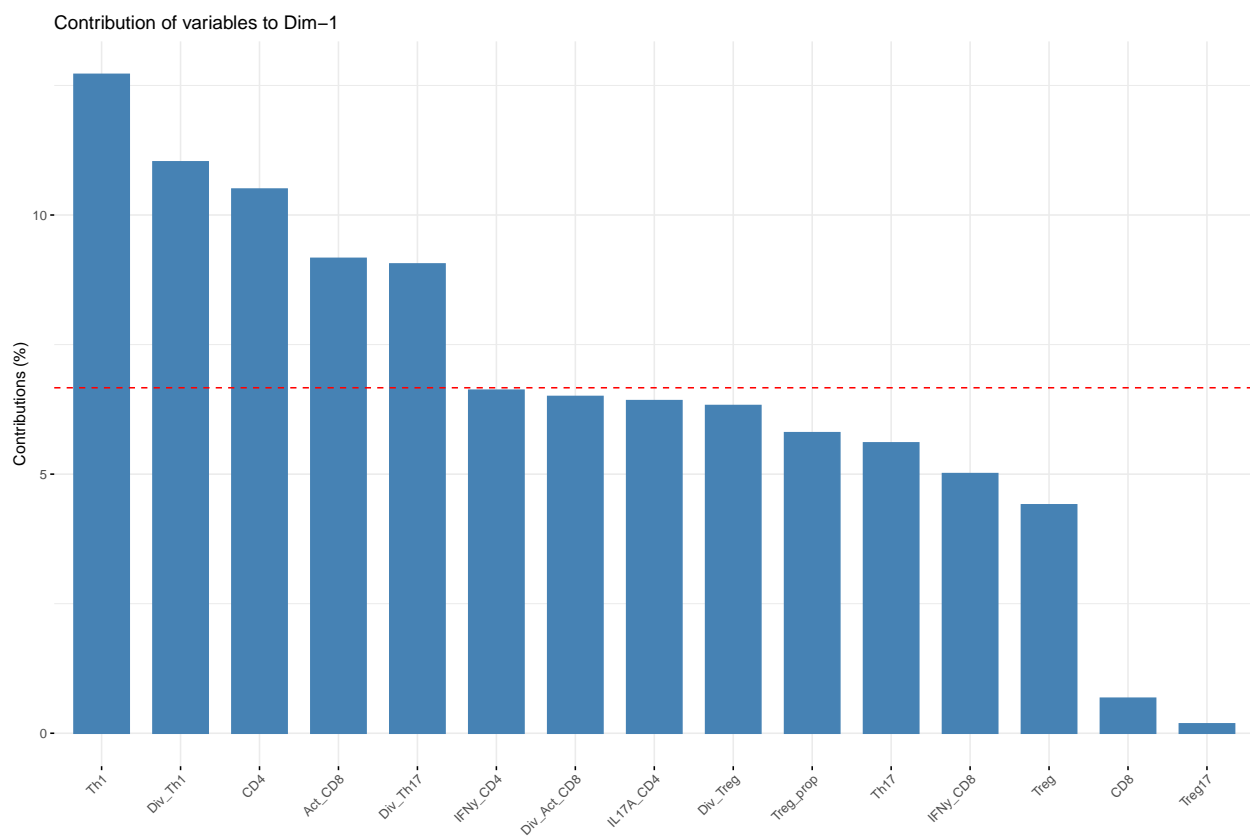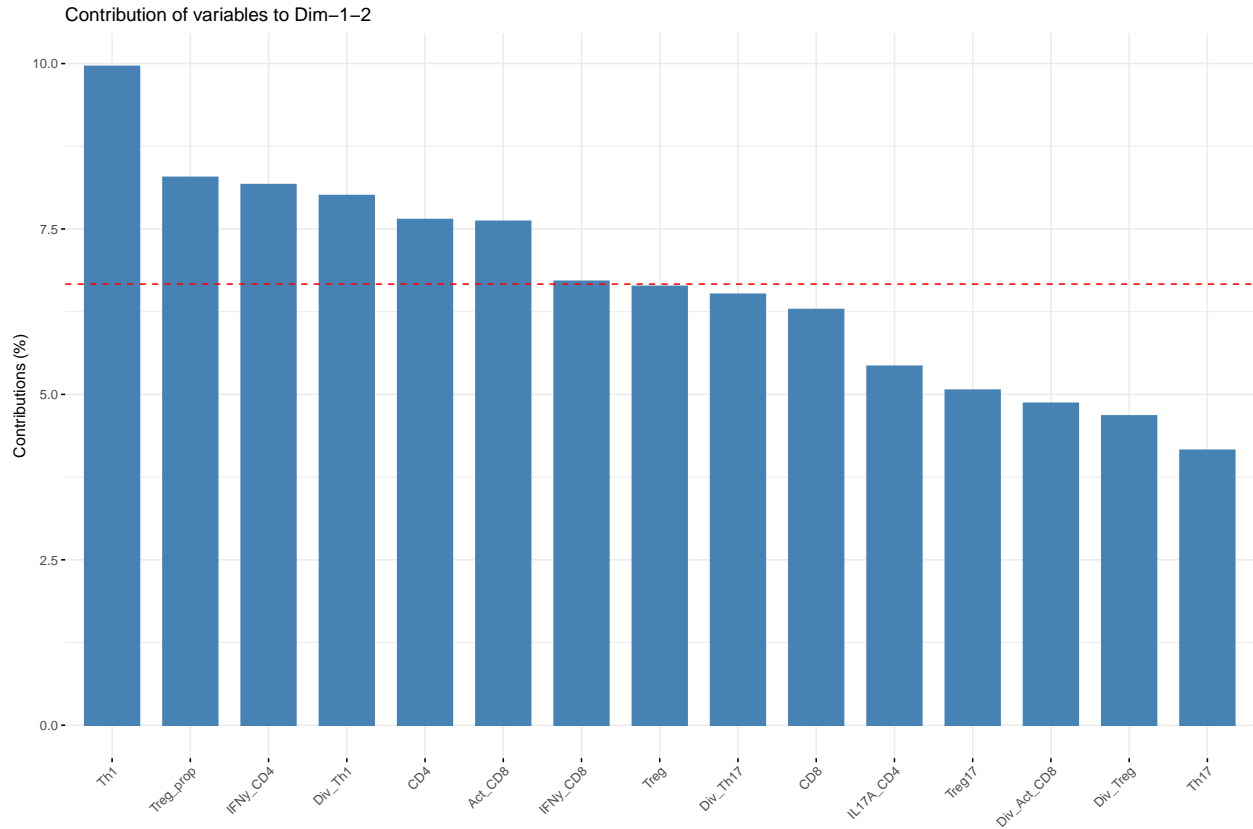
The red dashed line on the graph above indicates the expected average contribution. If the contribution of the variables were uniform, the expected value would be 1/length(variables) = 1/10 = 10%. For a given component, a variable with a contribution larger than this cutoff could be considered as important in contributing to the component.

Note that, the total contribution of a given variable, on explaining the variations retained by two principal components, say PC1 and PC2, is calculated as contrib = [(C1 * Eig1) + (C2 * Eig2)]/(Eig1 + Eig2), where

C1 and C2 are the contributions of the variable on PC1 and PC2, respectively Eig1 and Eig2 are the eigenvalues of PC1 and PC2, respectively. Recall that eigenvalues measure the amount of variation retained by each PC. In this case, the expected average contribution (cutoff) is calculated as follow: As mentioned above, if the contributions of the 10 variables were uniform, the expected average contribution on a given PC would be 1/10 = 10%. The expected average contribution of a variable for PC1 and PC2 is : [(10* Eig1) + (10 * Eig2)]/(Eig1 + Eig2)
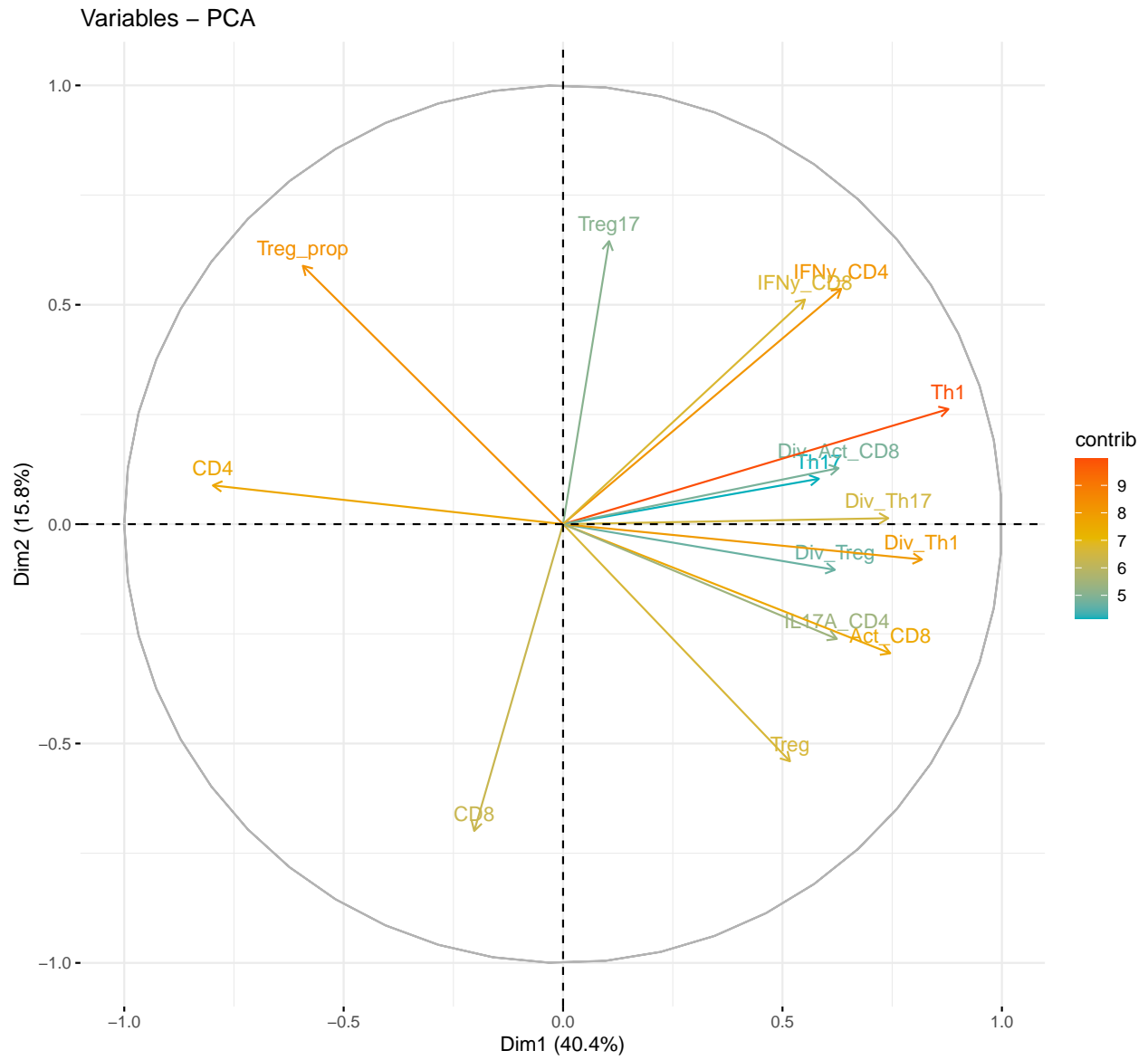
Variables – PCA

To visualize the contribution of individuals to the first two principal components:

Contribution of individuals to Dim-1-2
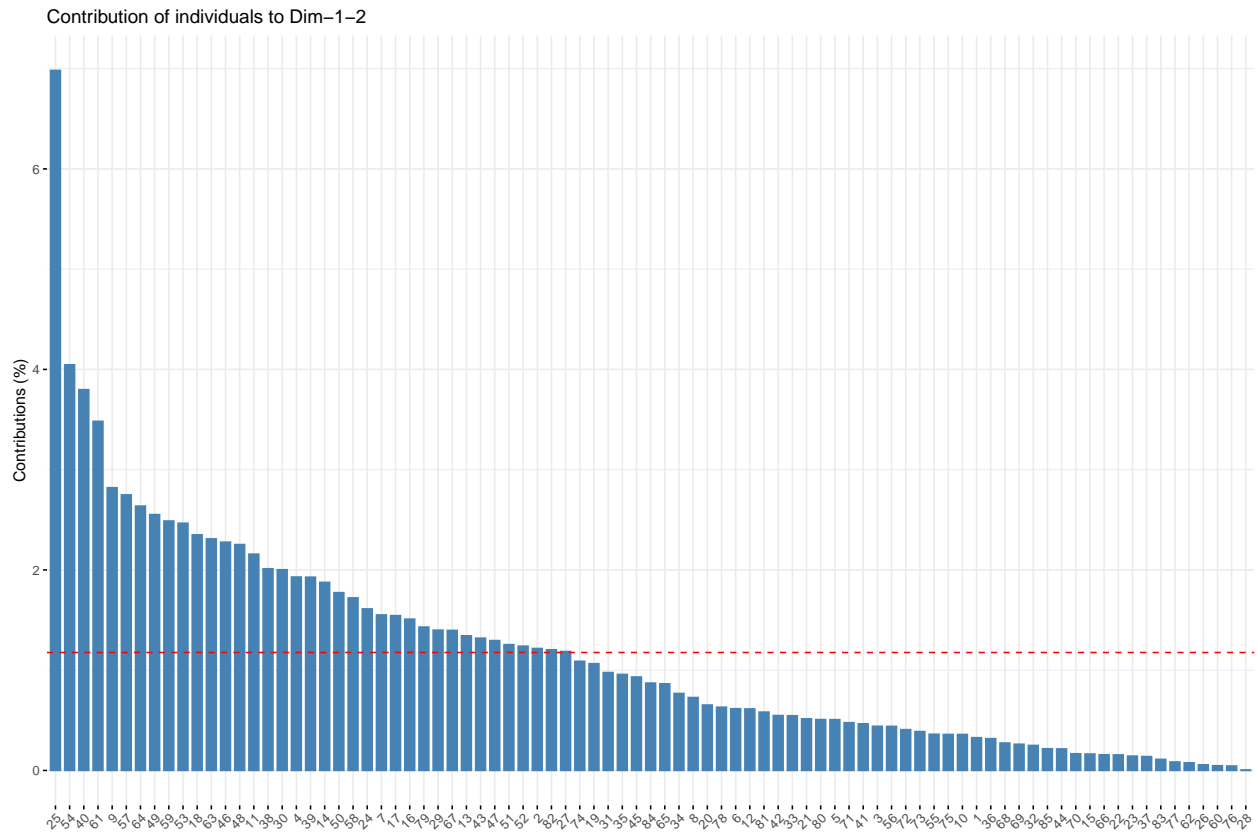
PCA + Biplot combination

PCA – Biplot
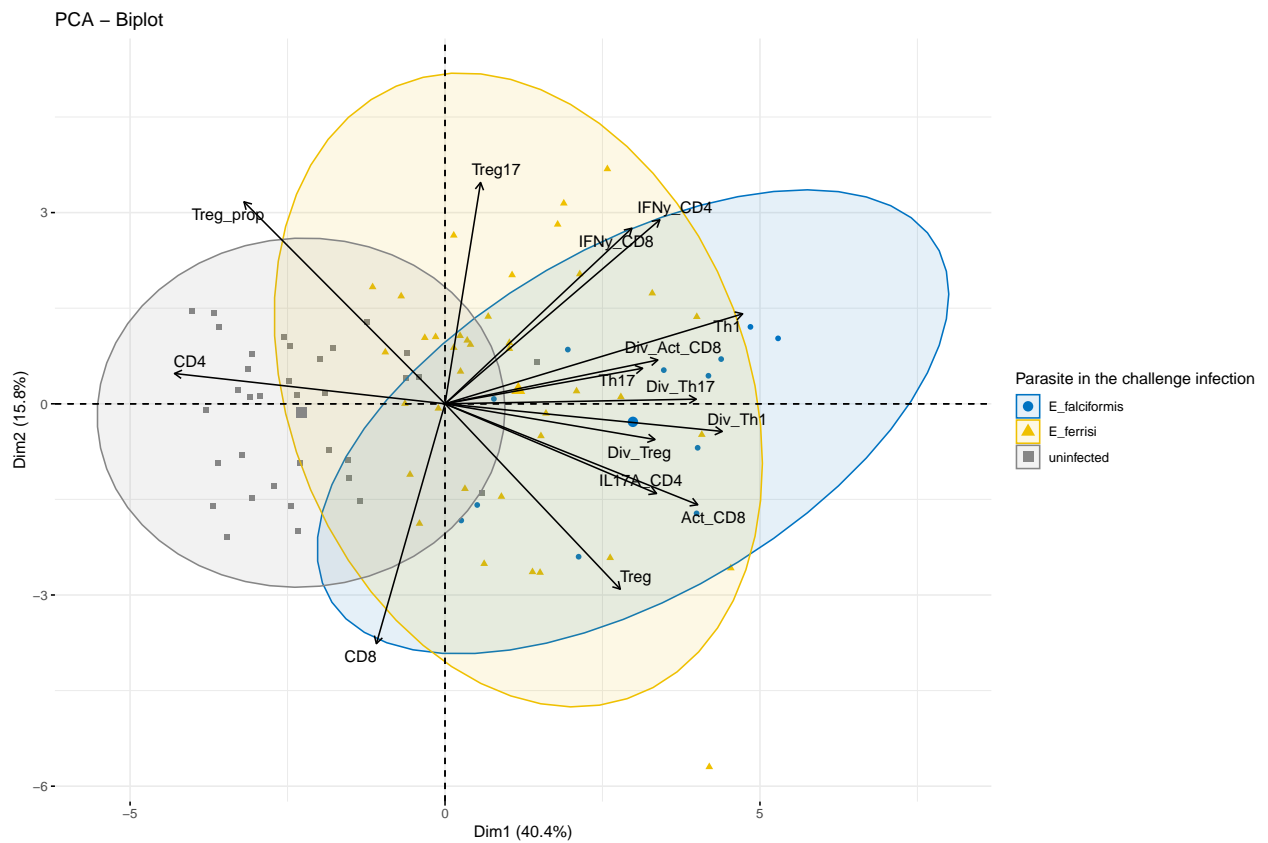
In the following example, we want to color both individuals and variables by groups. The trick is to use pointshape = 21 for individual points. This particular point shape can be filled by a color using the argument fill.ind. The border line color of individual points is set to "black" using col.ind. To color variable by groups, the argument col.var will be used.

Linear models:

```
##
## Call:
## lm(formula = max_WL ~ pc1 + pc2 + Parasite_challenge, data = imputed_facs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.5521  -2.9263   0.1456   3.6345   9.9540
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   80.9739     1.8237  44.401  < 2e-16 ***
## pc1                            0.9508     0.3772   2.521   0.0137 *
## pc2                           -0.1475     0.3542  -0.416   0.6782
## Parasite_challengeE_ferrisi   11.0501     1.7898   6.174 2.60e-08 ***
## Parasite_challengeuninfected  17.1202     2.5903   6.609 3.95e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.975 on 80 degrees of freedom
## Multiple R-squared:  0.4265, Adjusted R-squared:  0.3978
## F-statistic: 14.87 on 4 and 80 DF,  p-value: 3.976e-09

## [1] 520.8357

##
## Call:
## lm(formula = max_WL ~ pc1 + pc2 + Parasite_challenge + hybrid_status,
##     data = imputed_facs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.9842  -3.1816   0.5616   3.5633   9.4894
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    79.8812     2.1017  38.007  < 2e-16 ***
## pc1                             0.9103     0.4255   2.140   0.0356 *
## pc2                             0.3741     0.5915   0.633   0.5290
## Parasite_challengeE_ferrisi    10.6313     1.9281   5.514 4.76e-07 ***
## Parasite_challengeuninfected   16.6312     2.5658   6.482 8.55e-09 ***
## hybrid_statusF0 M. m. musculus  1.5746     2.5368   0.621   0.5367
## hybrid_statusF1 hybrid          3.5399     1.6592   2.134   0.0362 *
## hybrid_statusF1 M. m. domesticus -0.9093    2.0039  -0.454   0.6513
## hybrid_statusF1 M. m. musculus  4.2002     2.7063   1.552   0.1249
## hybrid_statusother              1.6583     2.1713   0.764   0.4474
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.91 on 75 degrees of freedom
```

```
## Multiple R-squared:  0.4763, Adjusted R-squared:  0.4134
## F-statistic: 7.578 on 9 and 75 DF,  p-value: 8.217e-08

## [1] 523.1115

##
## Call:
## lm(formula = max_WL ~ pc1 + pc2 + hybrid_status, data = imputed_facs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.0089  -3.8001   0.7976   4.3336  11.4052
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      91.3120     1.4723  62.020   <2e-16 ***
## pc1                              -0.6637     0.3125  -2.124   0.0369 *
## pc2                               0.4622     0.7284   0.635   0.5276
## hybrid_statusF0 M. m. musculus    1.6599     3.1596   0.525   0.6009
## hybrid_statusF1 hybrid            4.5068     2.0539   2.194   0.0312 *
## hybrid_statusF1 M. m. domesticus -0.5451     2.4849  -0.219   0.8270
## hybrid_statusF1 M. m. musculus    5.0932     3.3323   1.528   0.1305
## hybrid_statusother                0.7101     2.4448   0.290   0.7723
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.118 on 77 degrees of freedom
## Multiple R-squared:  0.1654, Adjusted R-squared:  0.08956
## F-statistic:  2.18 on 7 and 77 DF,  p-value: 0.04505

## [1] 558.7177

##
## Call:
## lm(formula = max_WL ~ pc1 + pc2 + infection_history, data = imputed_facs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.5834  -2.8158  -0.4039   3.2253   8.7078
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         84.39648    2.65405  31.799  < 2e-16
## pc1                                  0.98773    0.36148   2.732 0.007858
## pc2                                 -0.09311    0.38599  -0.241 0.810049
## infection_historyfalciformis_ferrisi  7.62628    2.63621   2.893 0.005011
## infection_historyfalciformis_uninfected 15.00027 3.40217   4.409 3.46e-05
## infection_historyferrisi_falciformis -4.18836    3.02105  -1.386 0.169791
## infection_historyferrisi_ferrisi      9.29476    2.78012   3.343 0.001301
## infection_historyferrisi_uninfected  13.15007    3.36444   3.909 0.000204
## infection_historyuninfected          13.65829    3.90798   3.495 0.000805
## infection_historyuninfected_falciformis -8.53966 3.95493  -2.159 0.034073
## infection_historyuninfected_ferrisi  -1.22495    3.30747  -0.370 0.712174
##
## (Intercept)                         ***
## pc1                                 **
```

```
## pc2
## infection_historyfalciformis_ferrisi      **
## infection_historyfalciformis_uninfected ***
## infection_historyferrisi_falciformis
## infection_historyferrisi_ferrisi          **
## infection_historyferrisi_uninfected       ***
## infection_historyuninfected               ***
## infection_historyuninfected_falciformis *
## infection_historyuninfected_ferrisi
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.48 on 74 degrees of freedom
## Multiple R-squared:  0.5699, Adjusted R-squared:  0.5117
## F-statistic: 9.804 on 10 and 74 DF,  p-value: 3.216e-10

## [1] 508.3793

##
## Call:
## lm(formula = max_WL ~ pc1 + pc2, data = imputed_facs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.168  -3.267   1.382   4.218   8.878
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  92.9634     0.6727 138.191  < 2e-16 ***
## pc1          -0.7582     0.2732  -2.776  0.00682 **
## pc2          -0.1078     0.4370  -0.247  0.80584
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.202 on 82 degrees of freedom
## Multiple R-squared:  0.08651,    Adjusted R-squared:  0.06423
## F-statistic: 3.883 on 2 and 82 DF,  p-value: 0.02448

##                    df      AIC
## weight_lm           6 520.8357
## weight_lm_exp_only  4 556.3975
```

repeating the heatmap on the now imputed data

```
facs <- imputed_facs %>%
  dplyr::select(c(EH_ID, all_of(CellCount.cols)))

 # turn the data frame into a matrix and transpose it. We want to have each cell
 # type as a row name
 facs <- t(as.matrix(facs))

 #switch the matrix back to a data frame format
 facs <- as.data.frame(facs)

 # turn the first row into column names
```

```
facs %>%
    row_to_names(row_number = 1) -> heatmap_data


table(rowSums(is.na(heatmap_data)) == nrow(heatmap_data))

##
## FALSE
##    15
```

```
# turn the columns to numeric other wise the heatmap function will not work
heatmap_data[] <- lapply(heatmap_data, function(x) as.numeric(as.character(x)))
# remove columns with only NAs
heatmap_data <- Filter(function(x)!all(is.na(x)), heatmap_data)

#remove rows with only Nas
heatmap_data <-  heatmap_data[, colSums(is.na(heatmap_data)) != nrow(heatmap_data)]
rownames(annotation_df) <- colnames(heatmap_data)
```

Heatmap on facs expression data: