

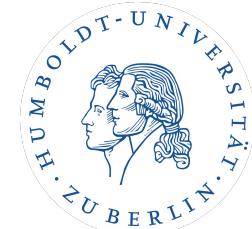


# Resistance and tolerance against infections in hybrid mice

Using immune parameters to predict the impact of  
hybridization on health

**Fay Webster**

Emanuel Heitlinger's Lab



# Outline

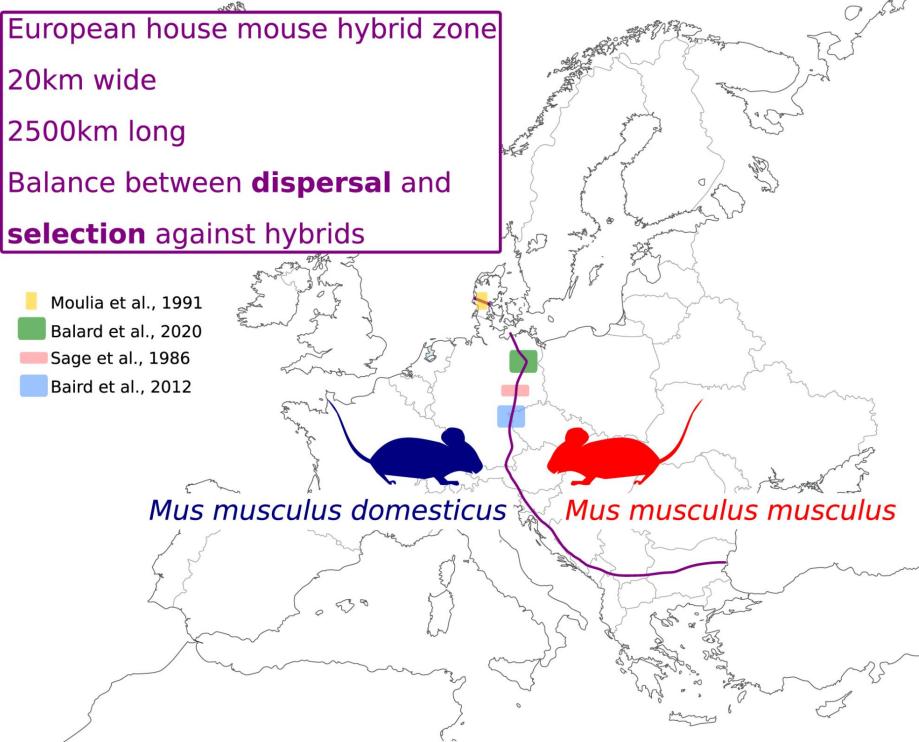
- Introduction
- Research question - Goals
- Methods
  - Experimental design
  - **Imputation with mice**
  - Exploratory analysis
  - Random forest model
- Results
- Discussion



# Introduction

# 1. Introduction:

The european house mouse hybrid zone

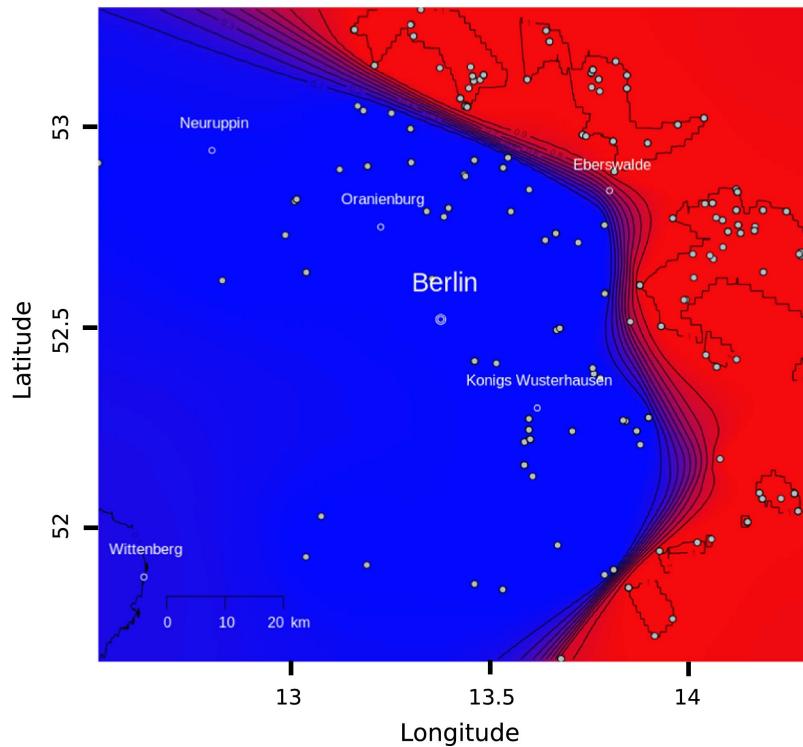


Balard et al., 2019

# Study area:

***M. musculus  
domesticus***

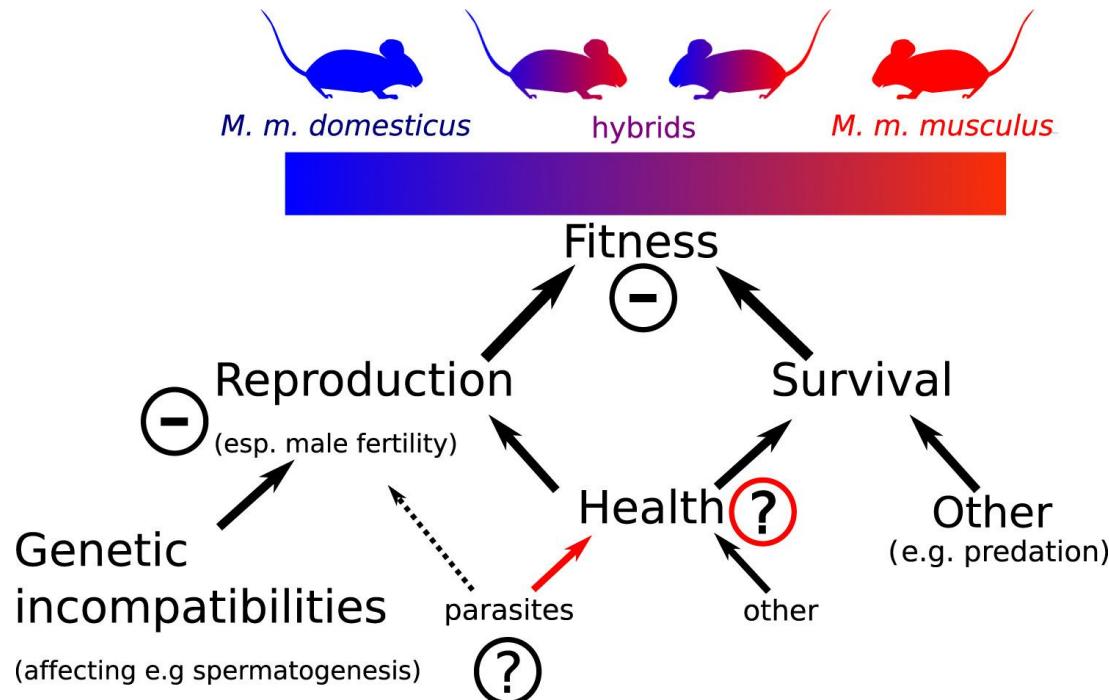
Map of posterior probability to belong to Mmm



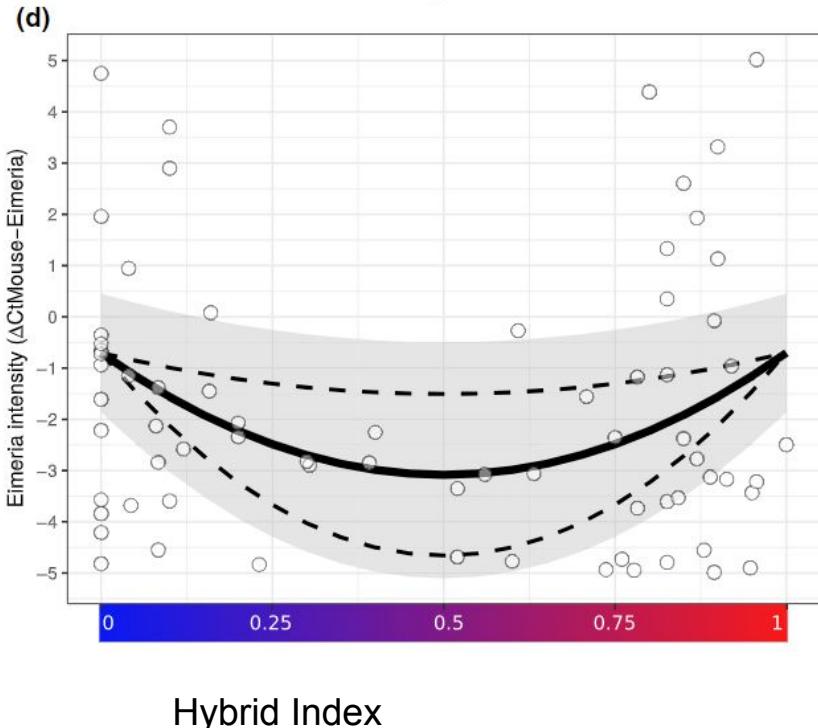
***M. musculus  
musculus***

Balard et al., 2019

# Fitness components of hybrid mice



# Intensity of infections with *Eimeria* are reduced in hybrids



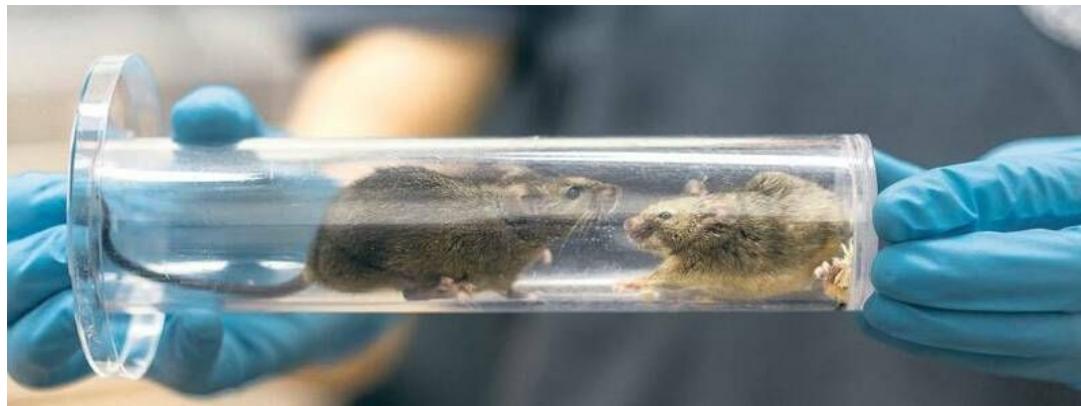
Graph: Balard et al., 2019

Method: Baird et al., 2012

# Research question:

# **Research question:**

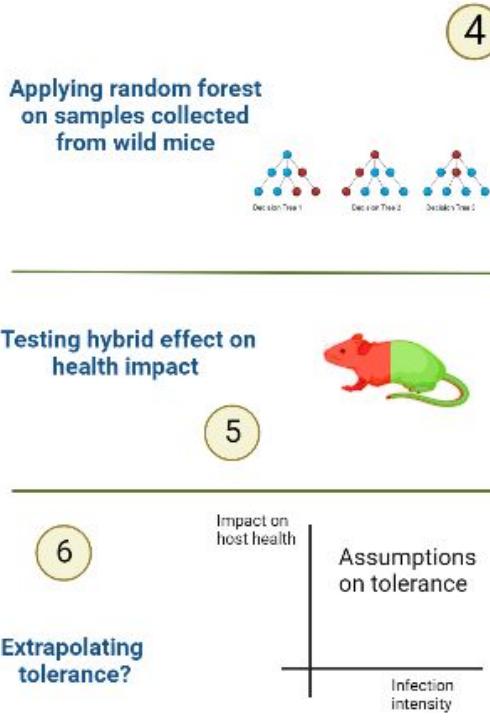
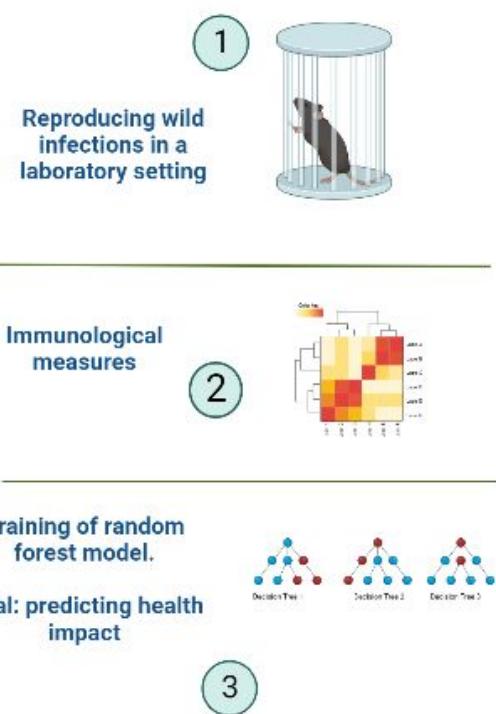
**Is tolerance modulating the fitness of hybrids?**



**Objective:**  
**Predicting health impact: An approach  
using commonly used immune parameters**

# Methods

# Graphical Abstract:



## 2. Methods - Experimental design

Laboratory infections

*E. falciformis:*  
22

*E. ferrisi:*  
47

Uninfected:  
47



N = 116

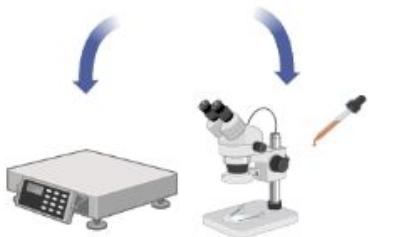
## 2. Methods: Experimental design

Day 0



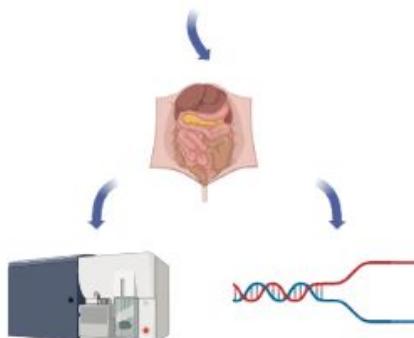
Infection with 150 sporulated oocysts

Day 1 - 8



1. Weight changes
2. Flotation (OPG)

Day 8

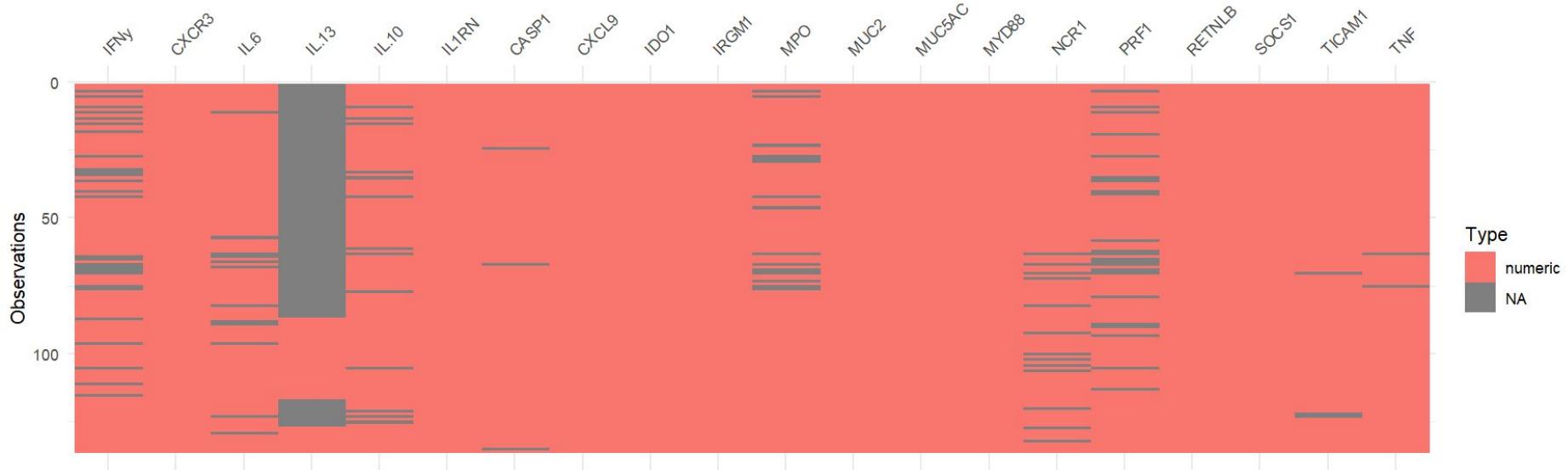


1. Cell sorting (FACS)
2. Immune gene expression Rt qPCR

Methods: Immune gene expression

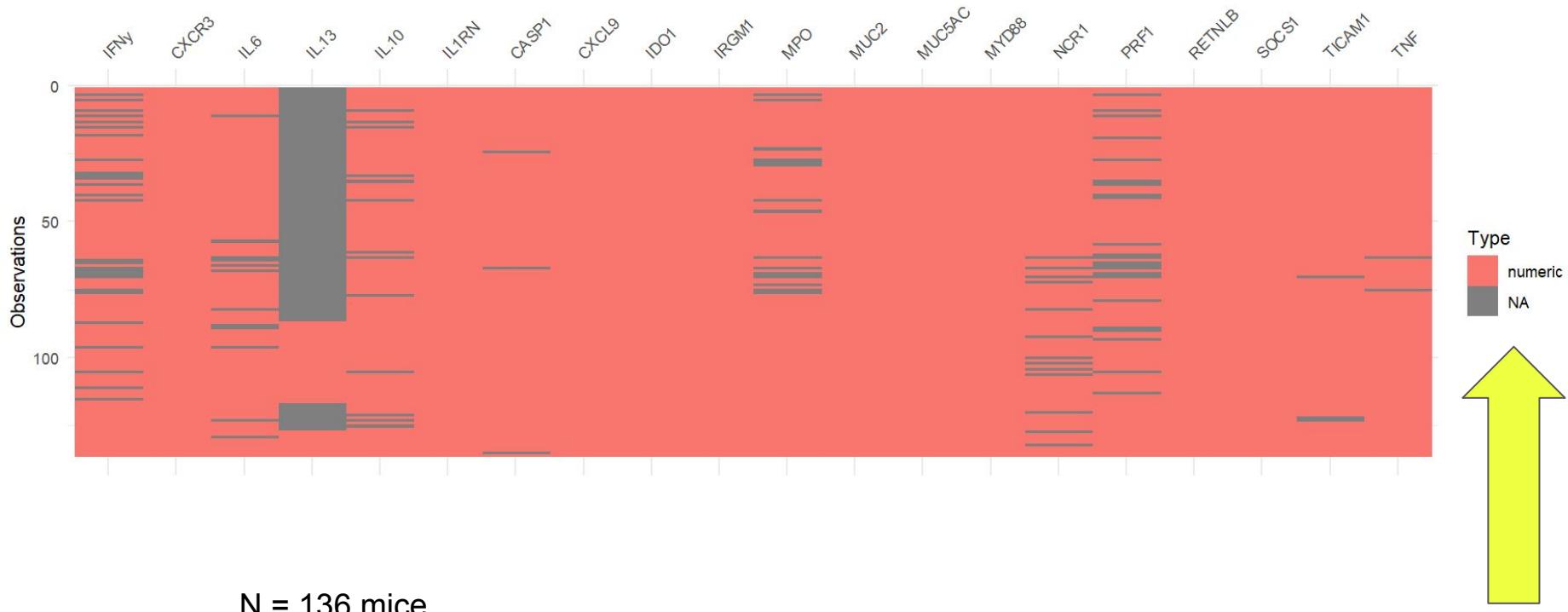
Predicting health impact of infections

# Laboratory infections: Gene expression data

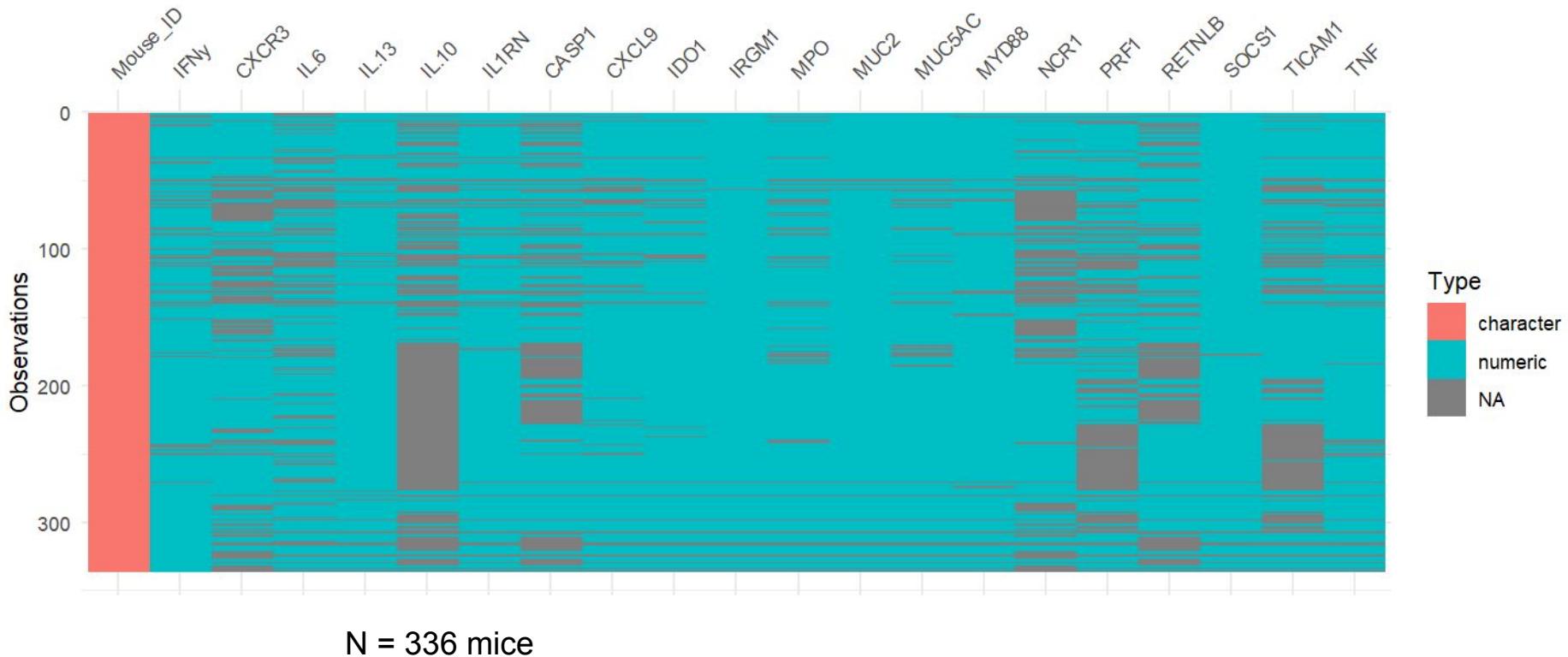


N = 136 mice

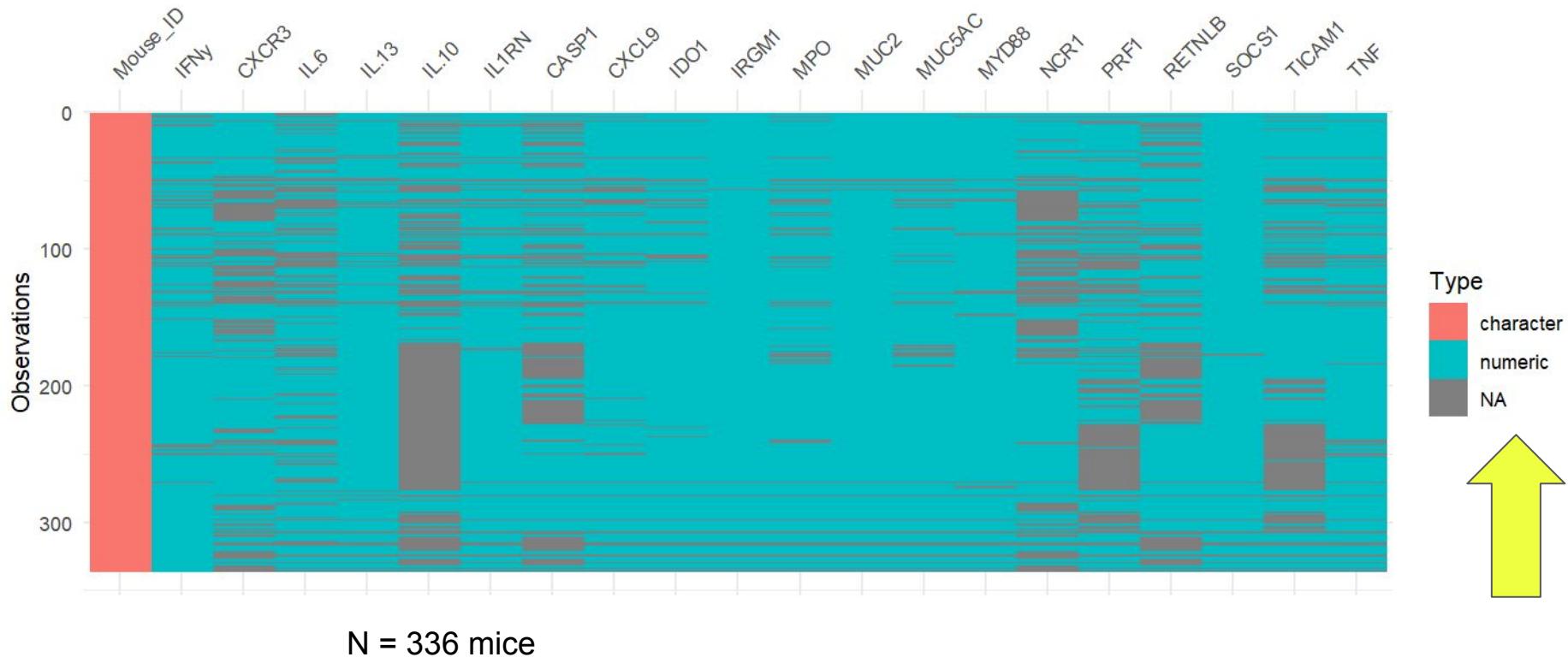
# Laboratory infections: Gene expression data



# Field data: Gene expression data



# Field data: Gene expression data



# **Methods:**

## Imputation of missing data

Why even **impute** missing data?



# Changing perspective on missing data

We should be suspicious of any dataset (large or small) which appears perfect.

— David J. Hand

Missing data are there whether we like it or not!

— Stef Van Buuren

The mean of the numbers 1, 2 and 4 can be calculated in R as

```
y <- c(1, 2, 4)  
mean(y)
```

```
[1] 2.33
```

where y is a vector containing three numbers, and where mean(y) is the R expression that returns their mean. Now suppose that the last number is missing. R indicates this by the symbol NA , which stands for “not available”:

```
y <- c(1, 2, NA)  
mean(y)
```

```
[1] NA
```

The mean is now undefined, and R informs us about this outcome by setting the mean to NA . It is possible to add an extra argument na.rm = TRUE to the function call. This removes any missing data before calculating the mean:

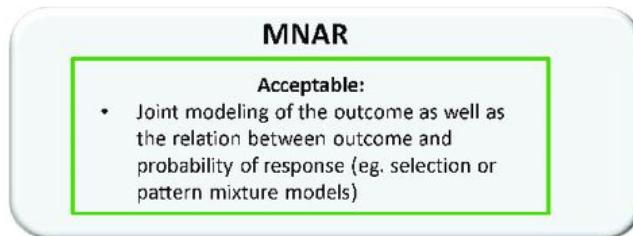
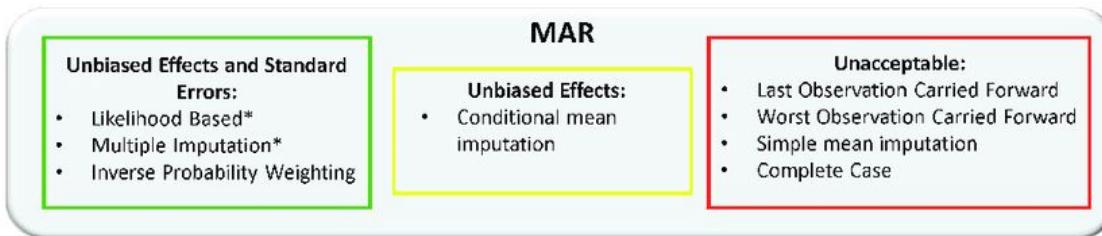
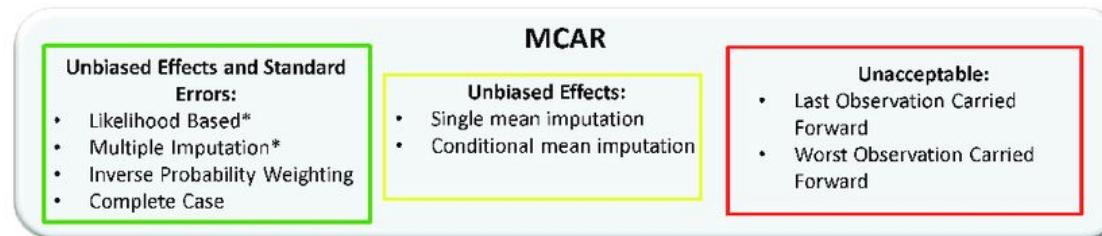
```
mean(y, na.rm = TRUE)
```

```
[1] 1.5
```

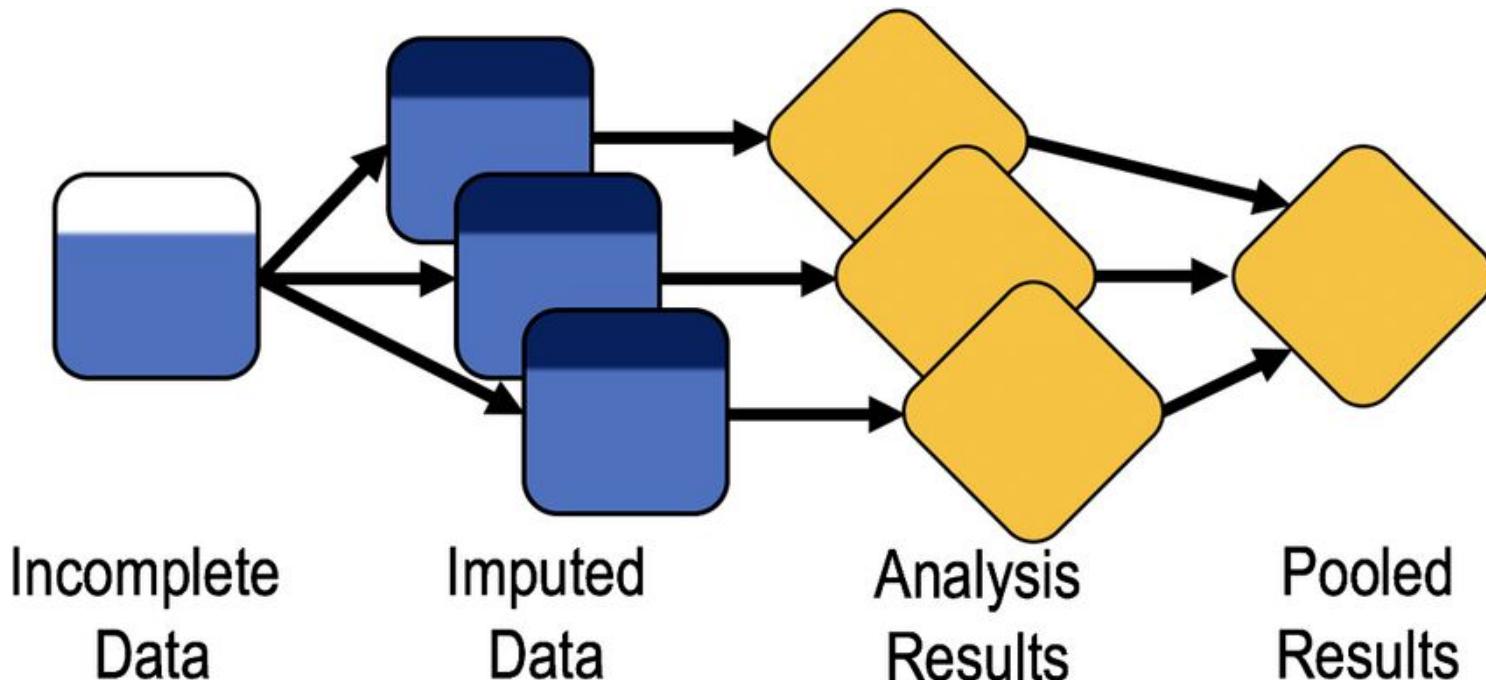
# Types of missing data

- Missing Completely at Random (MCAR)
- Missing at Random (MAR)
- Missing Not at Random (MNAR)

# Summary of acceptable and unacceptable analytic methods for types of missing data



# Multiple imputation in a nutshell



## 2.1.3 The expanding literature on multiple imputation

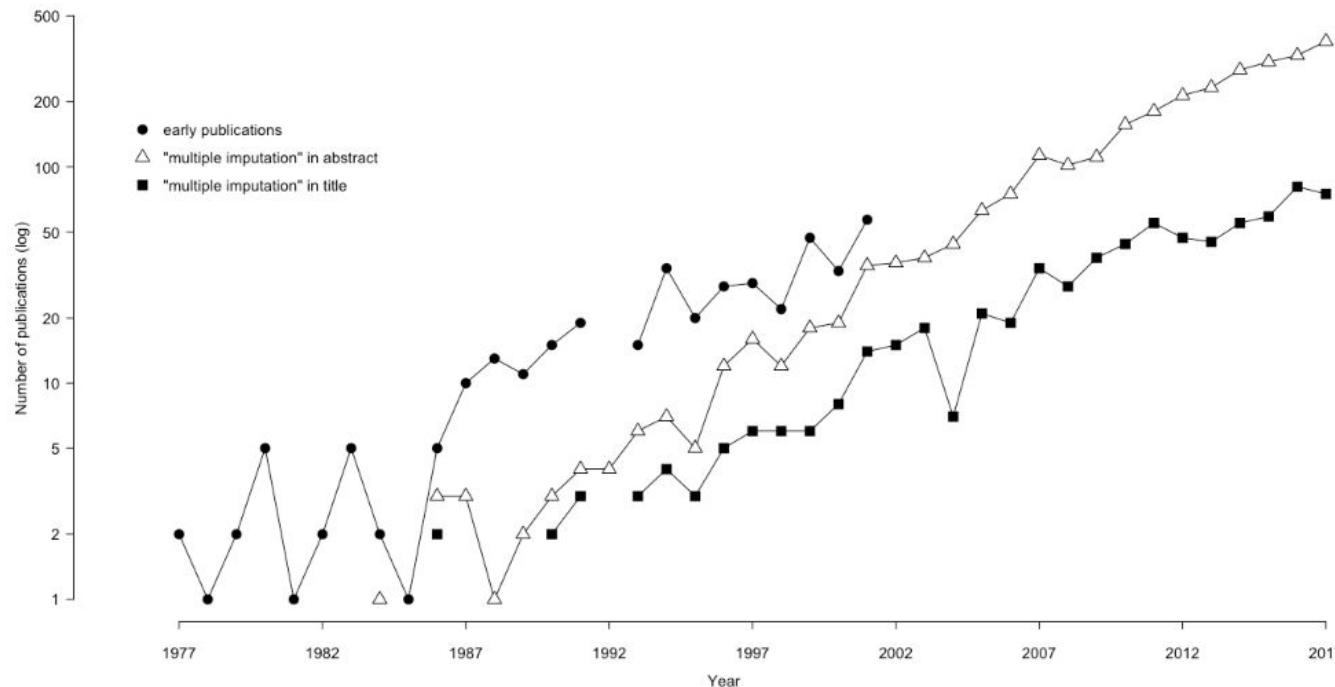
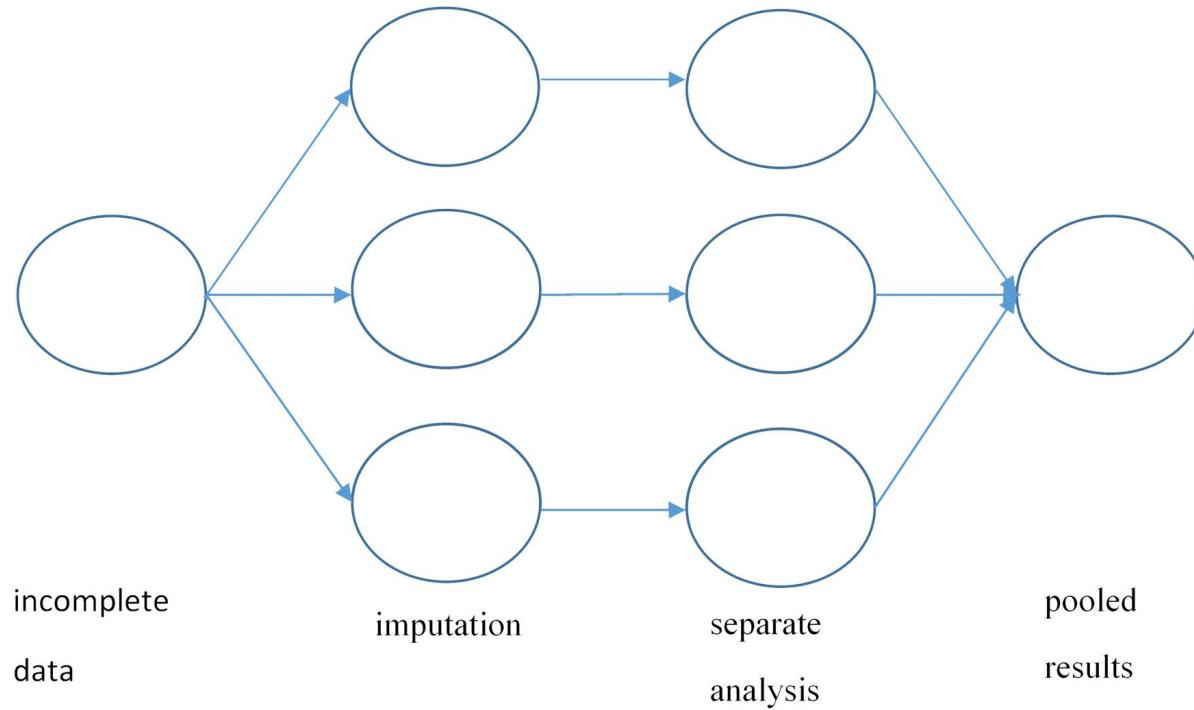
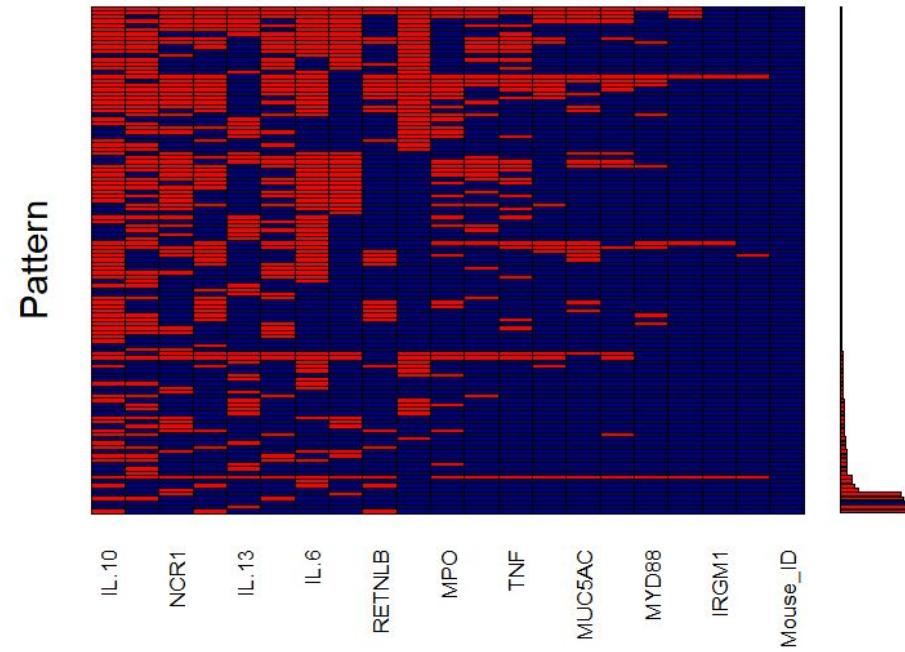
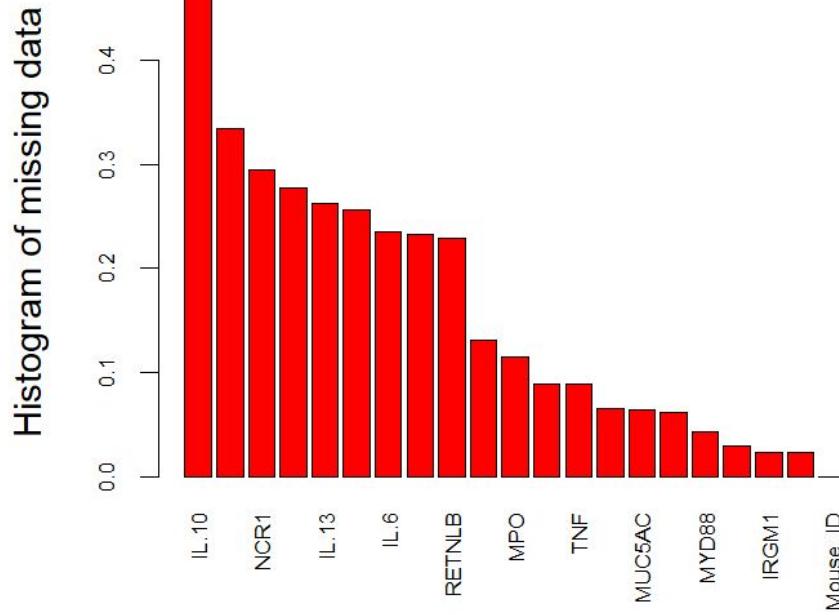


Figure 2.1: Multiple imputation at age 40. Number of publications (log) on multiple imputation during the period 1977–2017 according to three counting methods. Data source: <https://www.scopus.com> (accessed Jan 14, 2018).

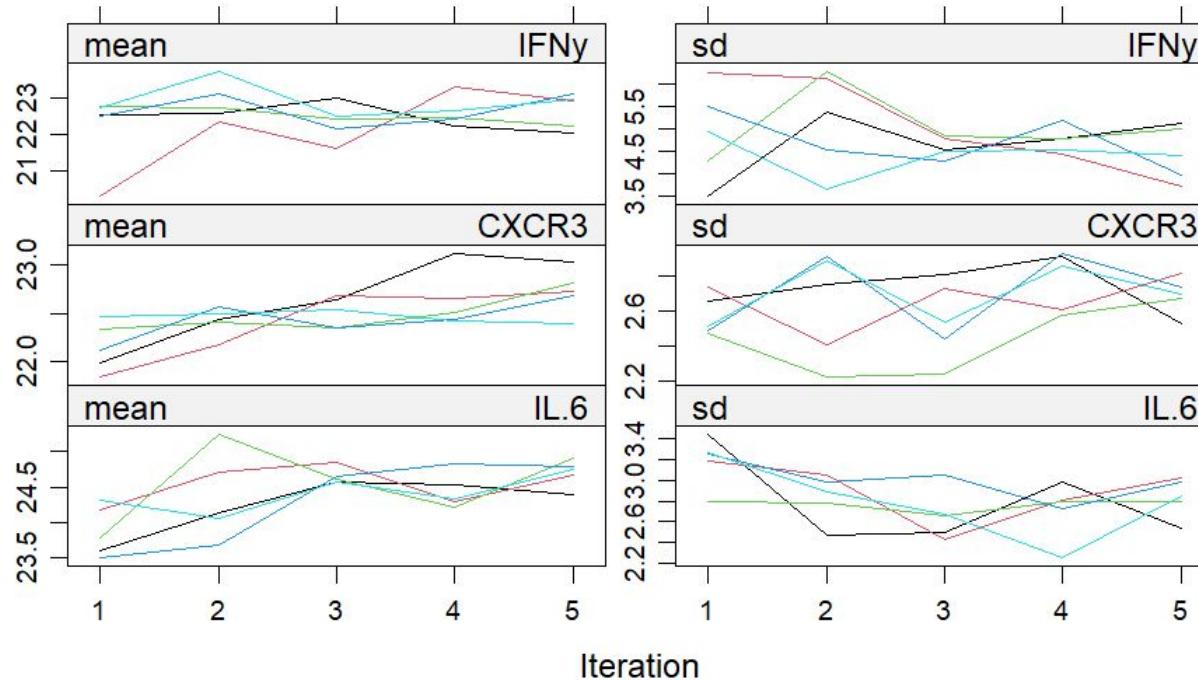
# MICE imputation



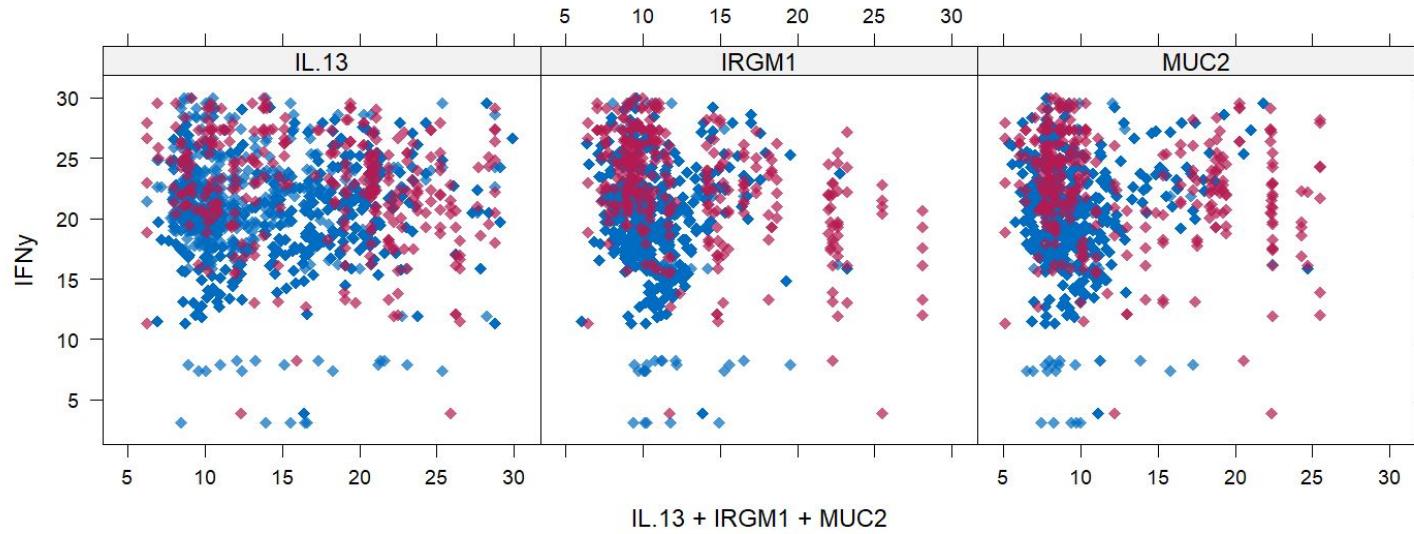
# MICE - Missingness in immune gene expression data



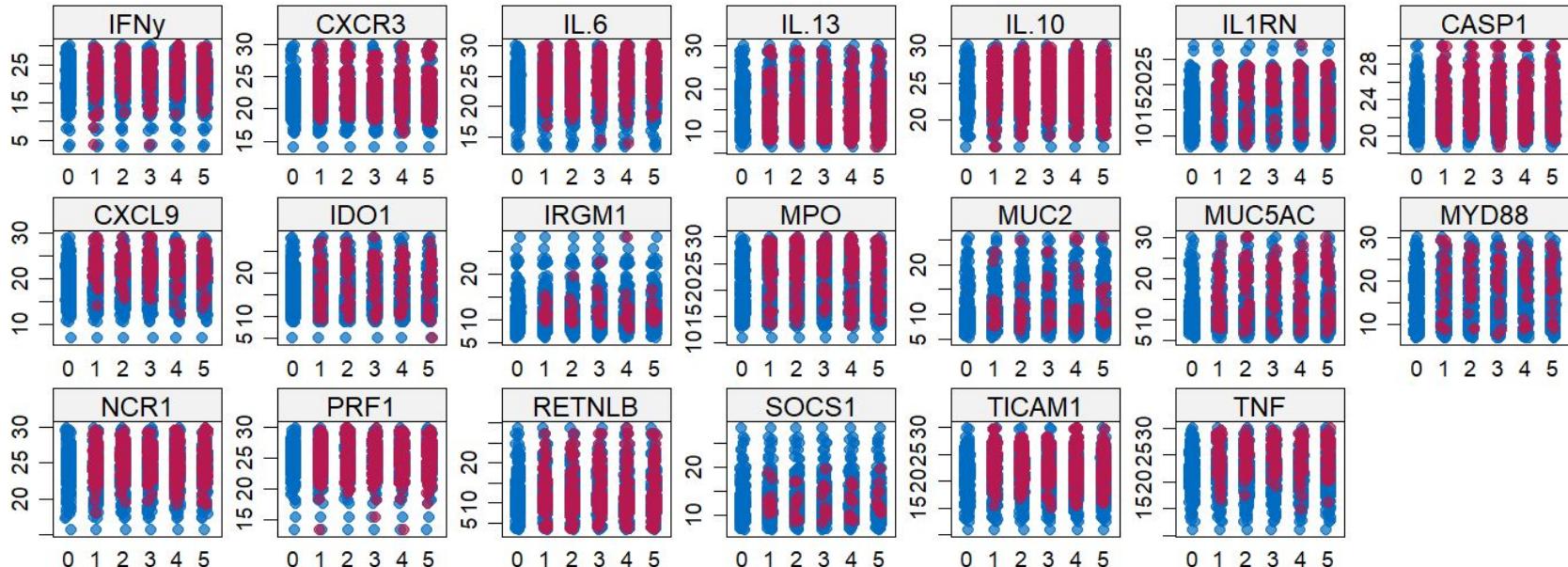
# MICE



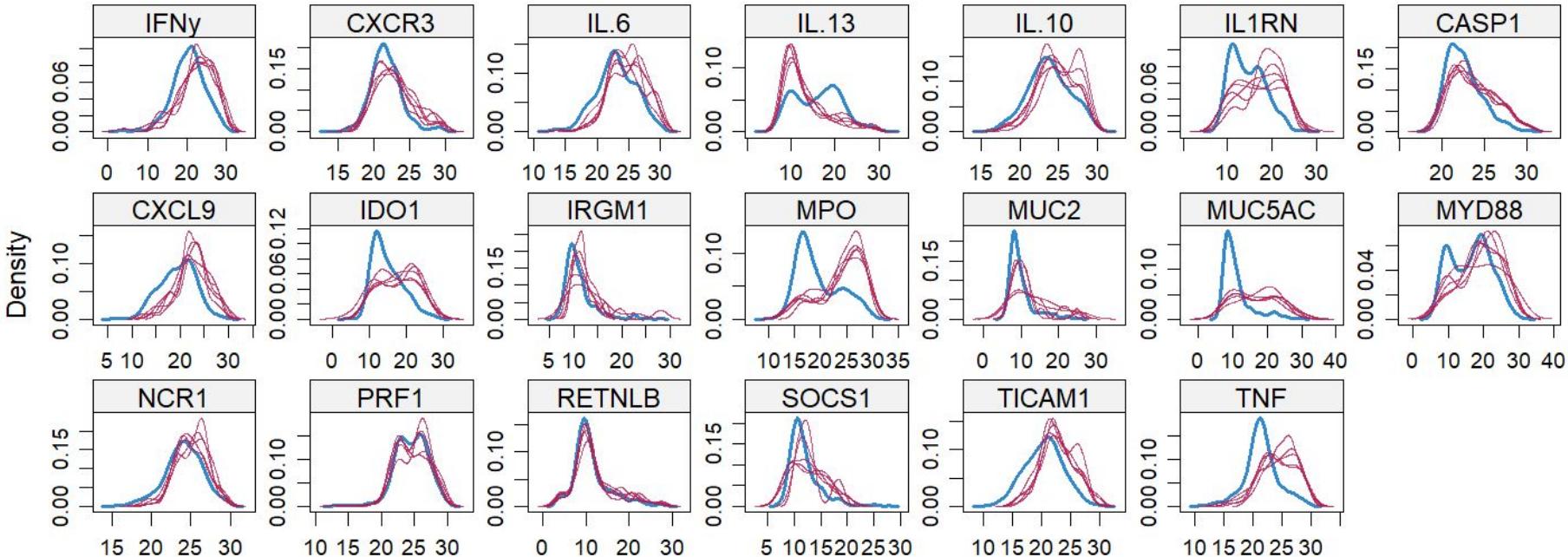
# MICE: *xyplot* - distribution comparison between imputed and observed data



# MICE: stripplot- distribution comparison between imputed and observed data



# MICE: stripplot- distribution comparison between imputed and observed data



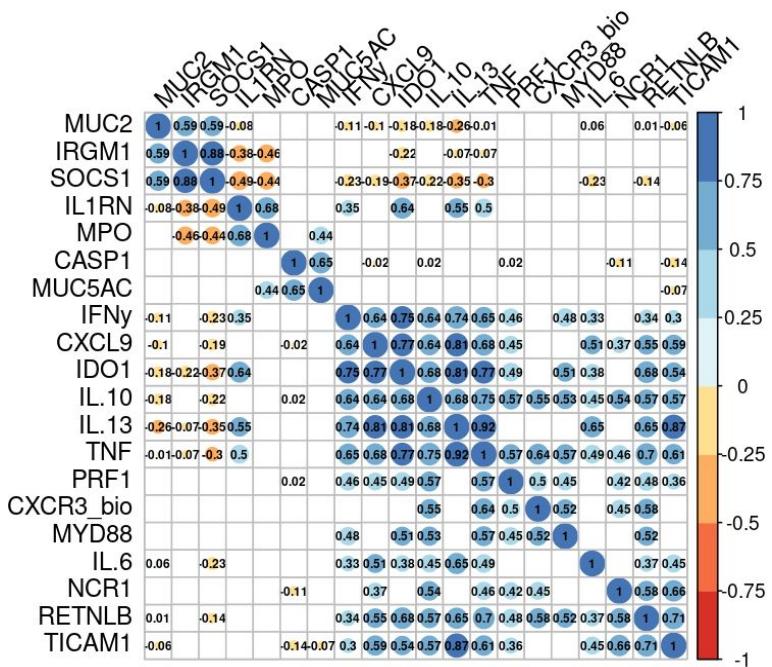
# MICE: Result



## **Methods:**

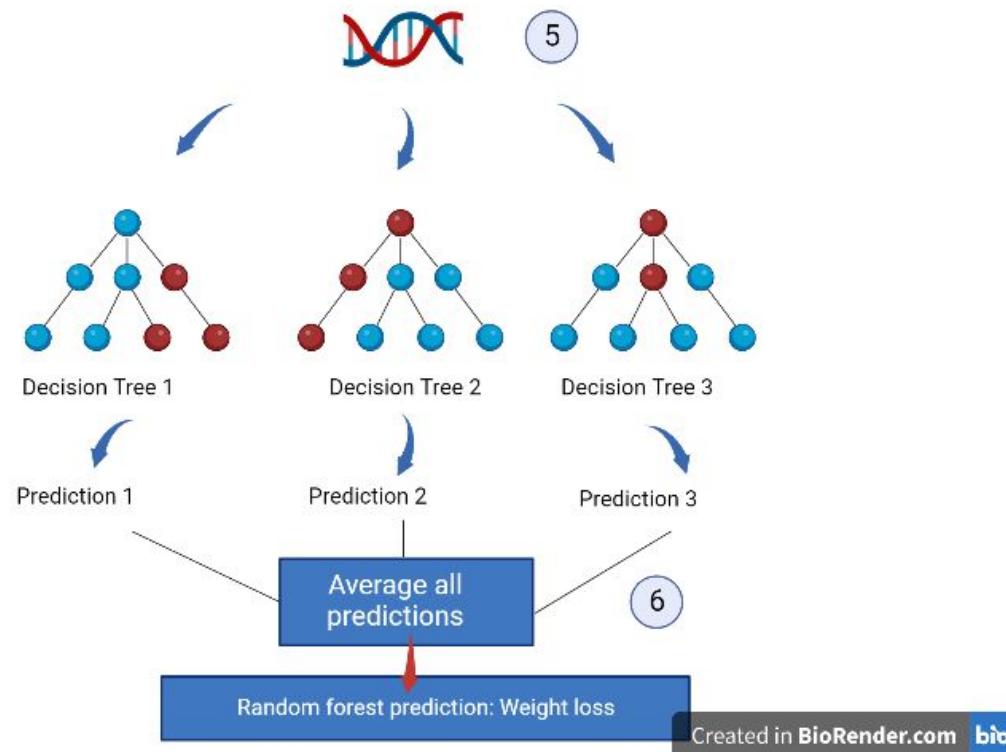
Random forest training on experimental  
laboratory infections

# Methods: Immune gene expression



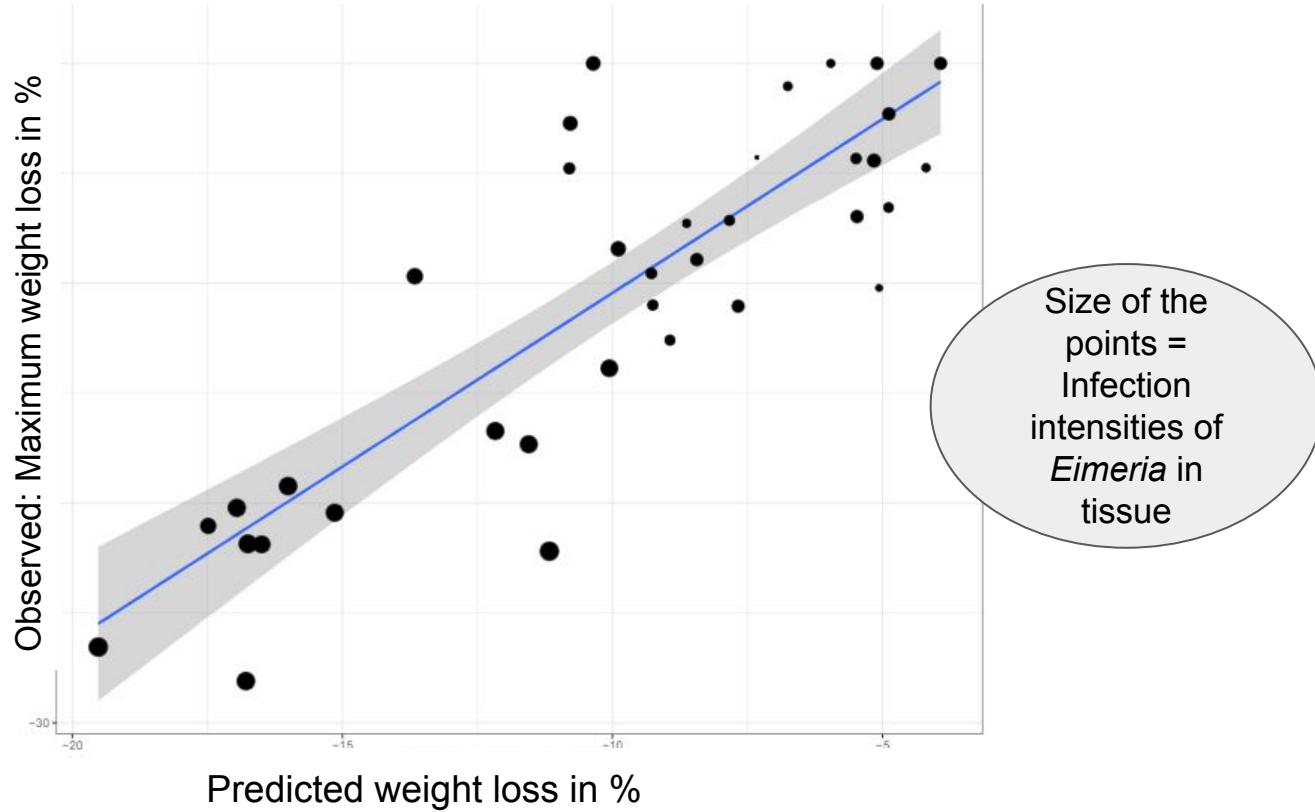
Can the immune measures predict the impact of the infection on host health?

# Random forest - Background



# Predicted and observed maximum weight loss

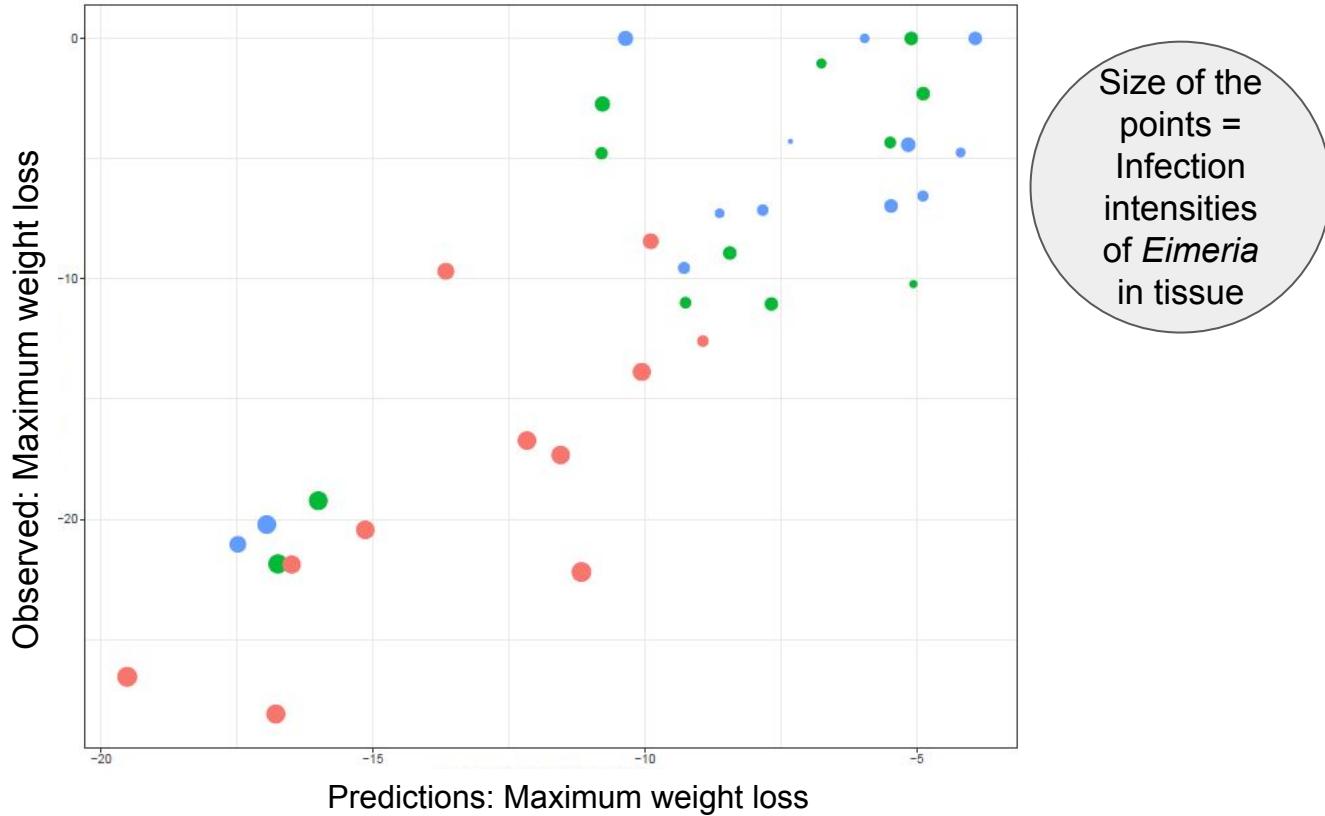
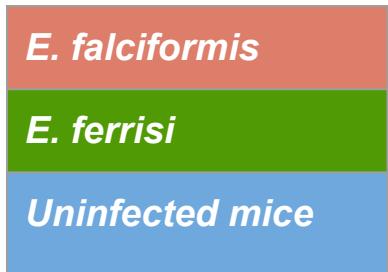
```
cor(result$WL_max,  
result$predictions,  
method = c("pearson",  
"kendall", "spearman"))  
0.8448673
```



# Predicted and observed maximum weight loss

Spearman's rank correlation

$S = 2661.2$ , p-value =  
 $2.48e-08$   
Rho: 0.750352

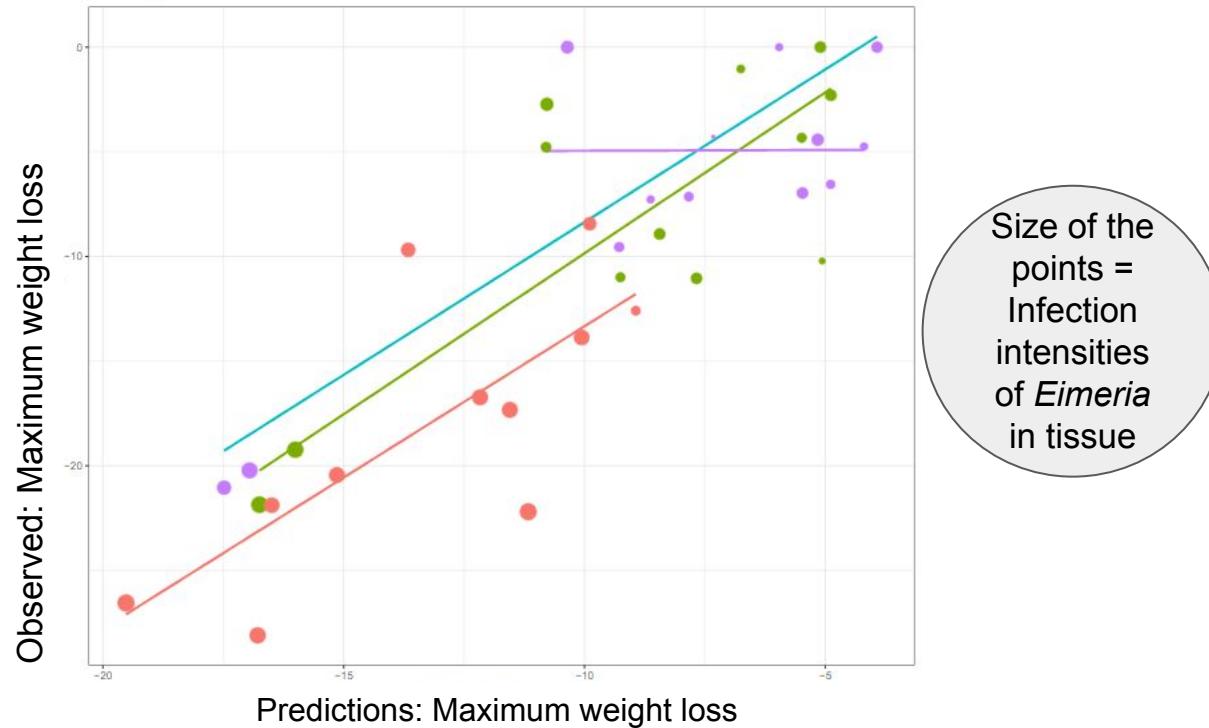


# Is the infection status of our mice reflecting the experiment planning?

```
# Here I create a new column, where we get the actual infection status
# According to the melting curve for eimeria
hm <- hm %>%
  dplyr::mutate(current_infection = case_when(
    Parasite_challenge == "E_ferrisi" & MC.Eimeria == "TRUE" ~ "E_ferrisi",
    Parasite_challenge == "E_ferrisi" & MC.Eimeria == "FALSE" ~ "uninfected",
    Parasite_challenge == "E_falciformis" & MC.Eimeria == "TRUE" ~ "E_falciformis",
    Parasite_challenge == "E_falciformis" & MC.Eimeria == "FALSE" ~ "uninfected",
    Parasite_challenge == "uninfected" & MC.Eimeria == "TRUE" ~ "infected_eimeria",
    Parasite_challenge == "uninfected" & MC.Eimeria == "FALSE" ~ "uninfected",
    TRUE ~ ""))
})
```

# Model predicts **false** positives

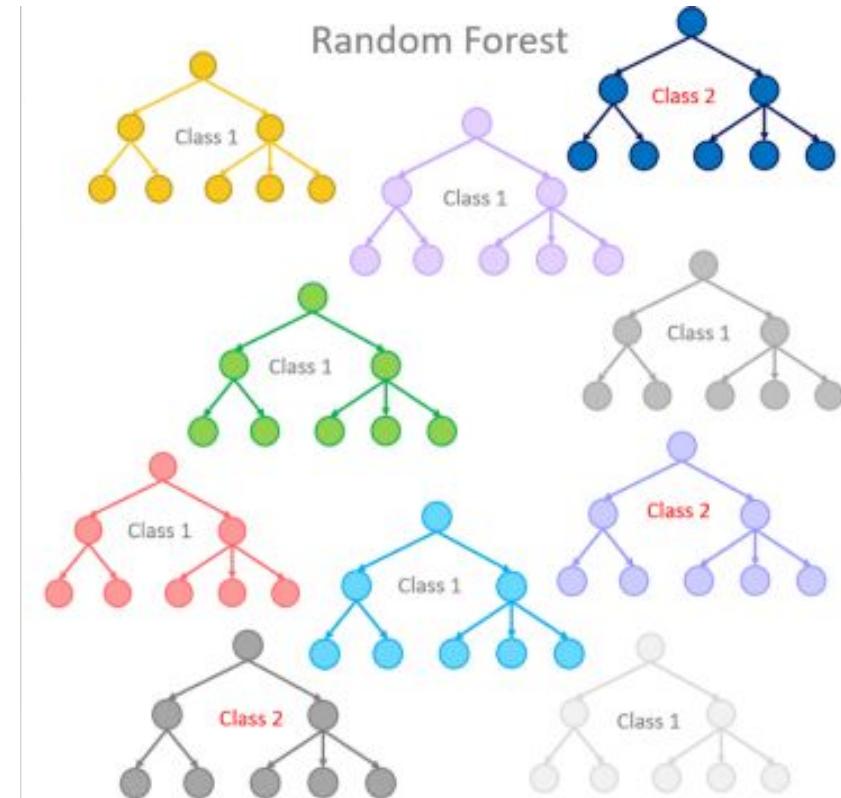
- E. falciformis*
- E. ferrisi*
- Infected with *eimeria*
- Actually uninfected



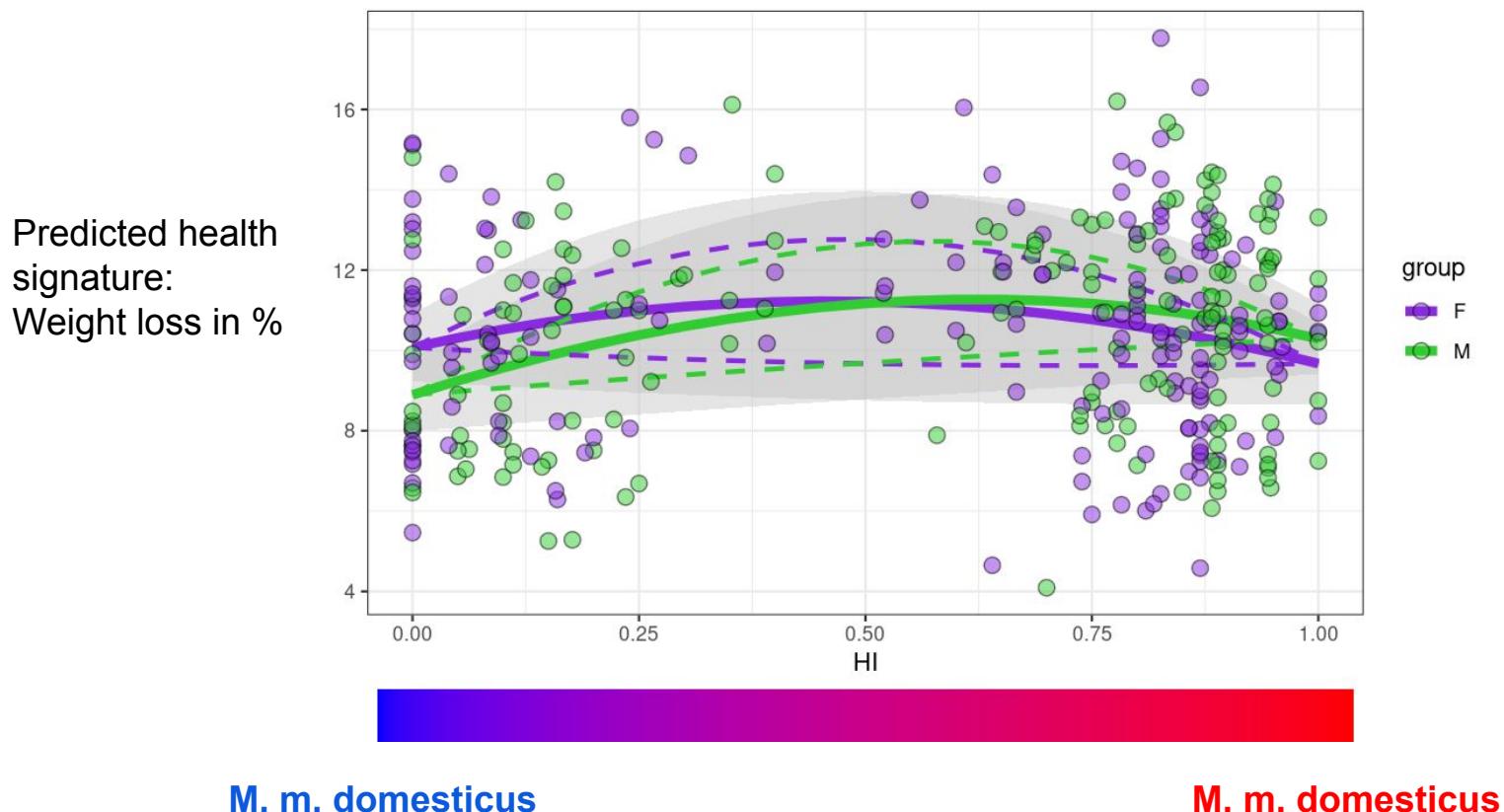
## Results

Application of random forest on wild data

# Random forest: Predicting health impact in the field

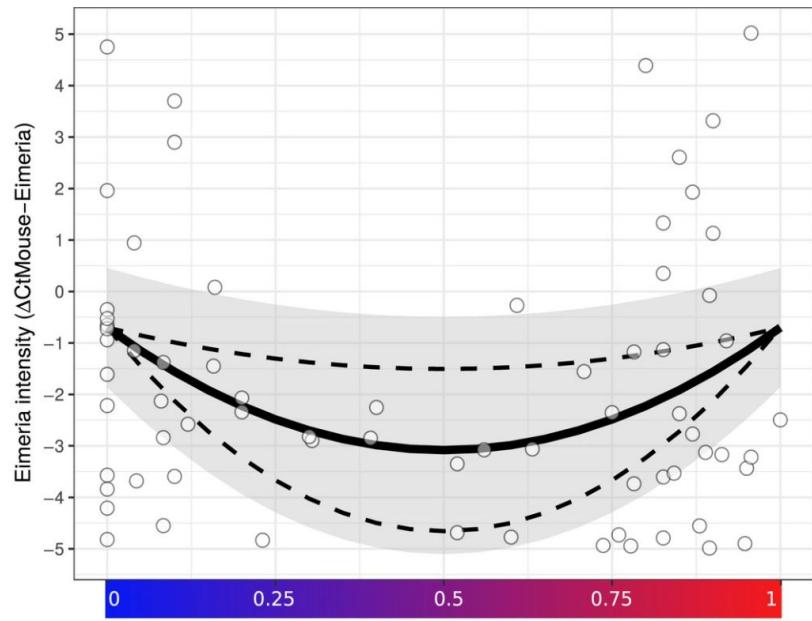


## Results: Effect of hybridization on predicted health impact

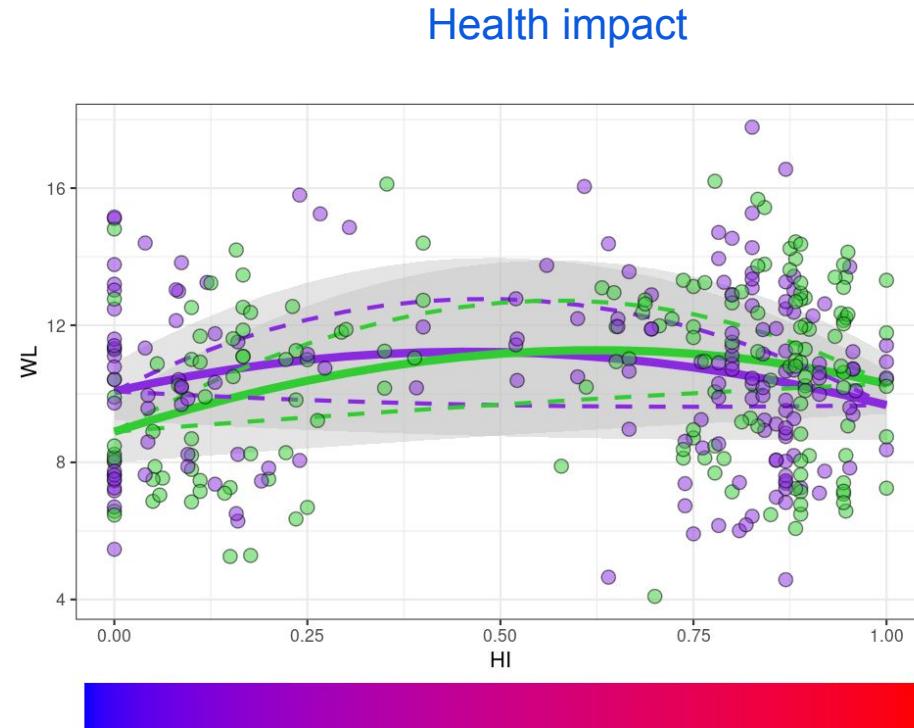


### 3. Results: Effect of hybridization on predicted health impact

Resistance



Health impact



group  
F  
M

# Acknowledgments

Berlin team:

- Emanuel Heitlinger
- Lubomir Bednar
- Finn Lobnow
- Victor Hugo Jarquín Díaz
- Alice Balard
- Sebastian Rausch
- Hongwei Zhang
- Laura Gramolini
- Tinghuan Song

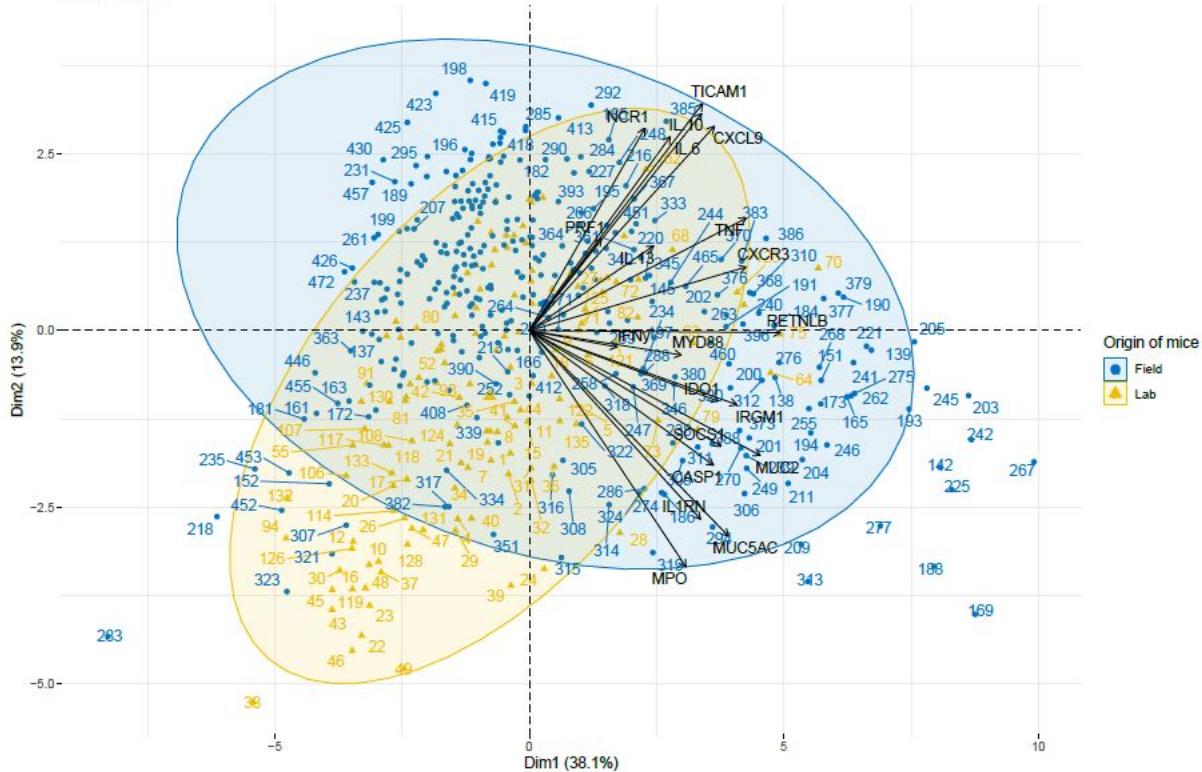


Czech team:

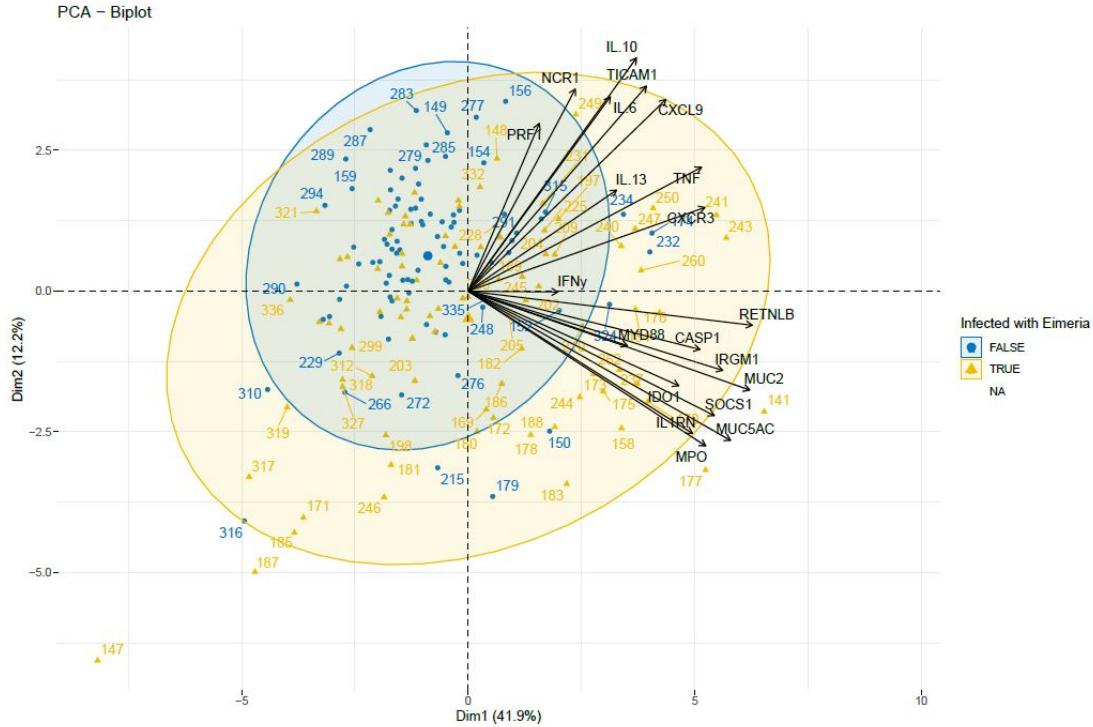
- Jaroslav Piálek
- Miloš Macholán
- Joëlle Goüy De Bellocq
- L'udovít Ďureje
- Iva Martincová
- Kristina Daniszová

genes

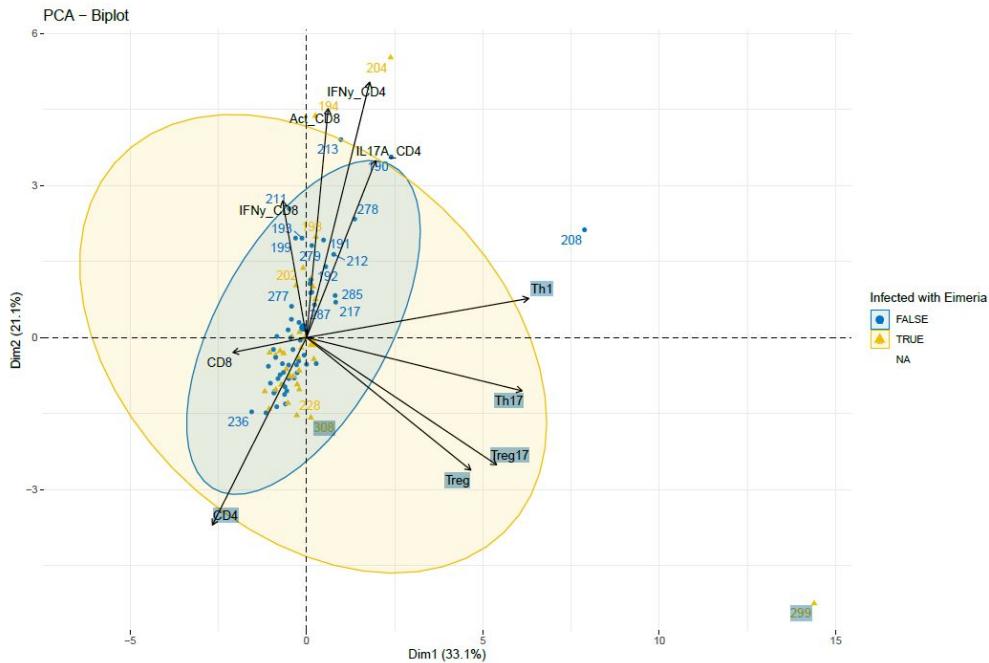
PCA - Biplot



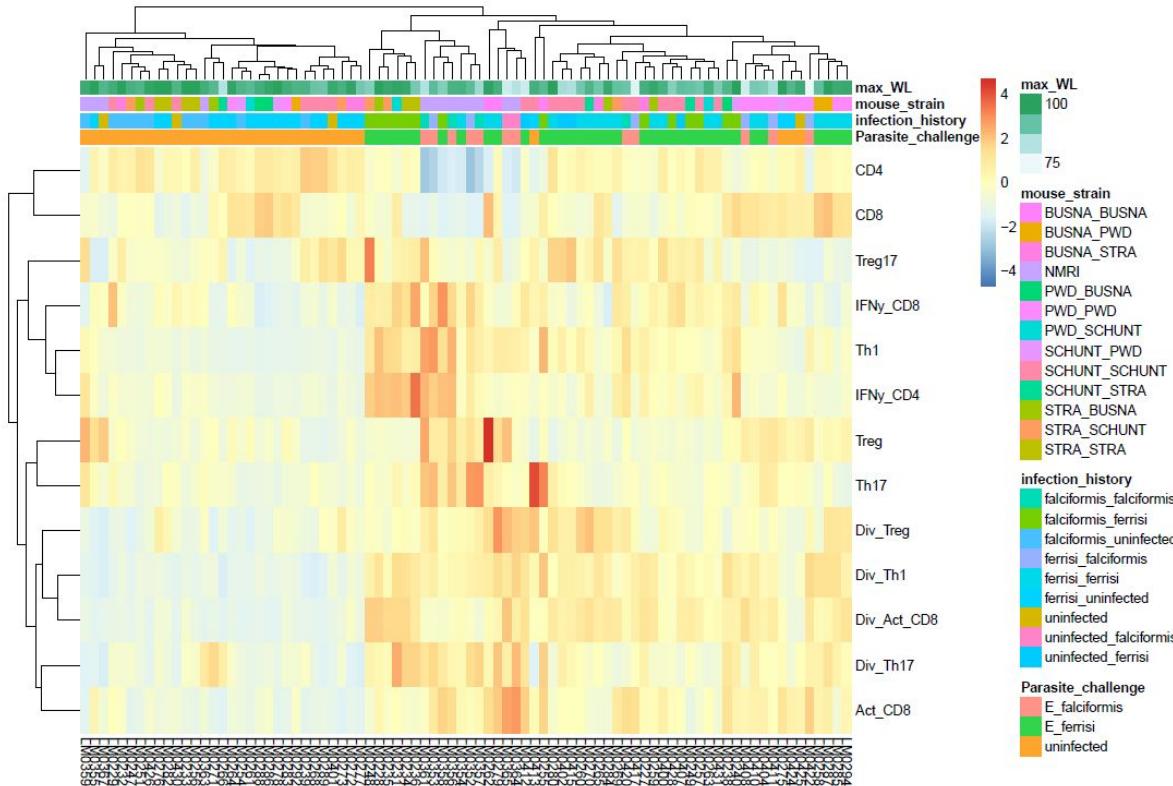
# Gene wild



# Facs wild

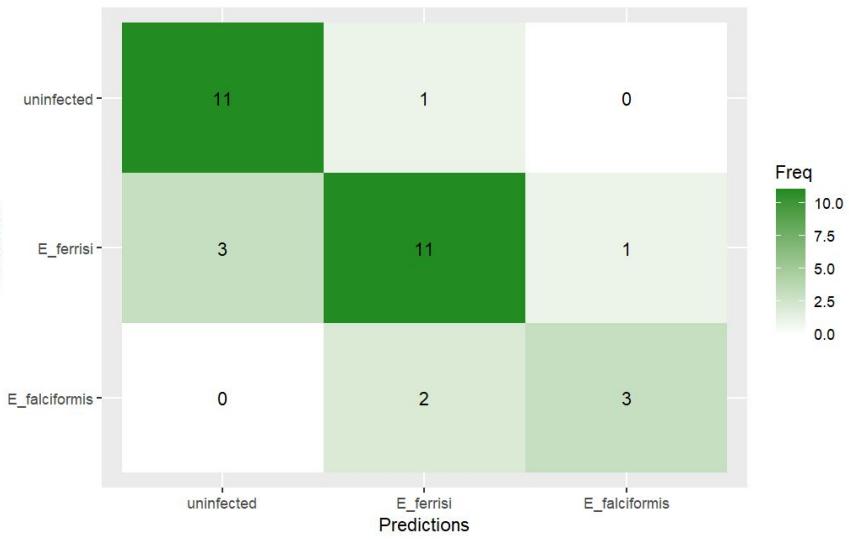


# Heatmap – Lab infections – facets data

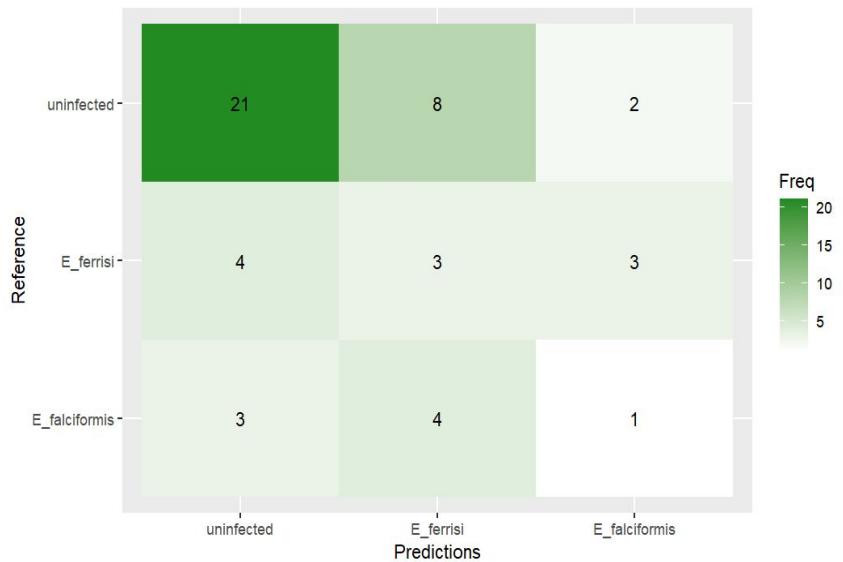


# Random forest: Using immune parameters to predict current infection

Lab data: predicting infective parasite

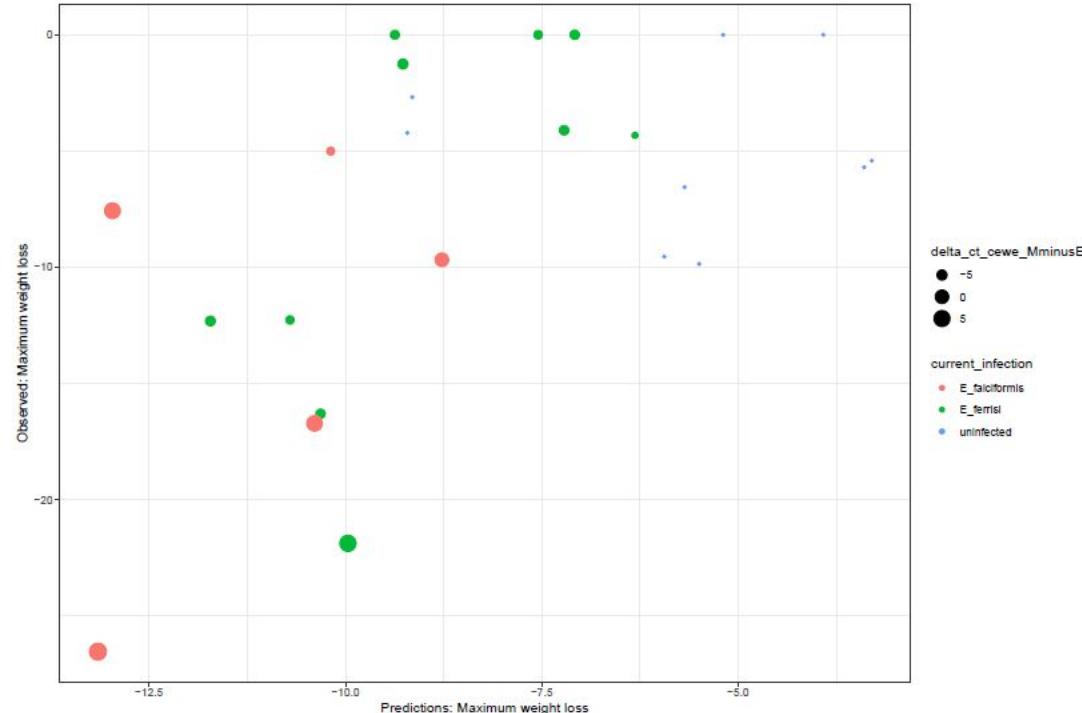


Field data: application of random forest

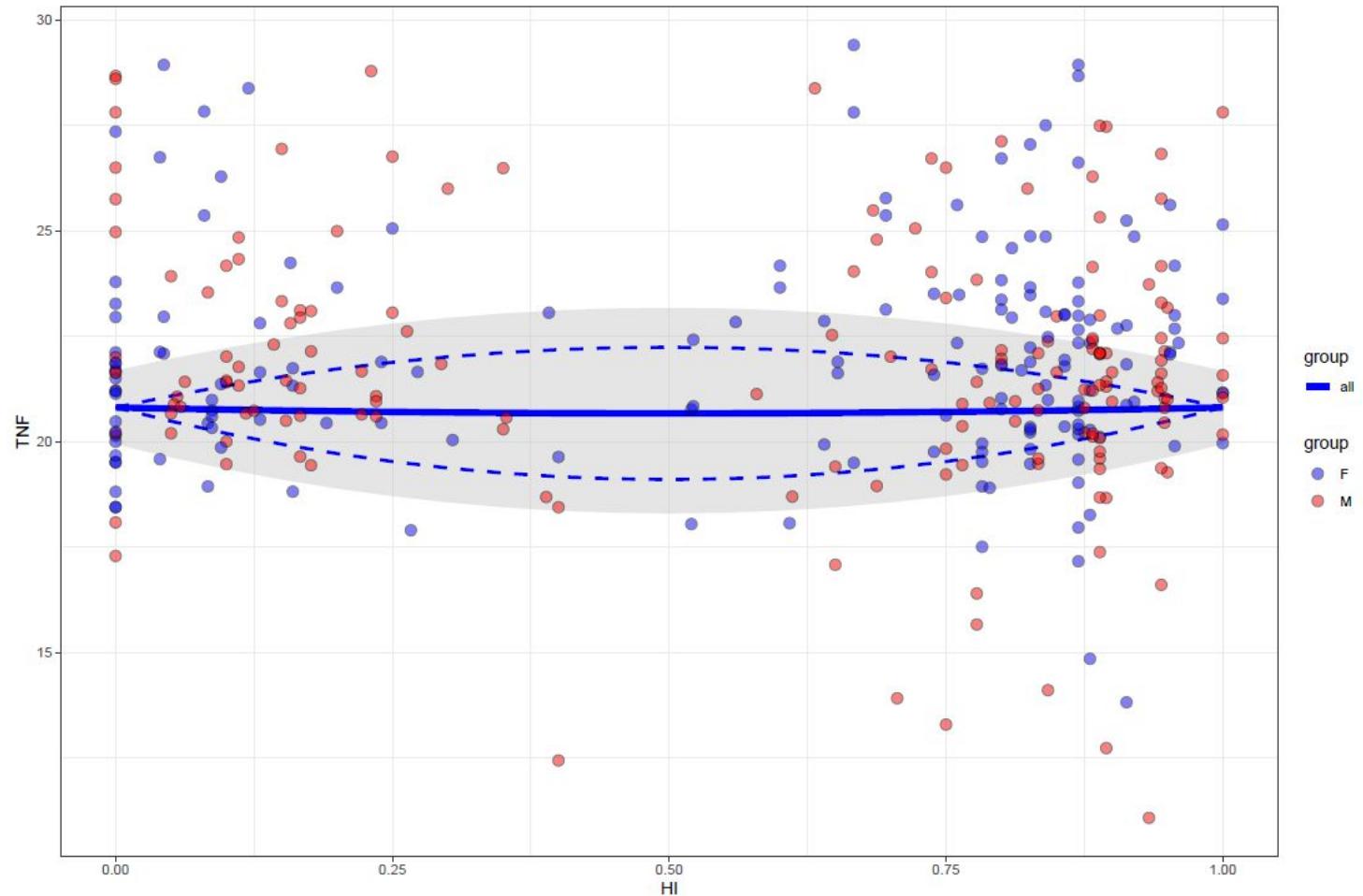


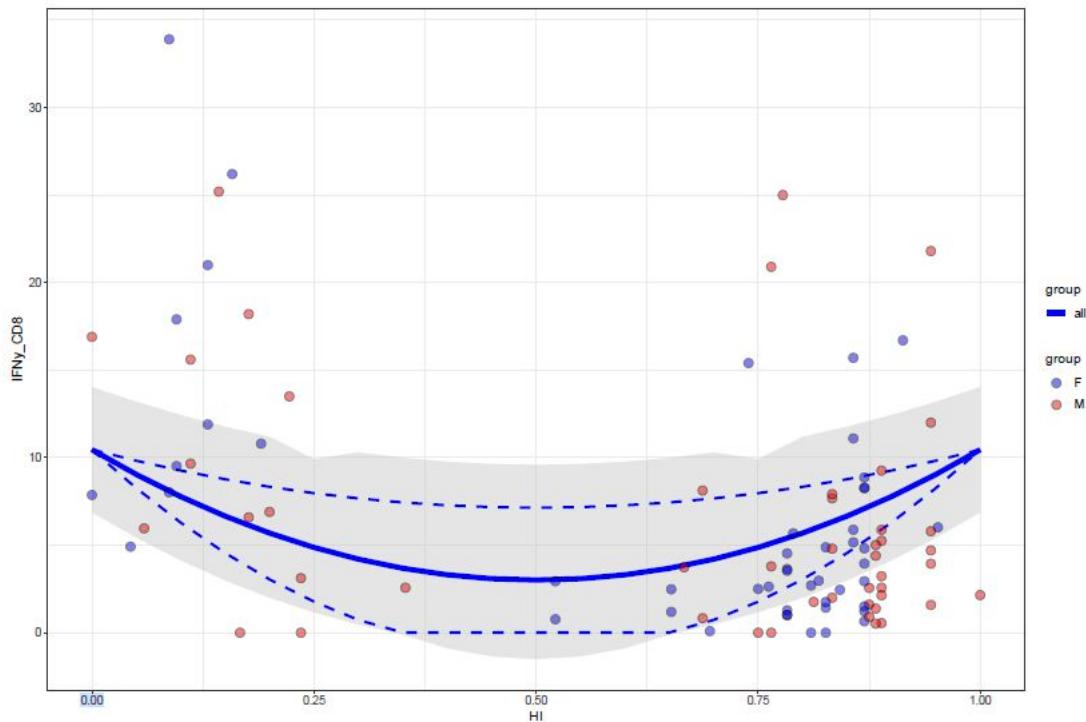
# Random forest facets

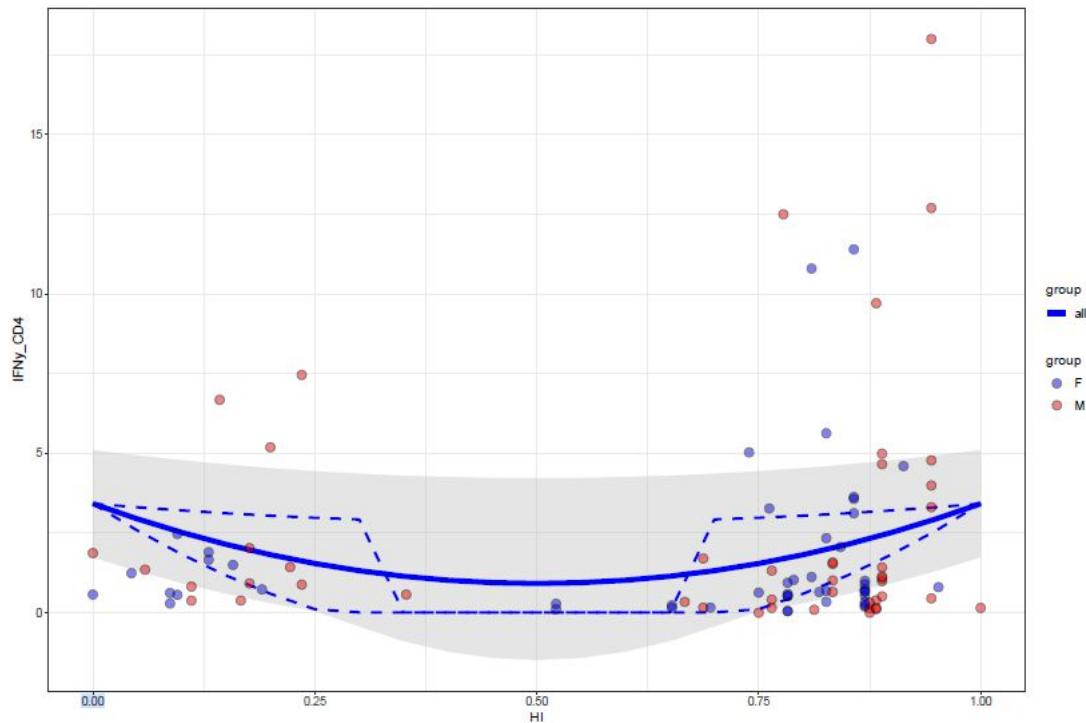
Rho =  
0.4908344



```
cor.test(result$predictions, result$WL_max, method = "spearman")
```

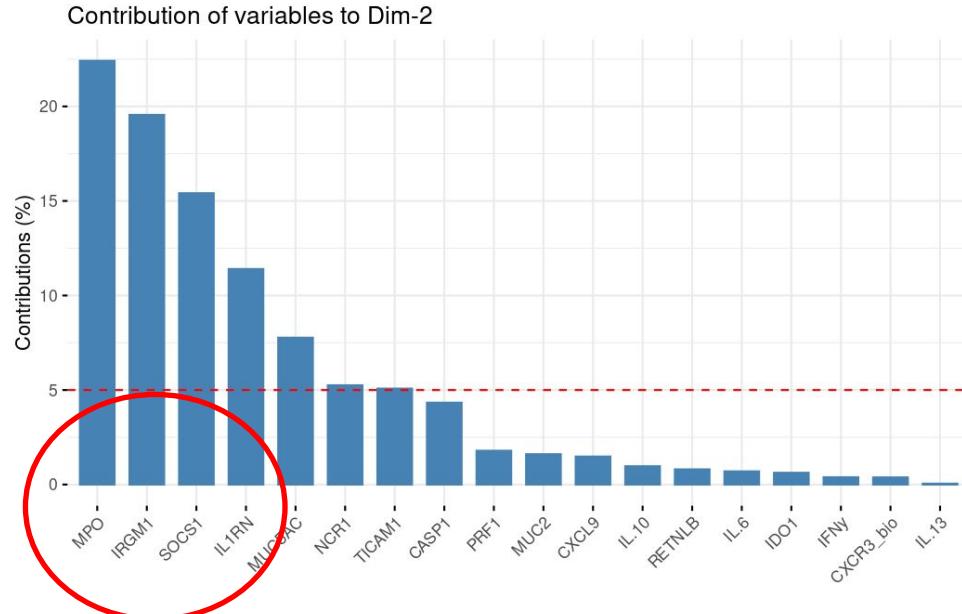




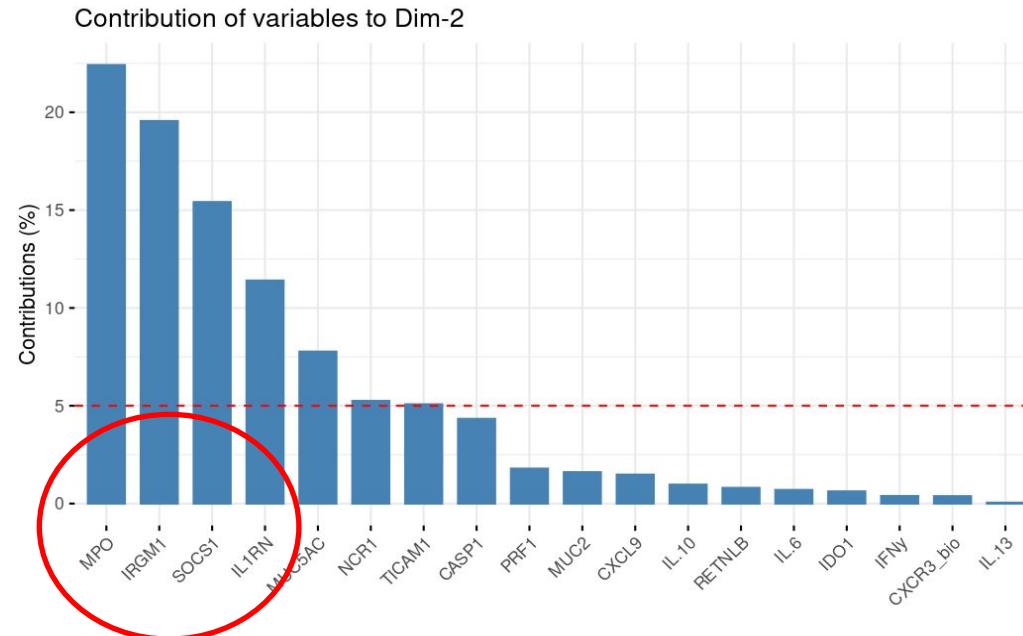
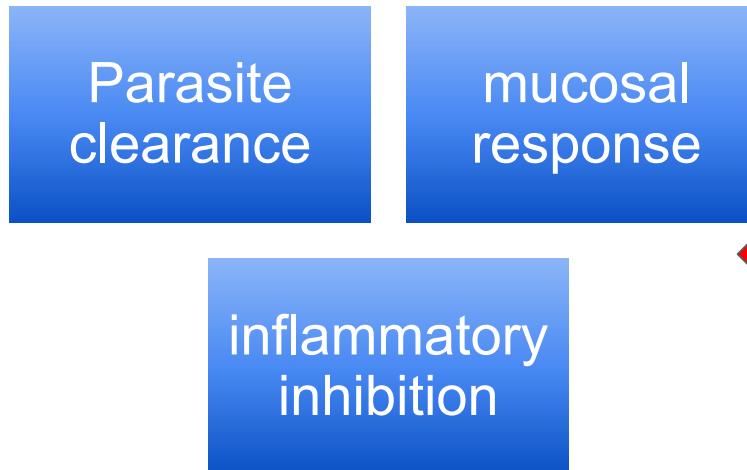


# Which principle component better predicts weight loss?

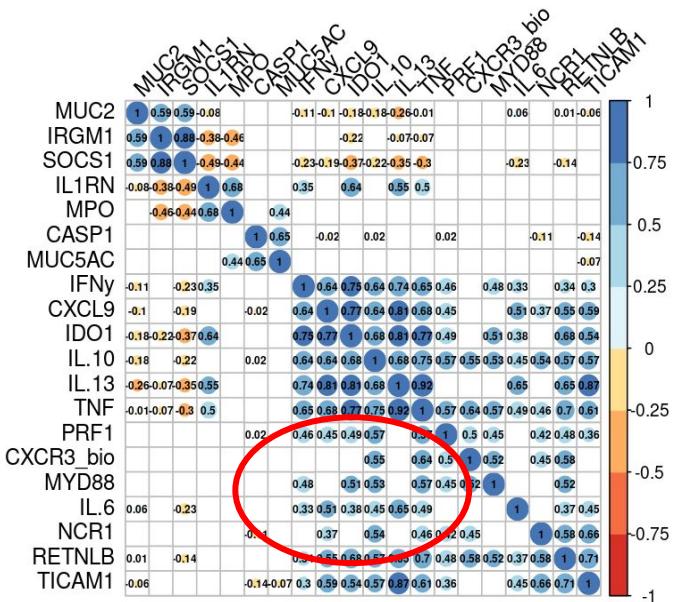
```
## Call:  
## lm(formula = max_WL ~ pc1 + pc2, data = imputed_expr)  
##  
## Residuals:  
##      Min        1Q    Median        3Q       Max  
## -16.913   -3.236    1.379    5.127   10.471  
##  
## Coefficients:  
##             Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 92.3746     0.6006 153.811 <2e-16 ***  
## pc1         0.1702     0.2061   0.826  0.4107  
## pc2        -0.7501     0.3448  -2.175  0.0317 *
```



# Which principle component better predicts weight loss?



# Inflammation, anti-fungal and microbial defense



PRF1

CXCR3

MYD88

IL.6

# Field data

Accuracy : 0.5102

95% CI : (0.3634, 0.6558)

No Information Rate : 0.6327

P-Value [Acc > NIR] : 0.9713

Kappa : 0.1191

McNemar's Test P-Value : 0.6422

Statistics by Class:

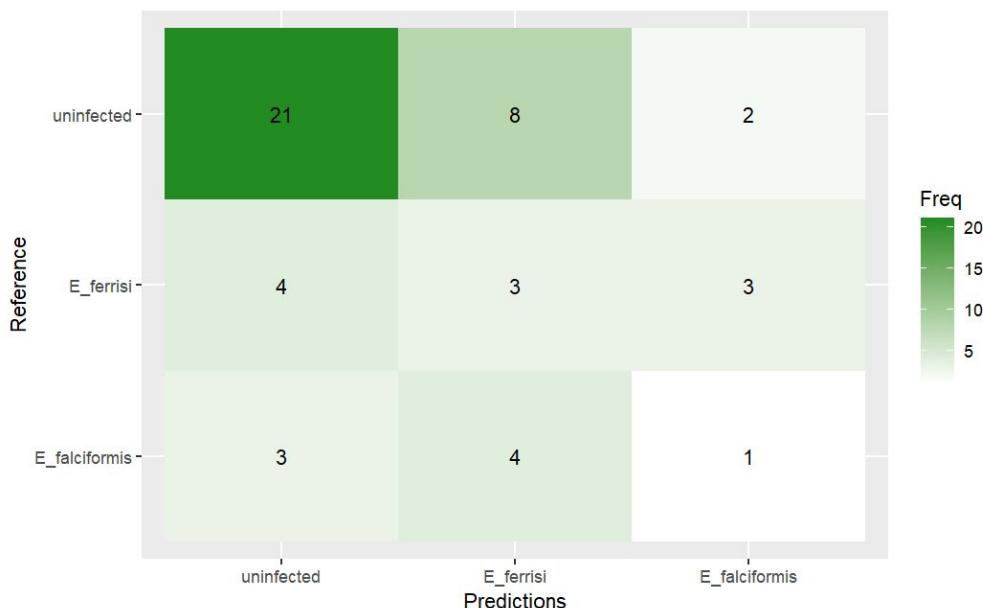
Class: E\_falciformis Class: E\_ferrisi Class: uninfected

Balanced Accuracy

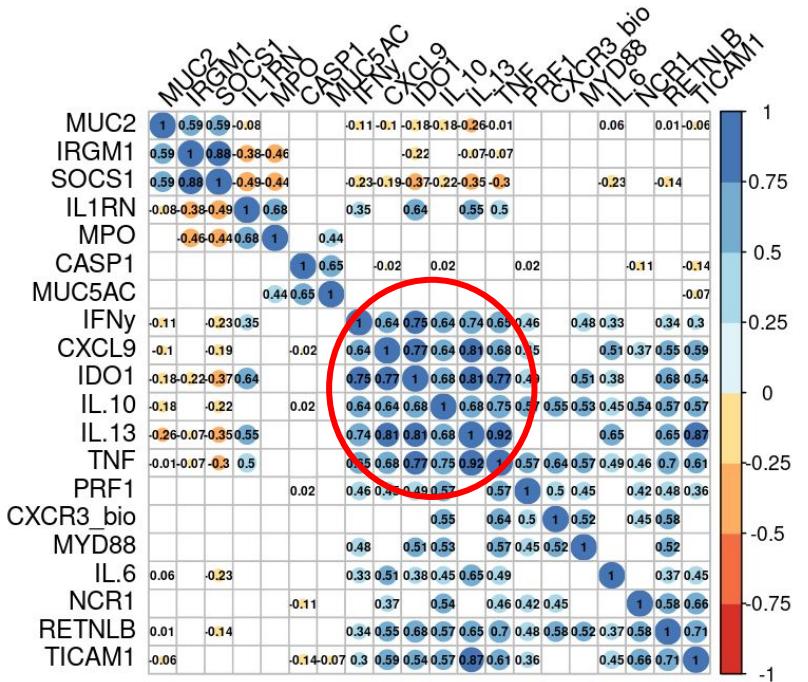
0.50152

0.49615

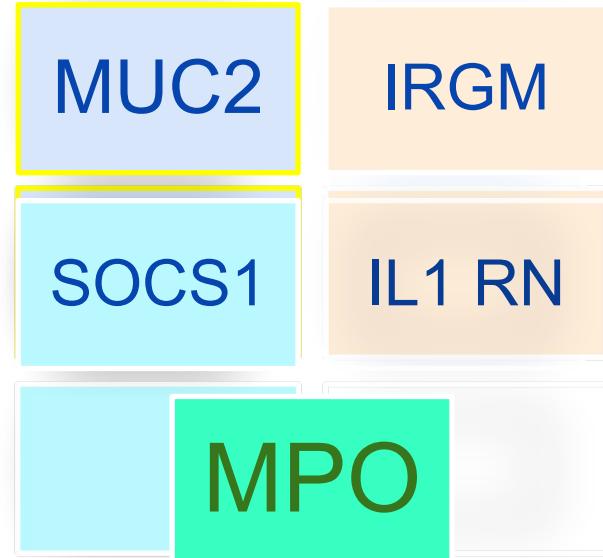
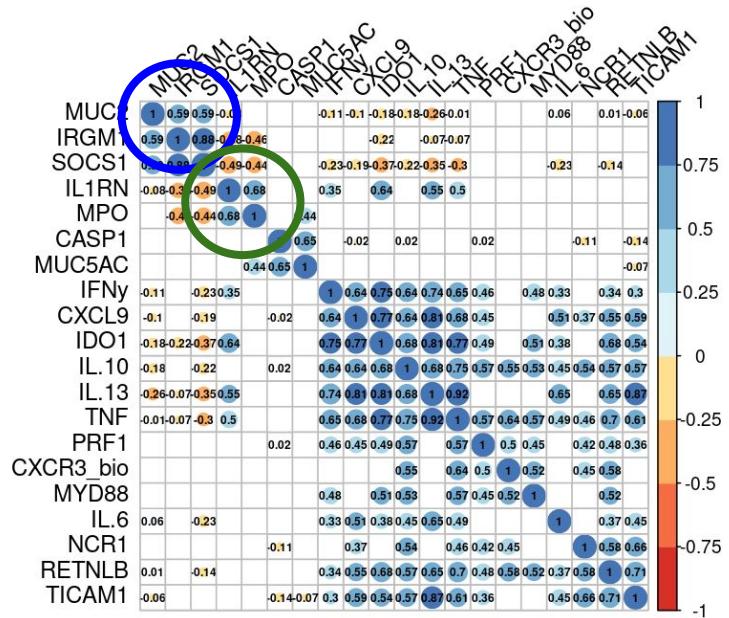
0.6443



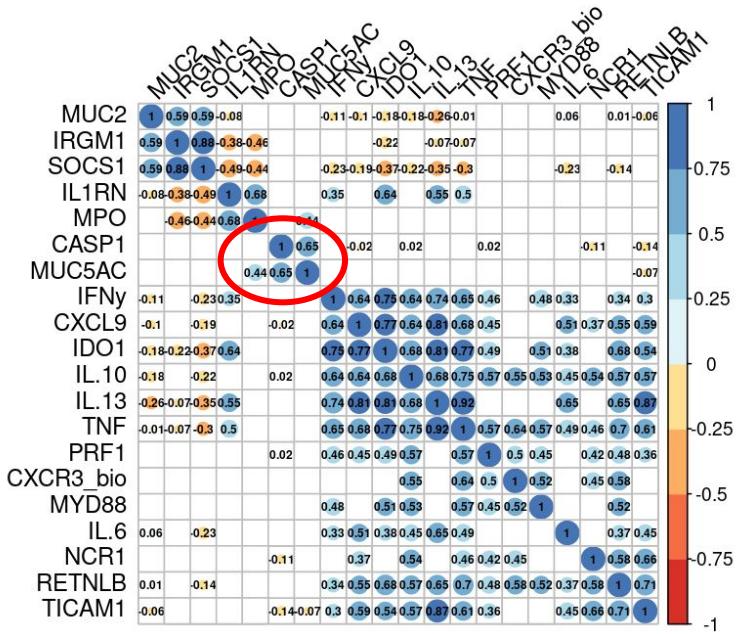
# Pro-inflammatory response → parasite clearance, mucus production



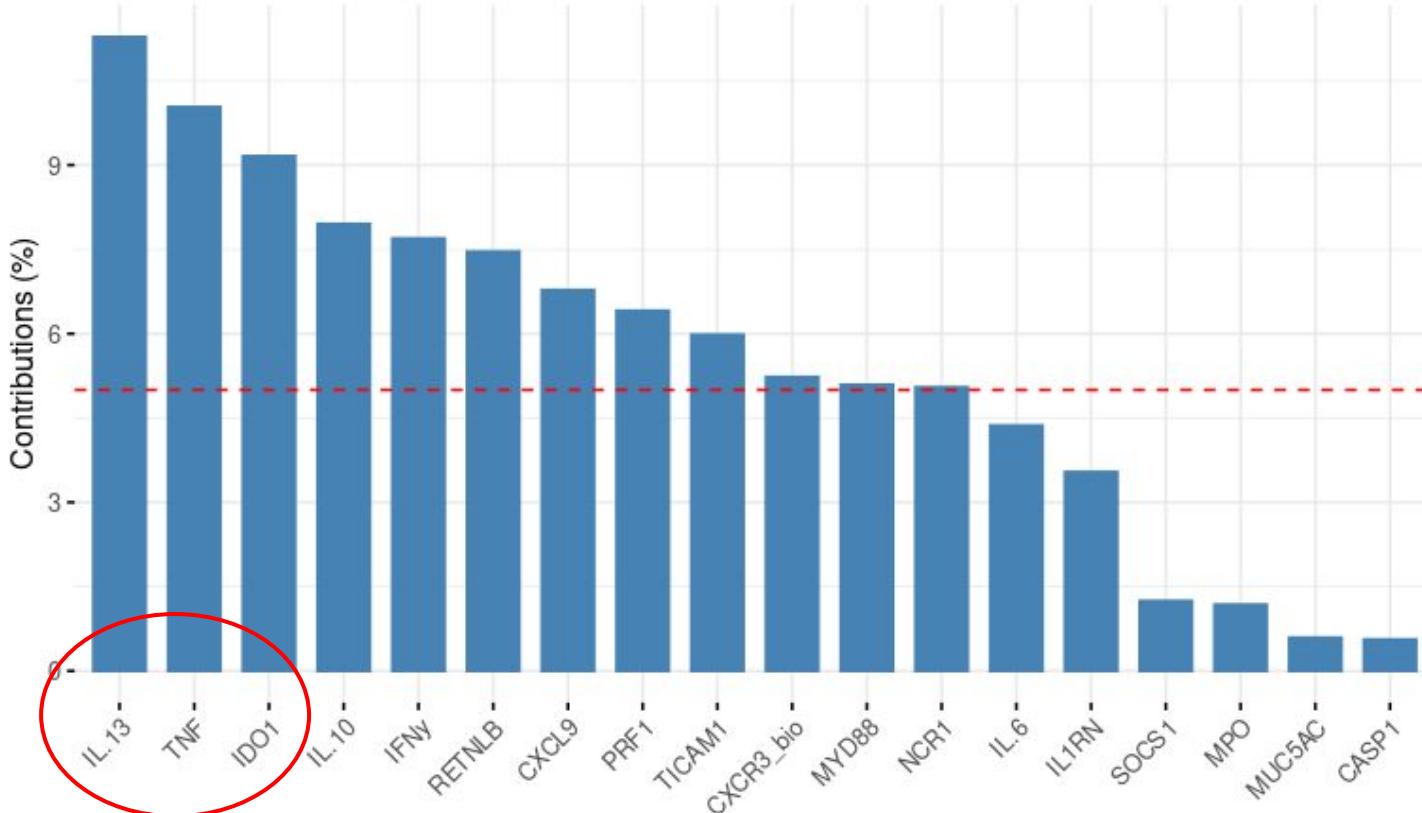
# Mucosal immune response + inflammatory inhibition:



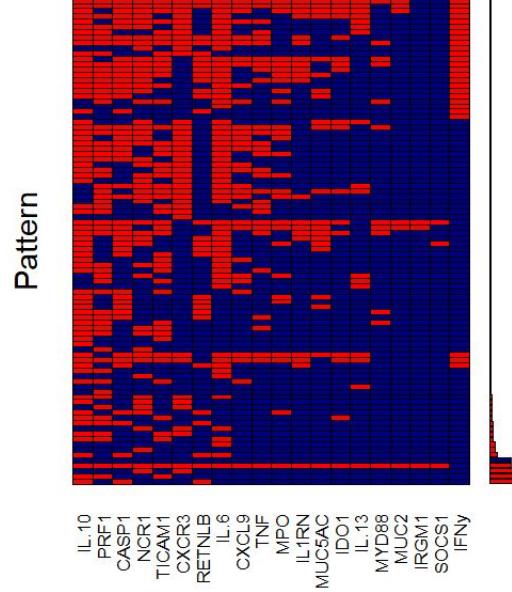
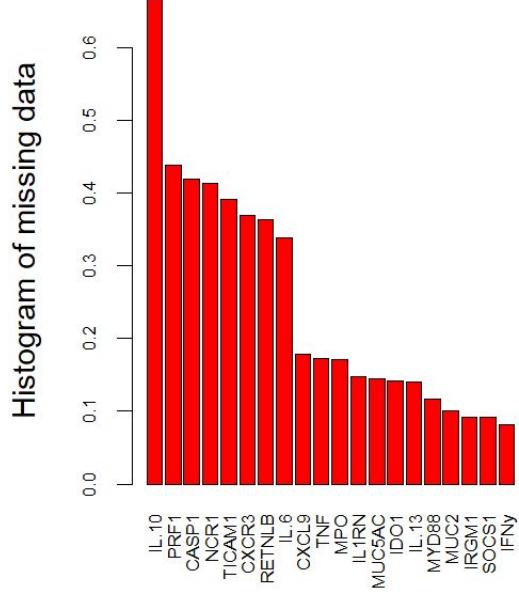
# Parasite clearance, Mucus production



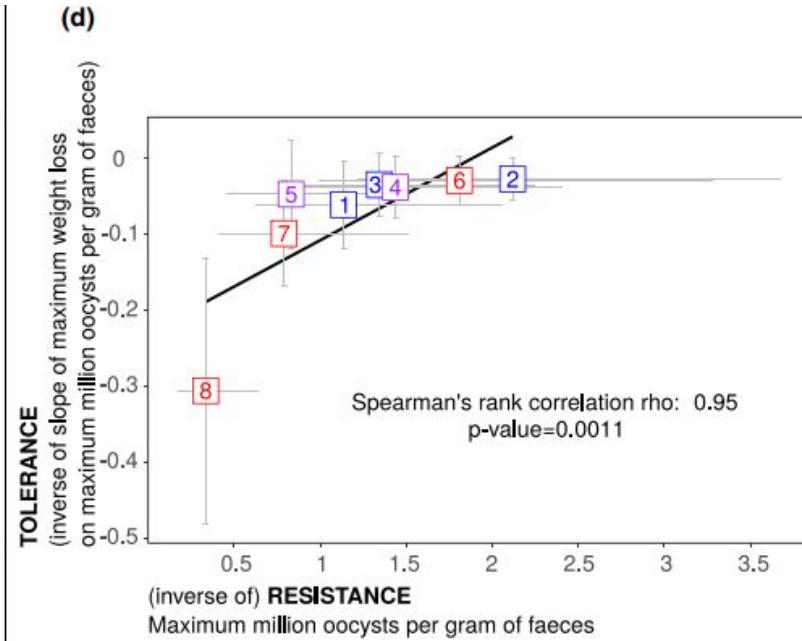
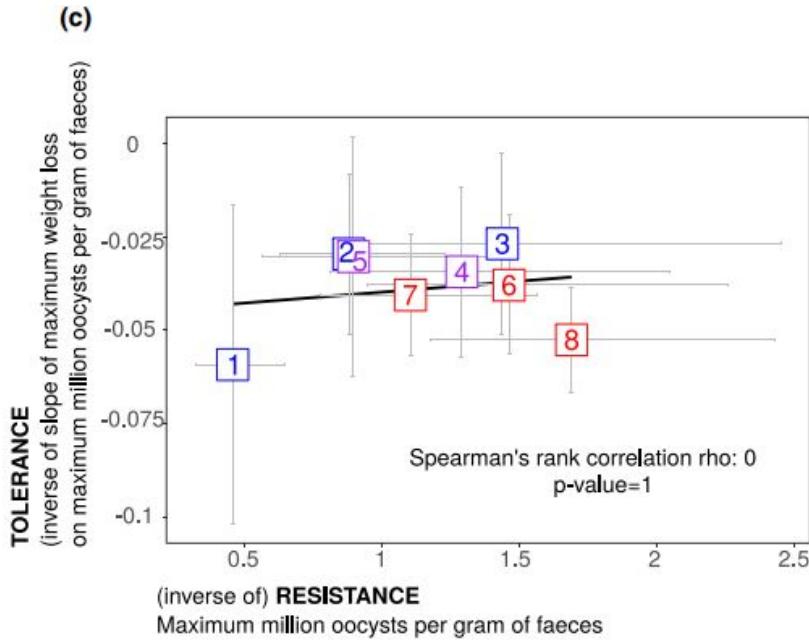
## Contribution of variables to Dim-1



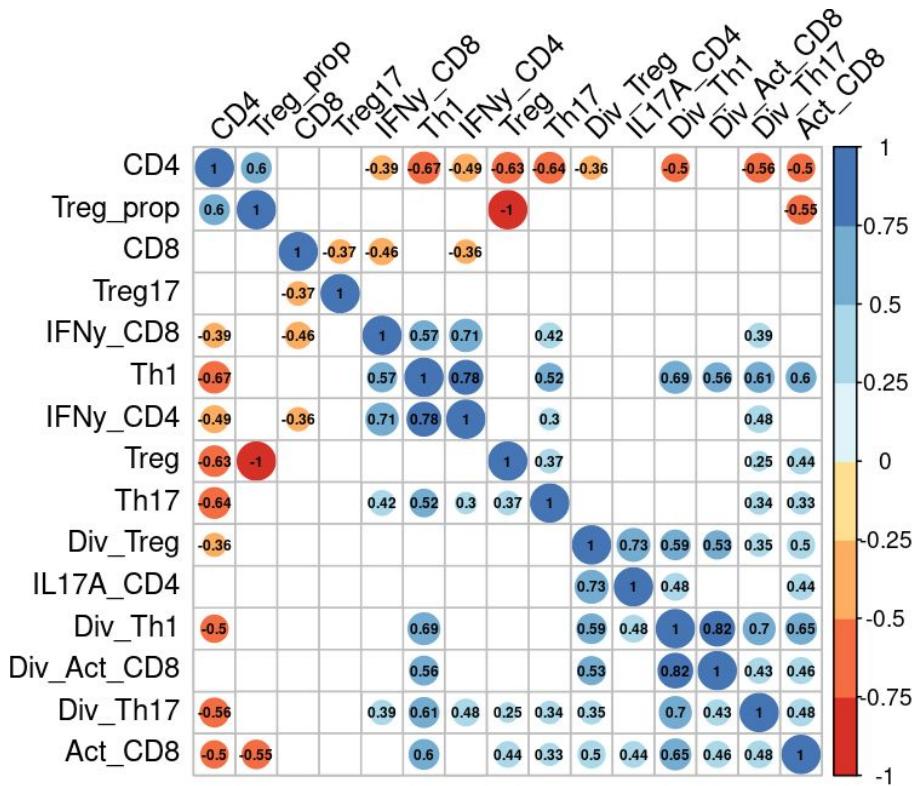
# Imputation mice

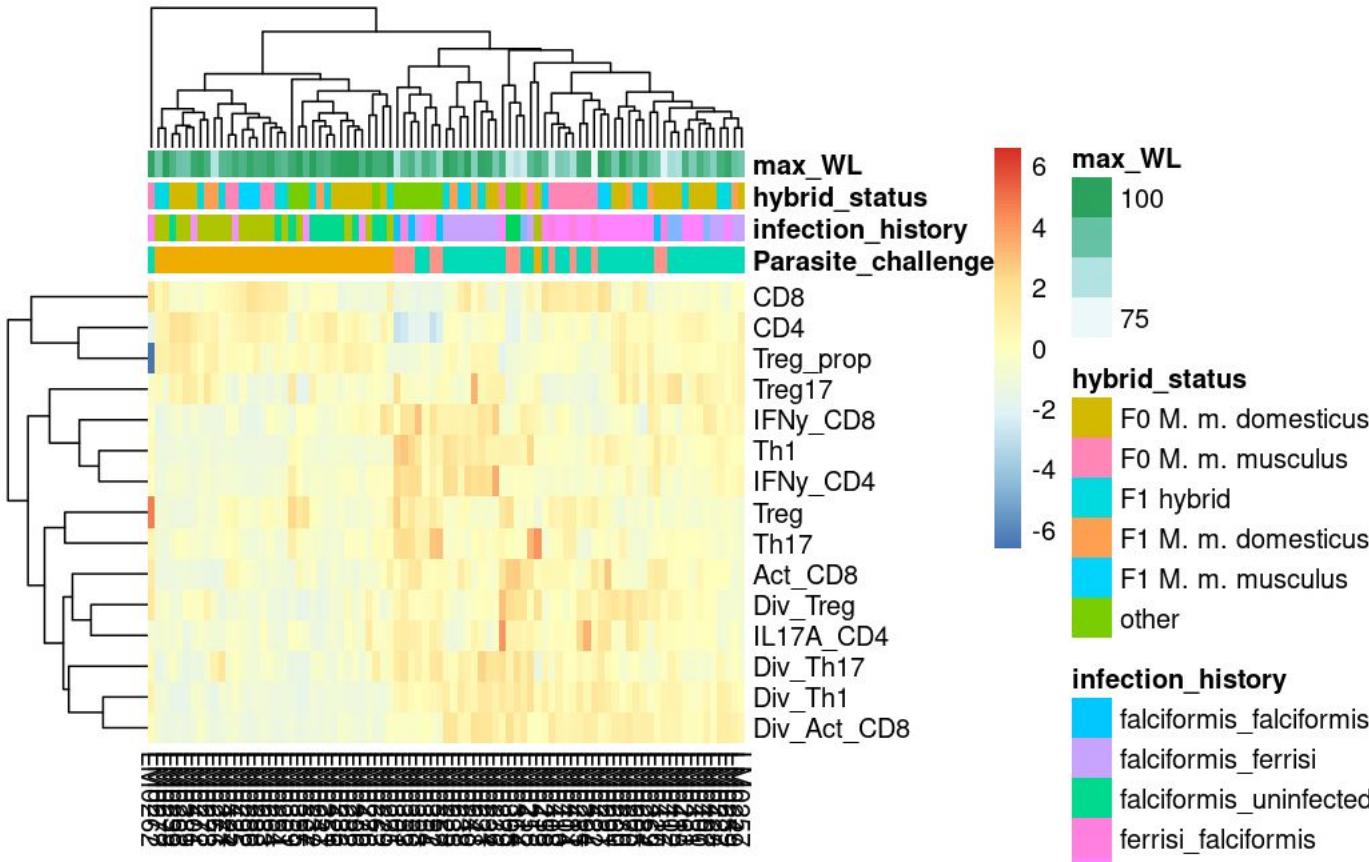


# *E. ferrisi* and *E. falciformis*: How does resistance and tolerance correlate in each species?

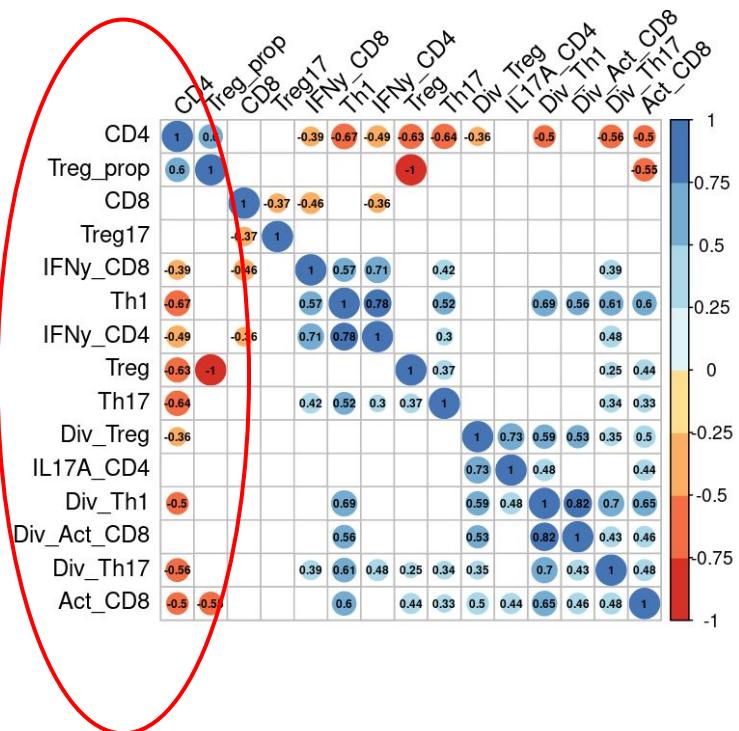


# FACS



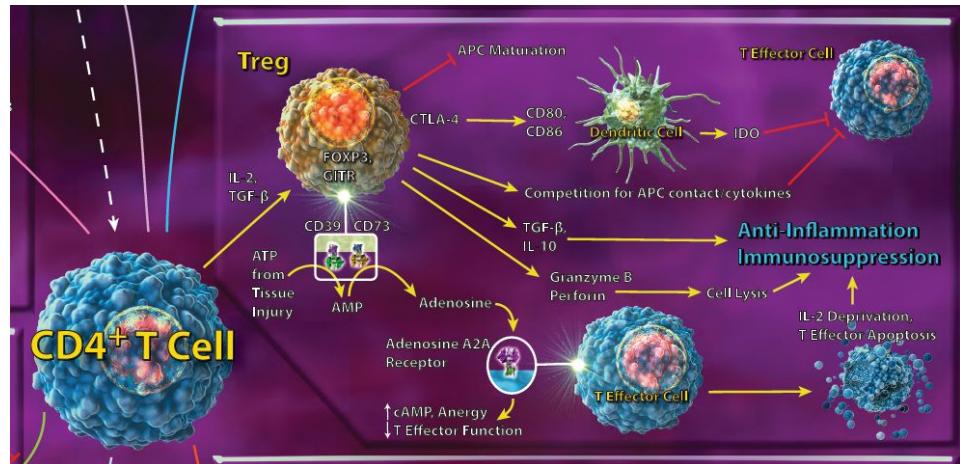


# CD4

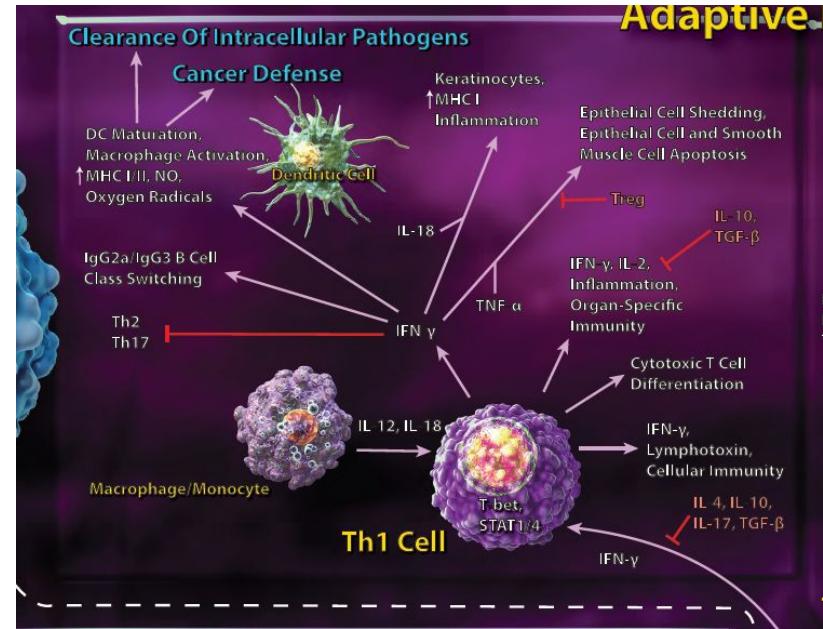
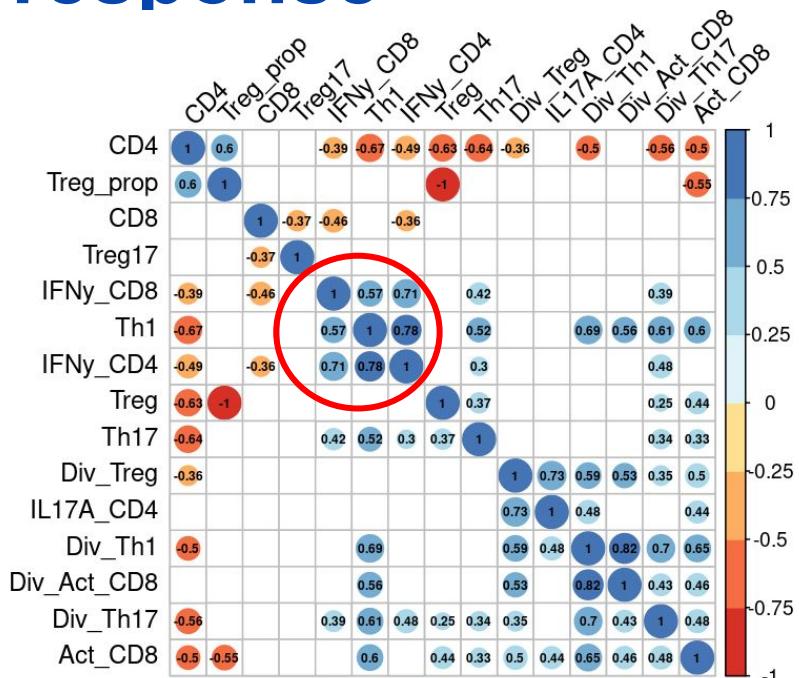


CD4+T:

- activation of the cells of the innate immune system
- critical role in the suppression of immune reaction

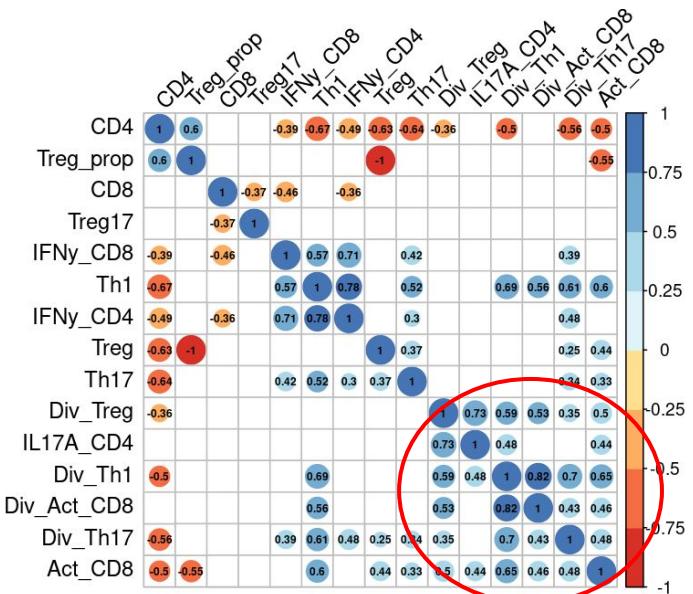


# Pro-inflammatory response

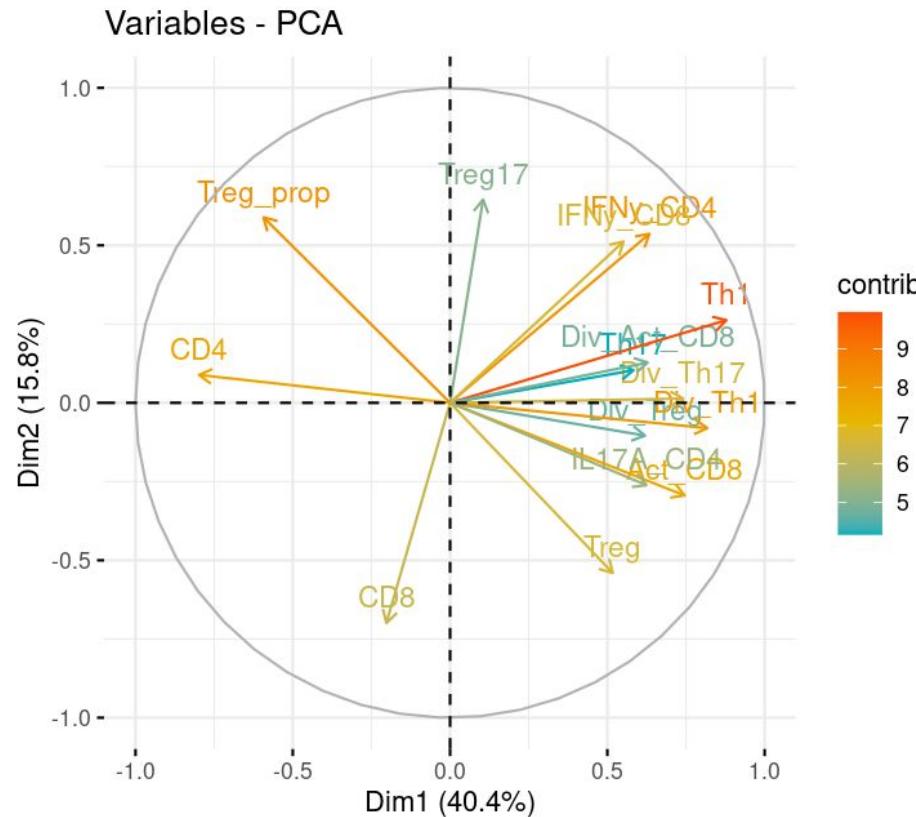


- **IFNy\_CD8: pro - inflammatory IFNy**
- **IFNy\_CD4: pro - inflammatory IFNy**
- **Th1: increased **cell-mediated** response**

# Dividing cells



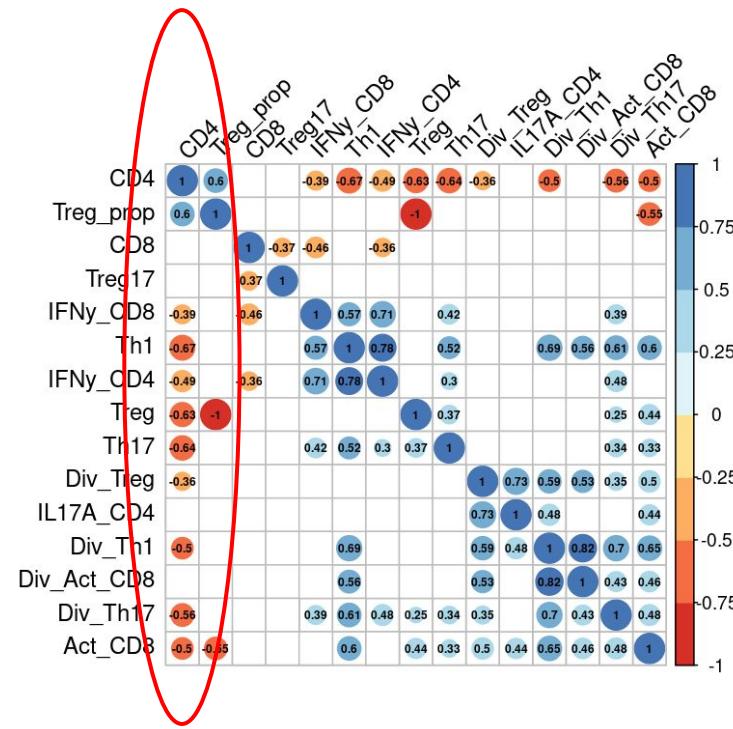
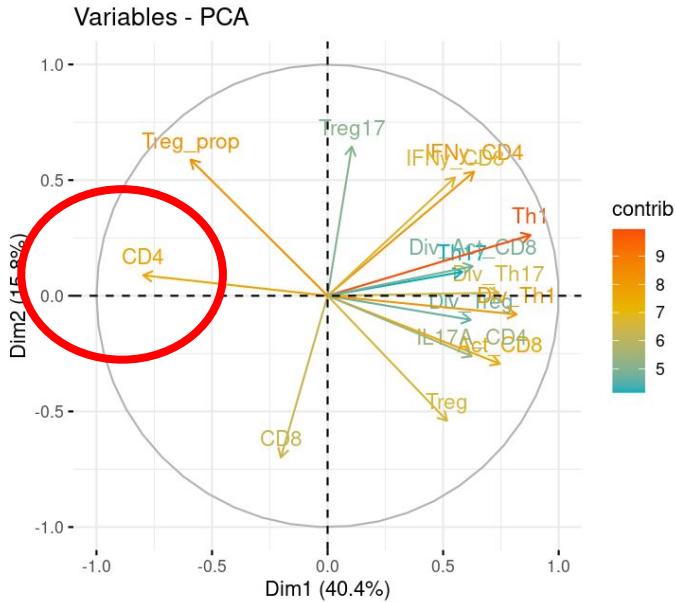
# PCA of FACS data



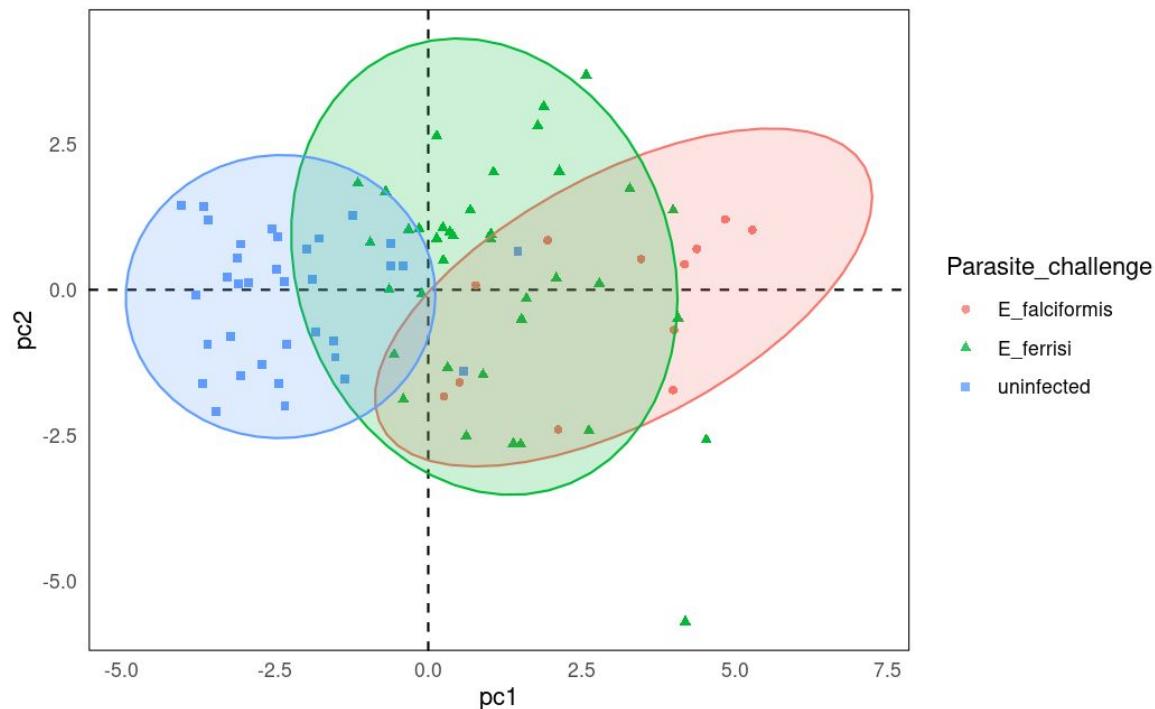
# CD4

CD4+T:

- activation of the cells of the innate immune system
- critical role in the suppression of immune reaction

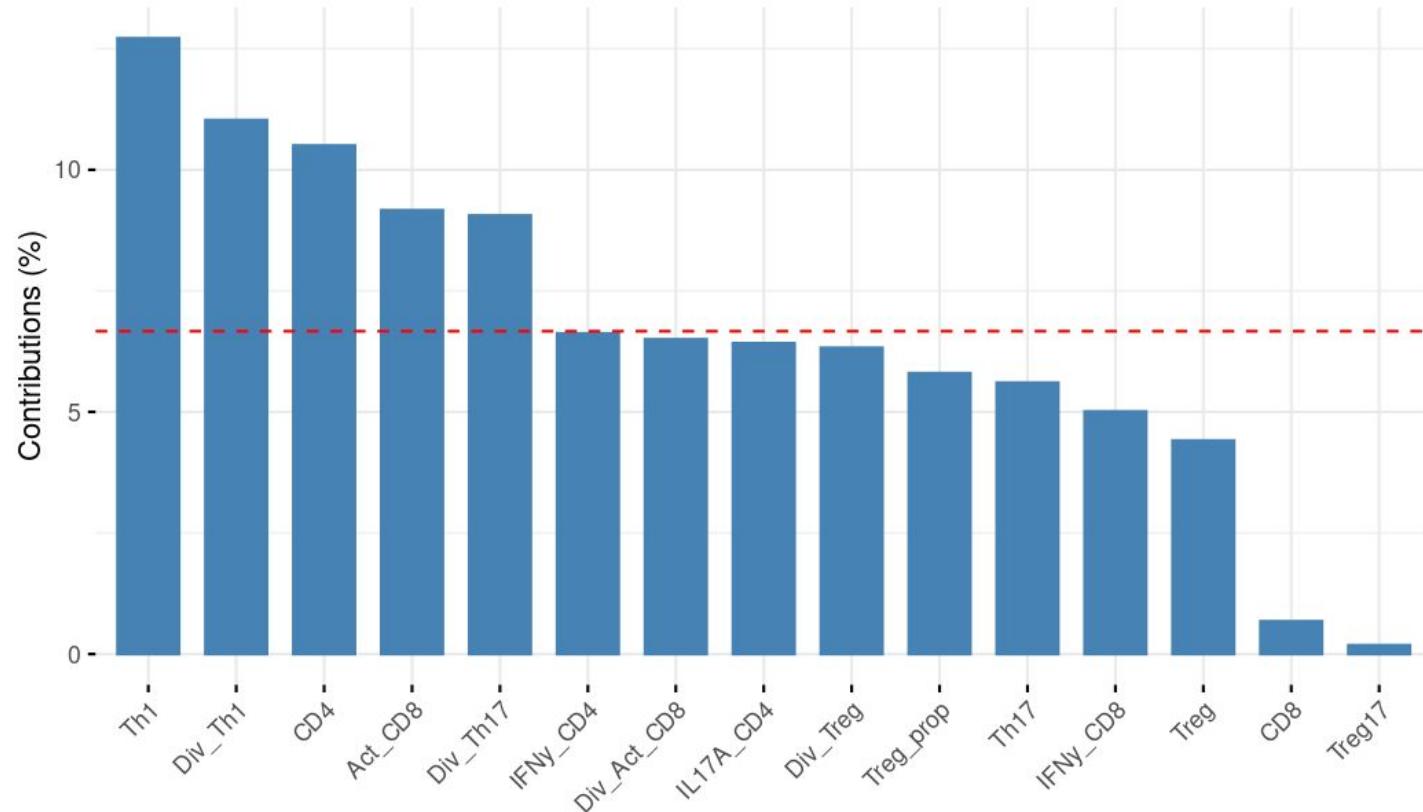


# FACS



# FACS

Contribution of variables to Dim-1



# PCA of immune gene expression

