Ongoing Microbiome Rsearch lines

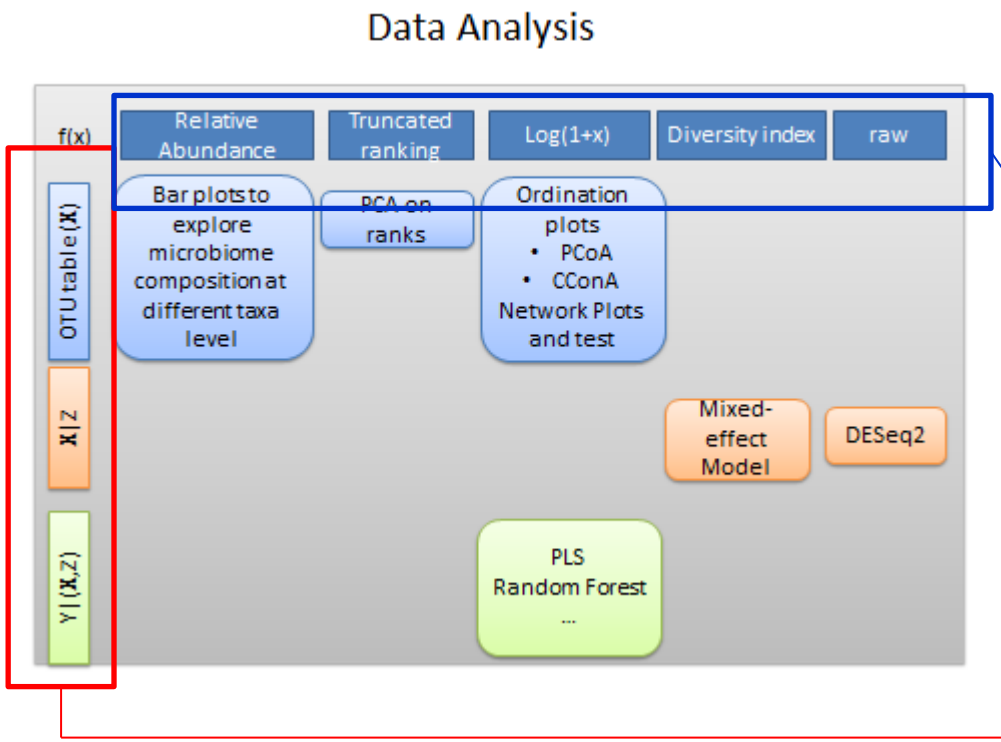
Ziv Shkedy

CenStat/I-BioStat, Hasselt University, Belgium

January, 16, 2017

Rsearch lines

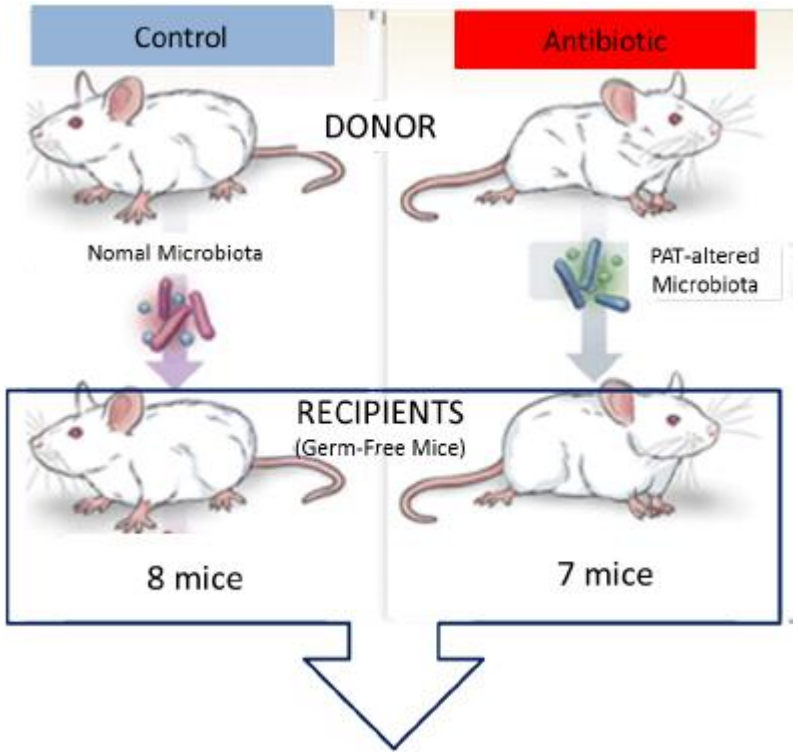
Nolen's presentation:



Similar to Omics data and other high dimensional data:

- Different type of data.
- Different rsearch settings.

Main interest: intervention studies

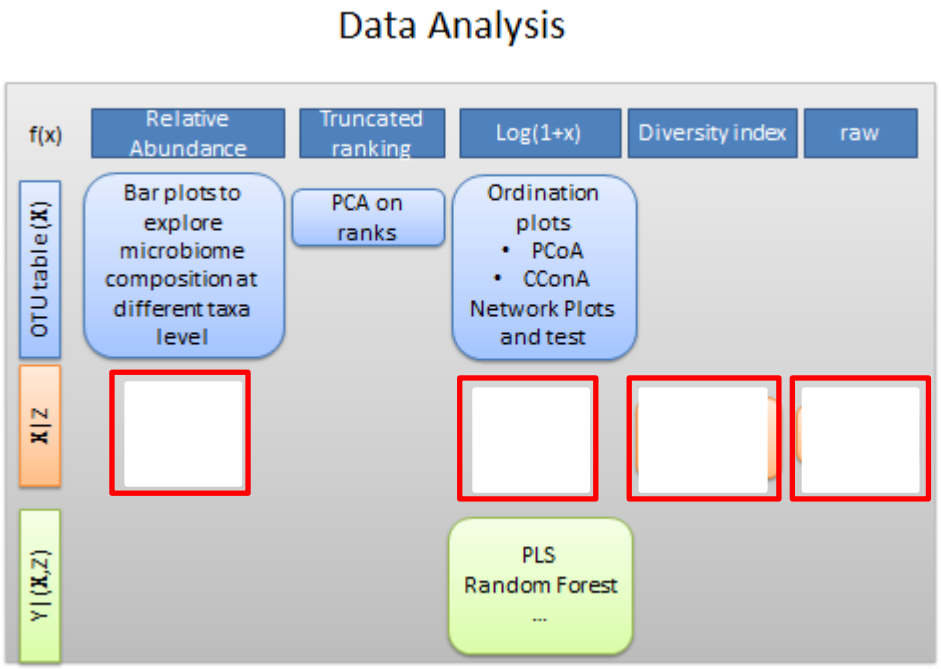


- Transfer microbiome of the donors to the recipient.
- Two groups of donors: treated and control.

Measurements:

1. Microbiome data.
2. Clinical response(s).

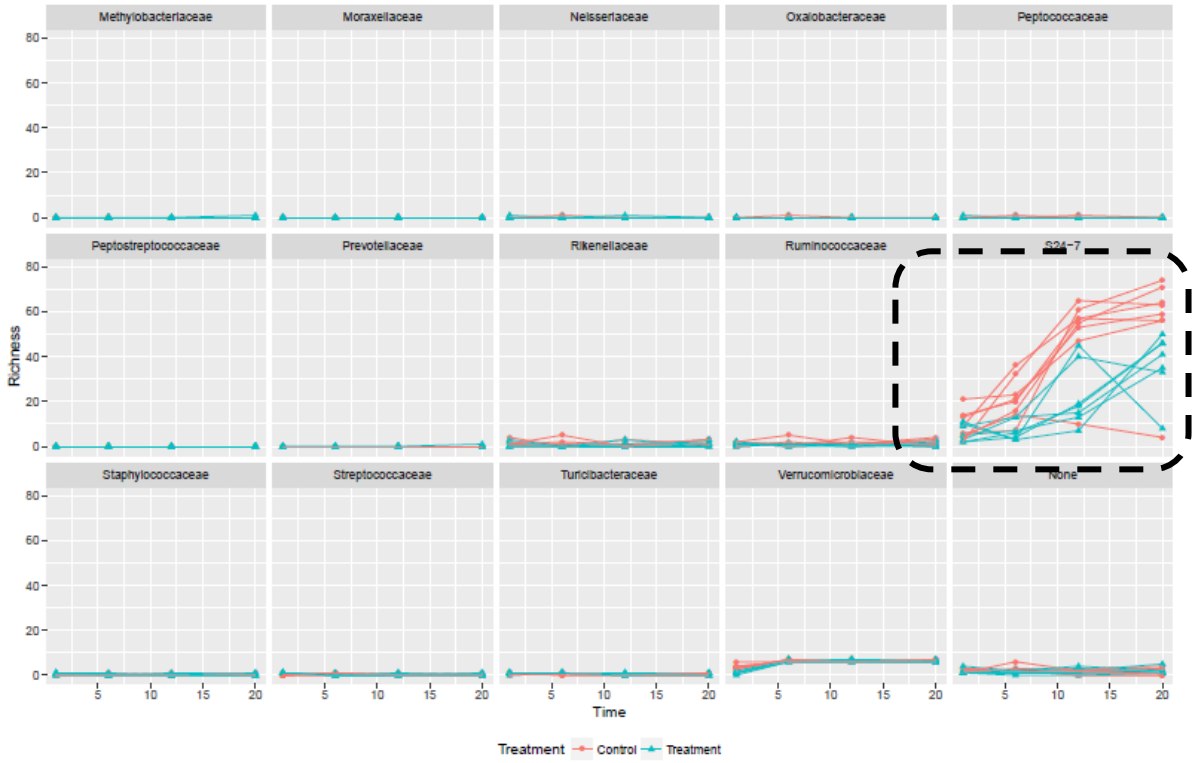
1. Longitudinal analysis of microbiome data



- X: microbiome data.
- Z: treatment.

- Repeated measurements over time.

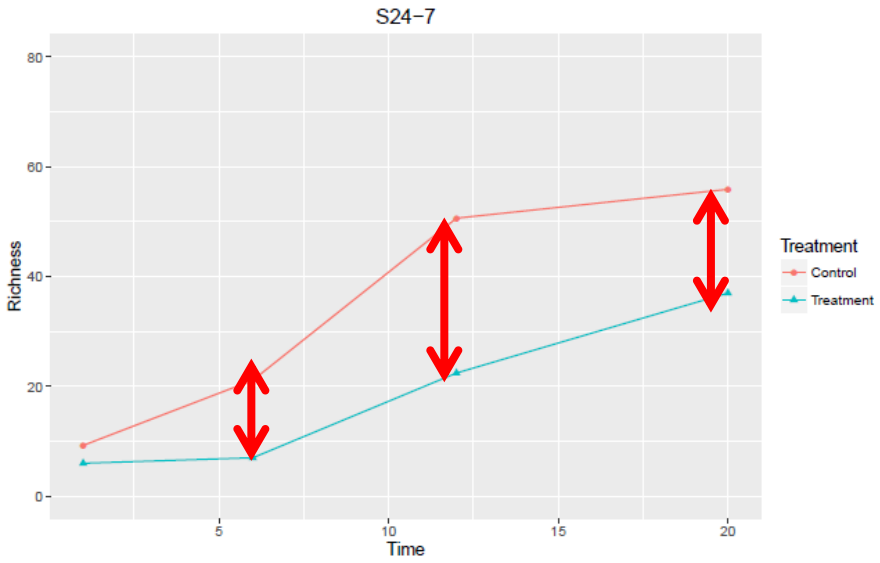
Family-level Richness – Subject Profiles



- Richness: the number of non zero OTUs per subject (per family).
- Two treatment groups.

Not all families are active !!

Modeling richness over time: family level

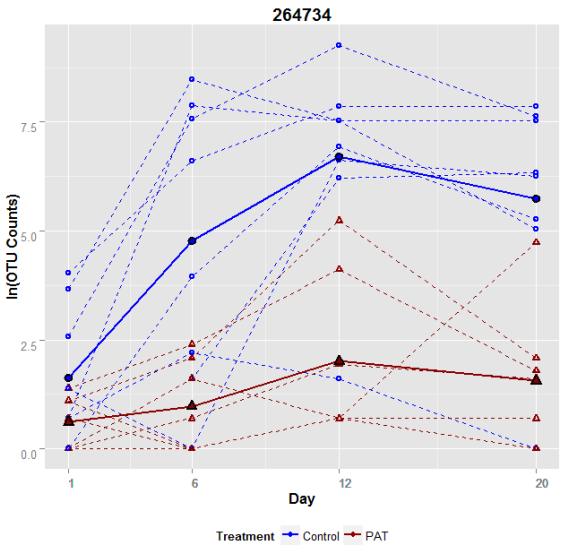


Response: number of active OTUs in the family.

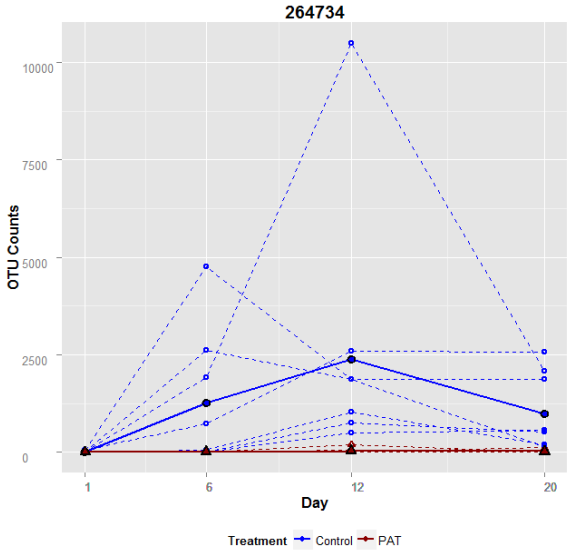
Treatment effect over time ?

Effect of microbiota transfer on OTU Level: 264734

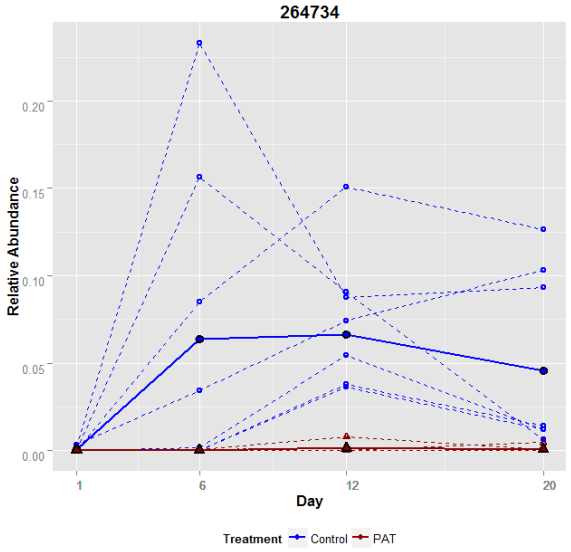
Log(Raw counts+1)



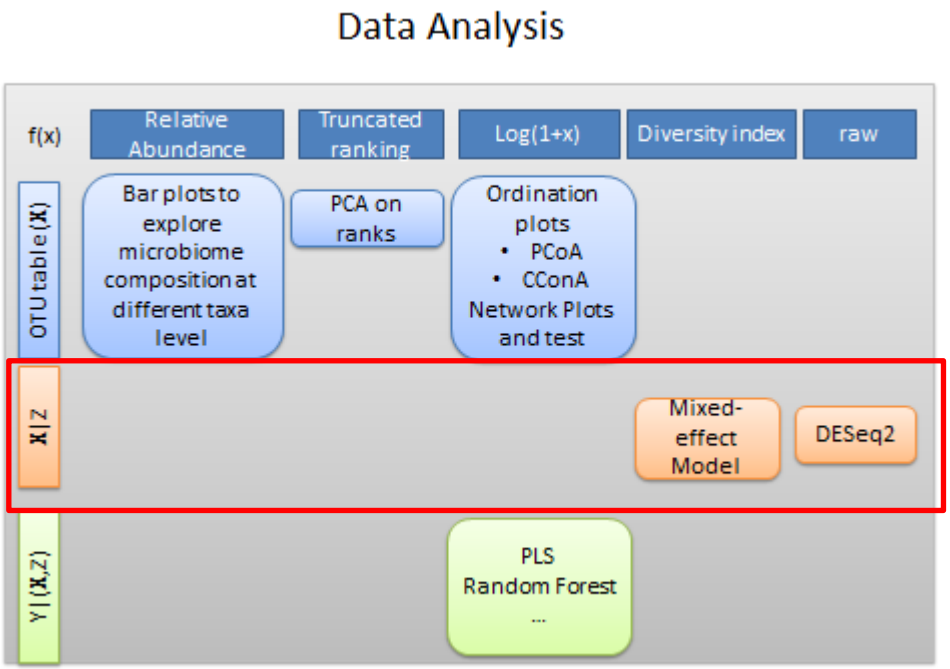
Raw counts



Relative abundance



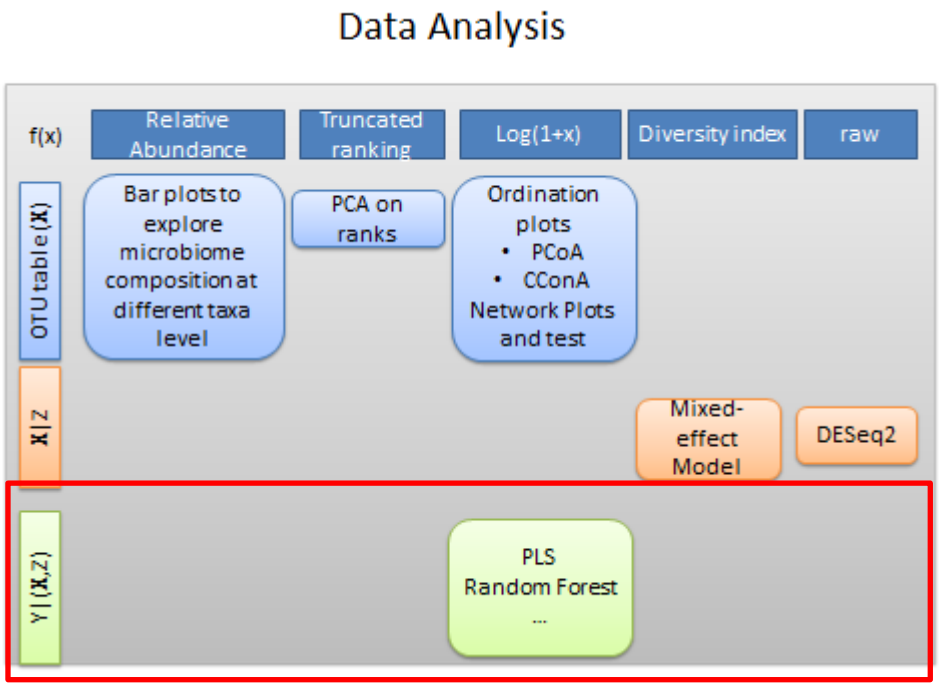
Longitudinal analysis of microbiome data



Issues:

- Zero infelted longitudinal data:
  - Continuous.
  - Counts.
  - Binary
- Over-dispersion data.

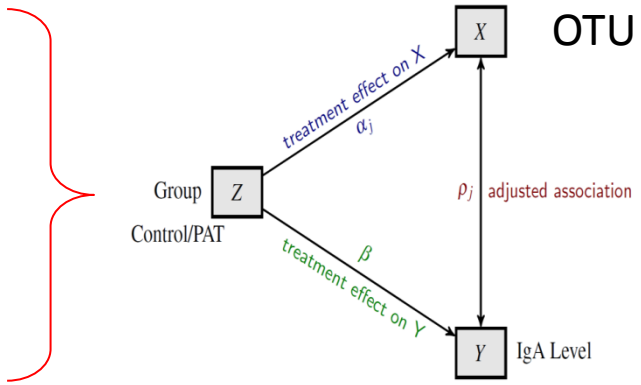
2. Development of Biomarkers



Modeling the effect of microbiome on clinical variables taking into account intervention.

The biomarker setting

- PAT study: IgA and microbiome.
- Analysis at each time point:
  - Joint modeling
    - Log(count)
    - Row counts.
    - Relative abundance.
  - Non-parametric methods.
  - Methods for zero-inflated count data.

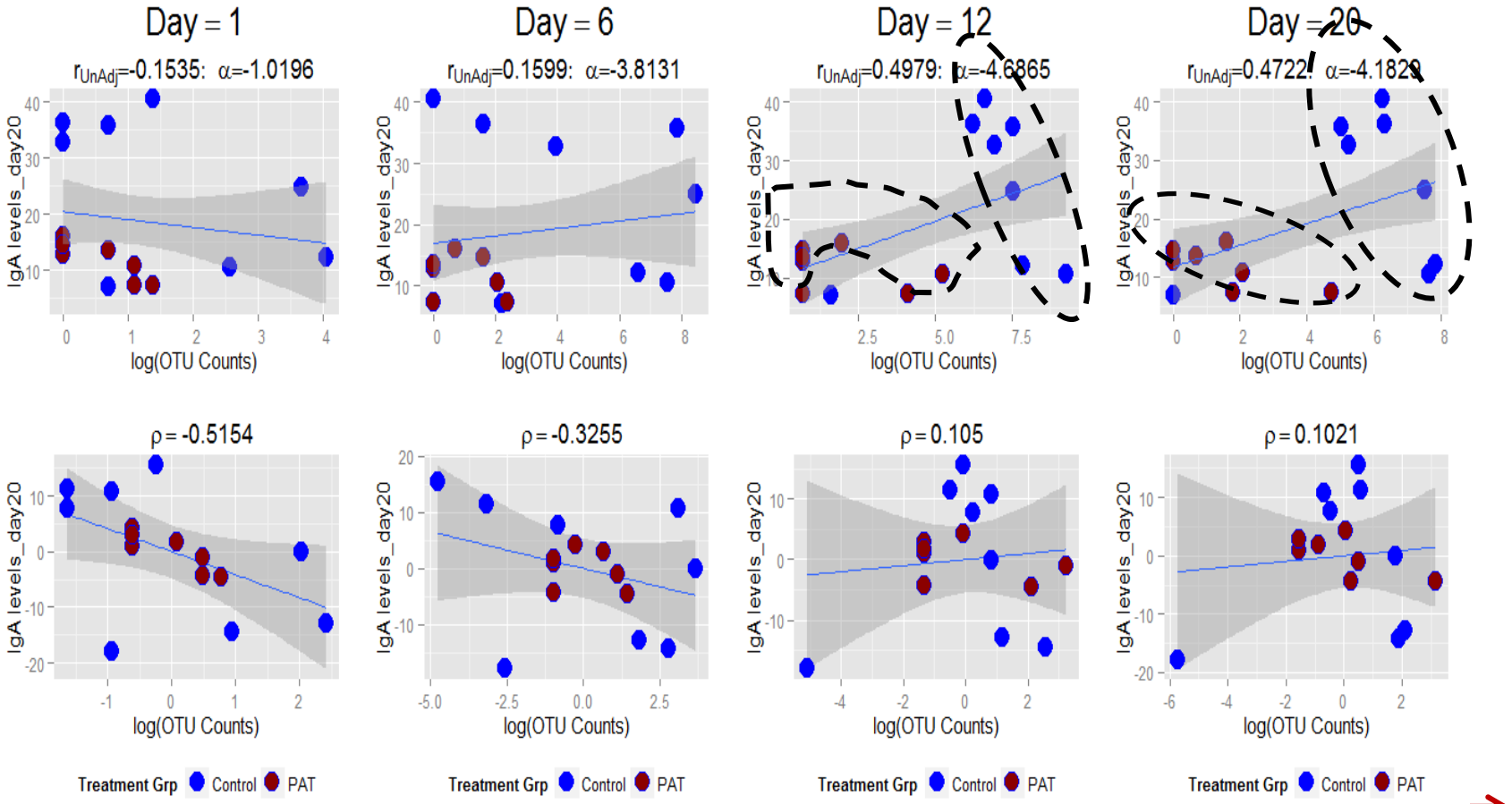


- T1D: survival/log(survival)

Feature specific models

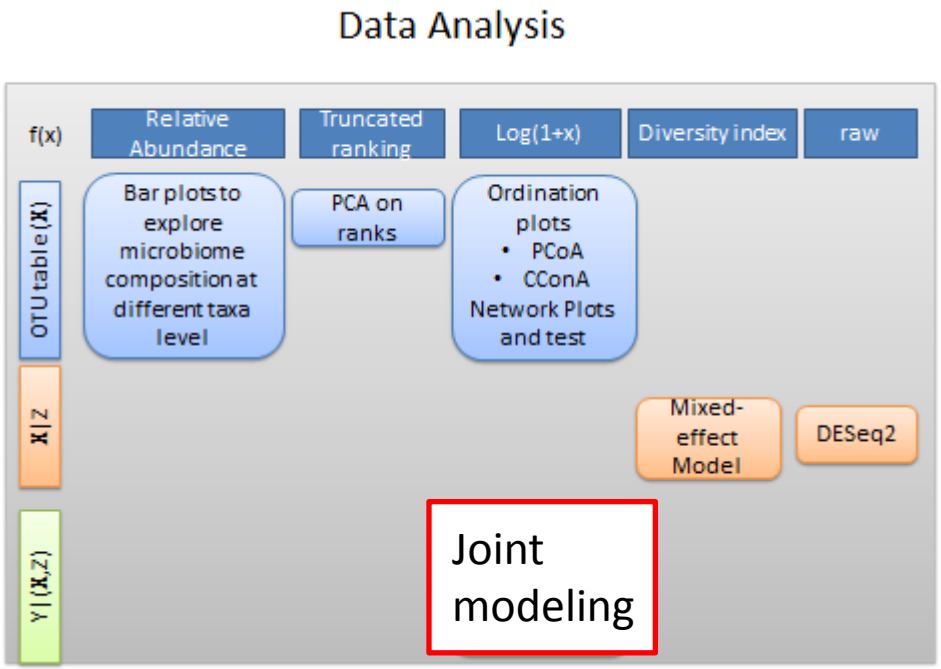
Results for OTU 264734

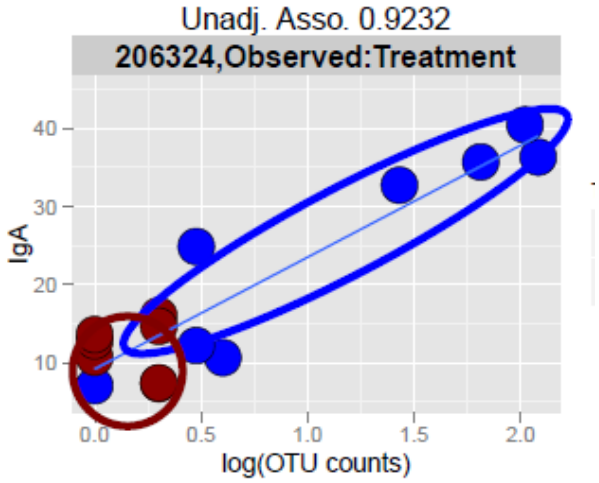
Evolution of treatment effect over time.

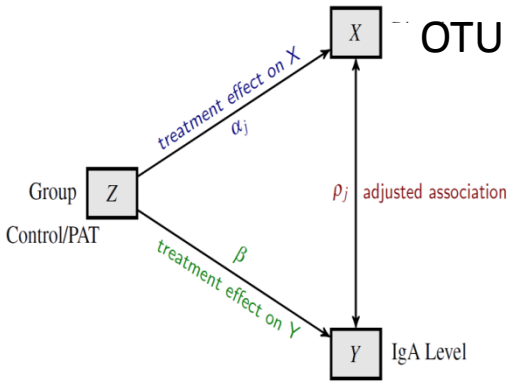


differentially abundant from day 12

2.1: A joint model: OUT 206324



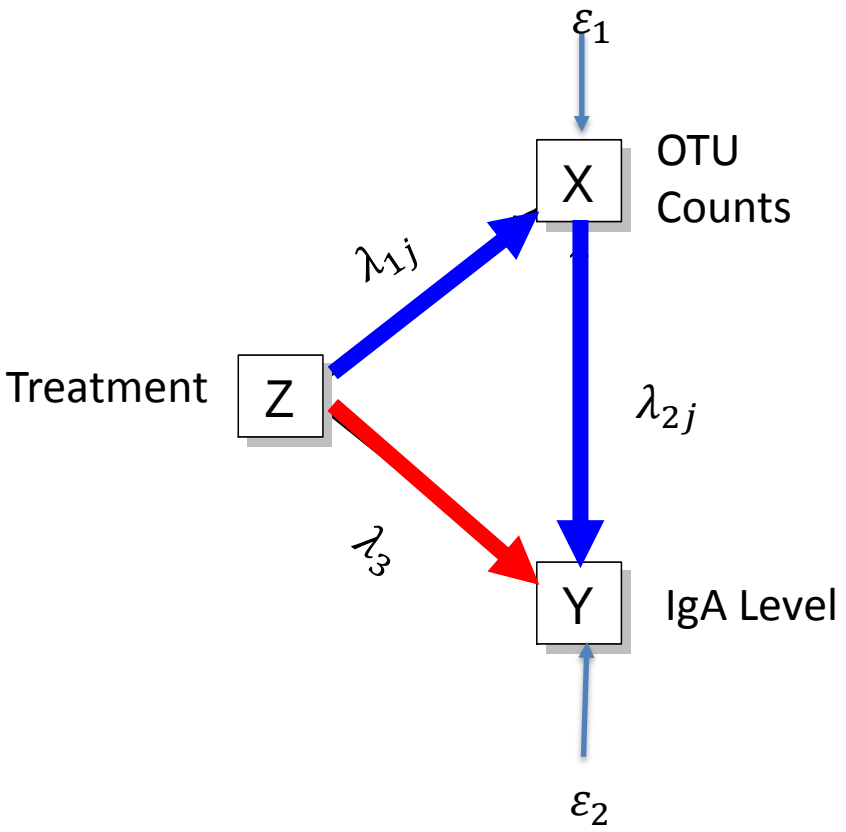




2.2: Structural equations modeling

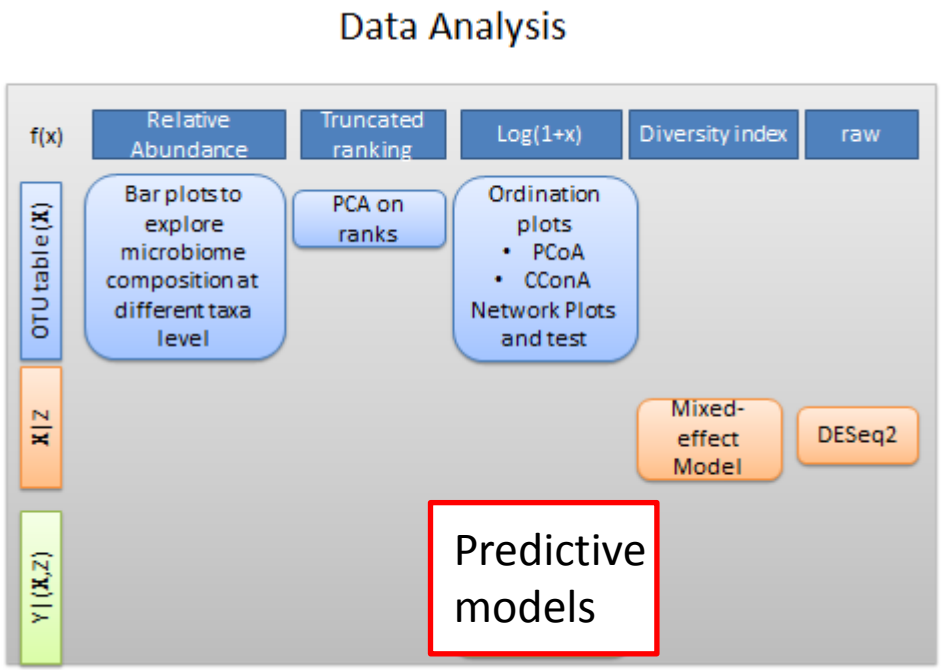
Modeling direct and indirect effect of the treatment on the clinical outcome:

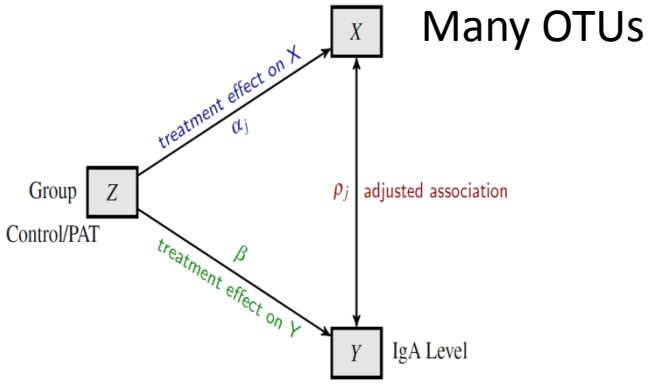
$$X_{ij} = \lambda_{1j} Z_i + \varepsilon_{1i},$$
$$Y_i = \lambda_3 Z_i + \lambda_{2j} X_{ij} + \varepsilon_{2i}.$$



⇨ Example: Rudradev

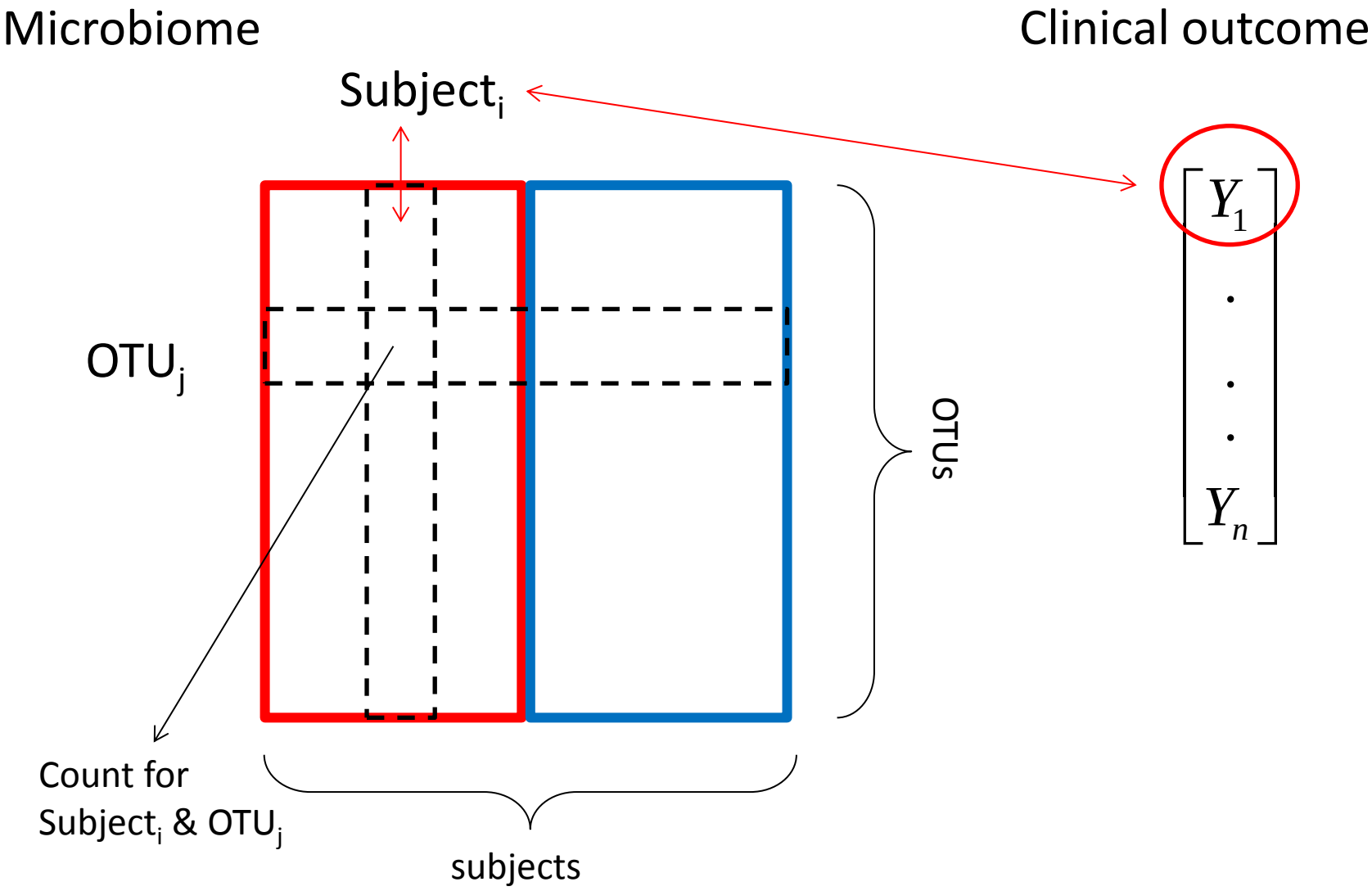
2.3: Predictive models





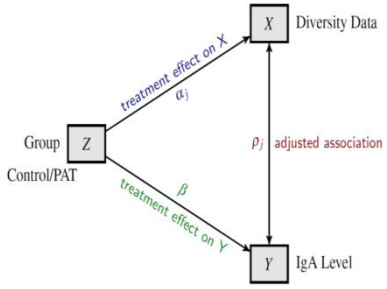
Can we predict the clinical outcome using microbiome data (taking into account the treatment effect) ?

Data Structure



Predictive models for clinical outcomes

Analysis at each time point:



$$Y_i \mid Z_i, OTU_1, \ldots OTU_M$$

Predictive models for clinical outcome:

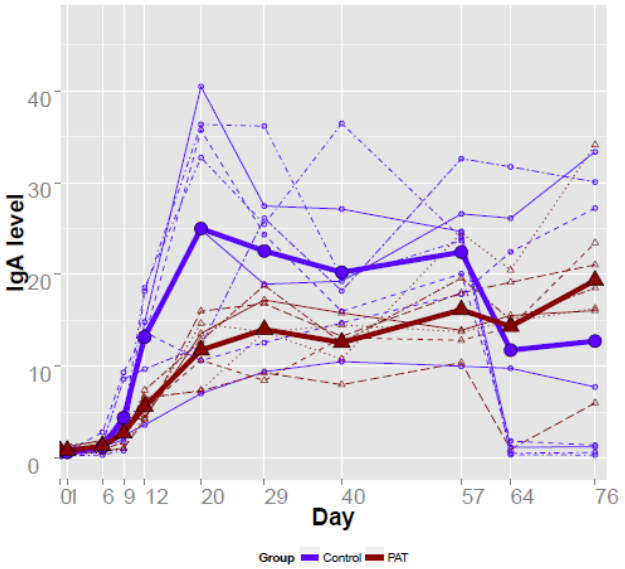
$$Y_i = Z\gamma + \sum_{j=1}^{M} \beta_j OTU_{ij} + error_i$$

$$E(Y_i) = f(treatment, OTU_1, OTU_2, ..OTU_K)$$

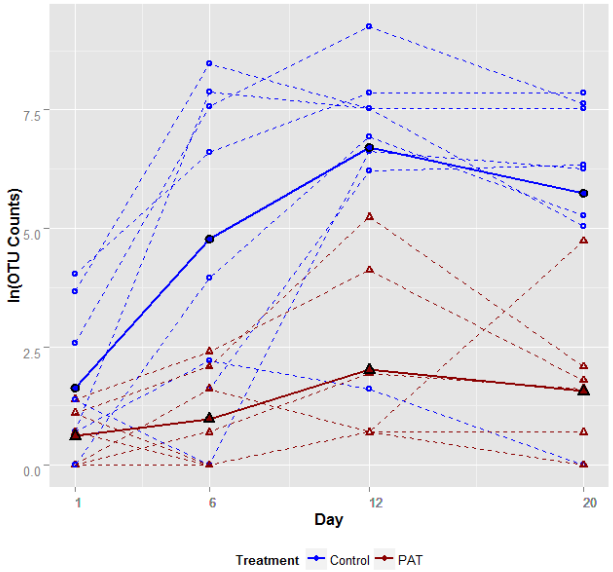
2.4: A joint modeling - longitudinal data

- Joint model for the longitudinal data.
  - Joint models for IgA and counts/relative abundance/abundance.
  - Evolution of treatment effects over time.

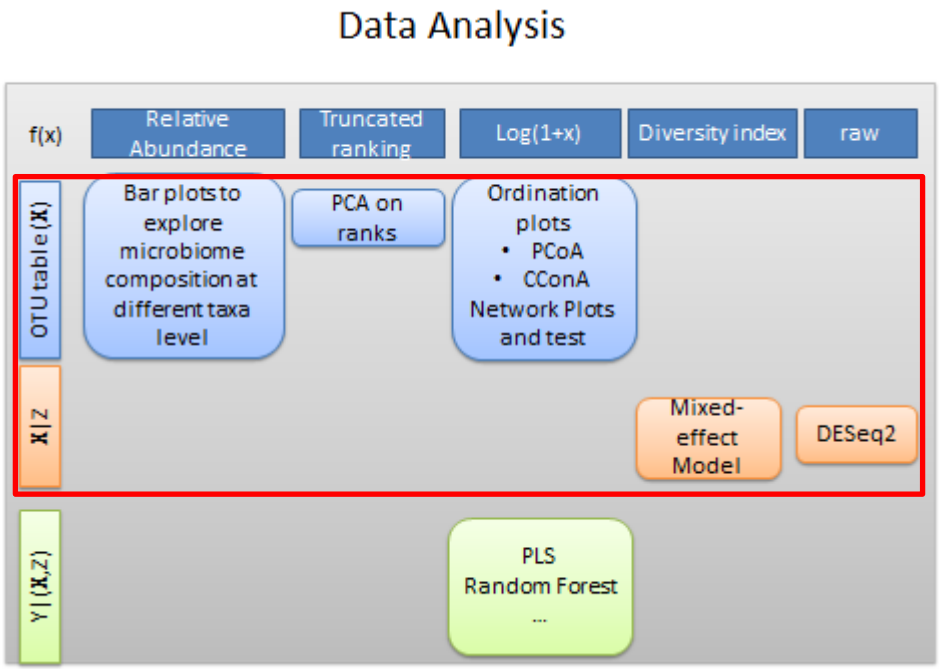
IgA level



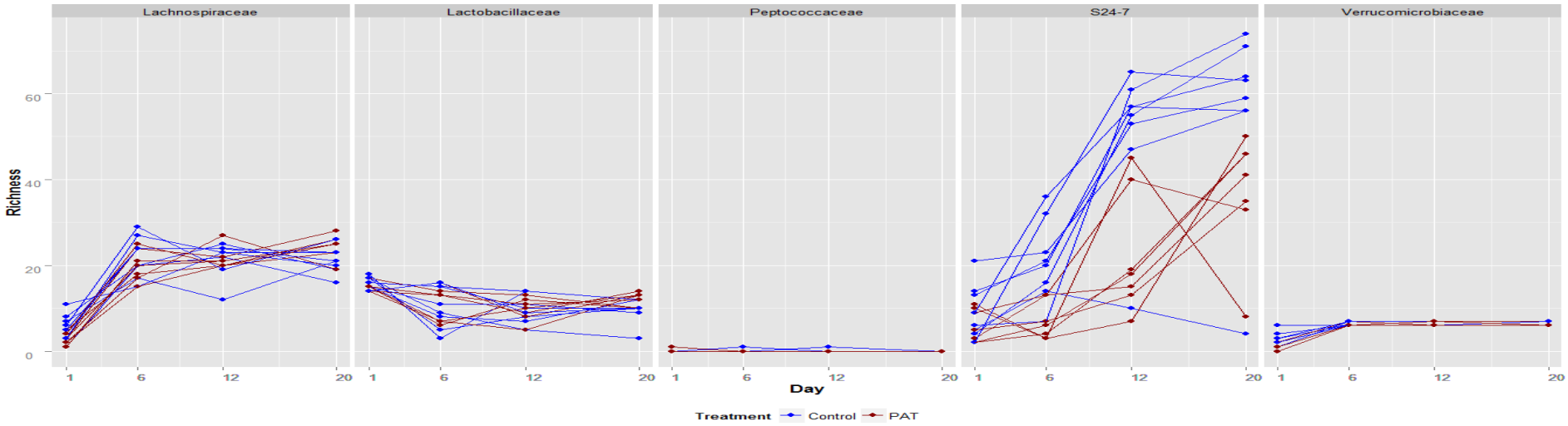
OTU 264734



3. Population models for the evolution over time by family



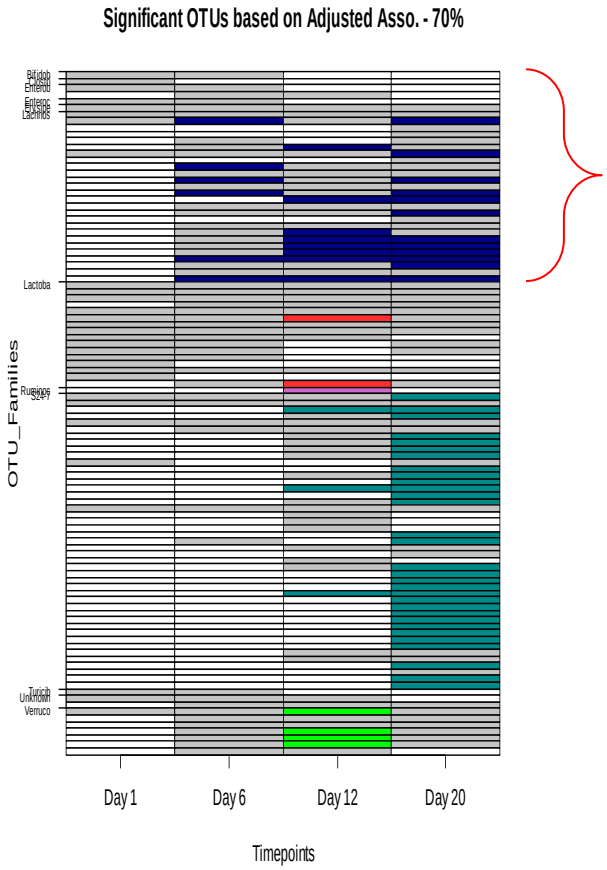
3.1: Family-based analysis: richness

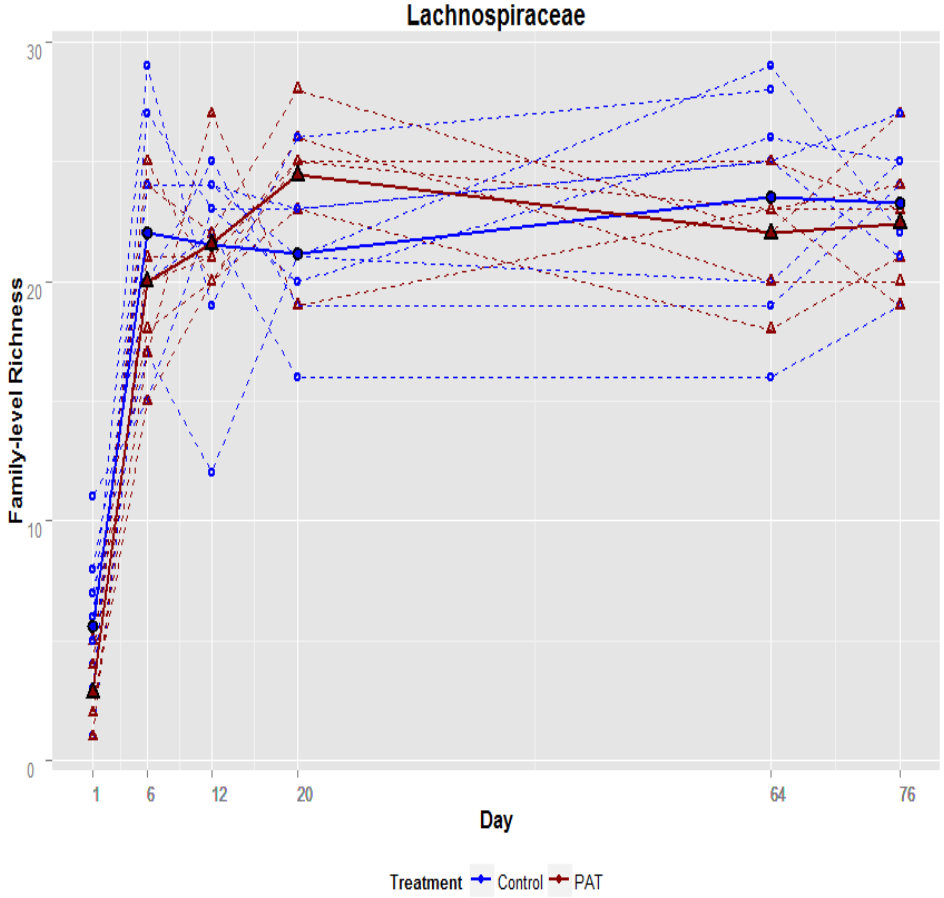


$$\frac{\partial N(t)}{\partial t} = f(time, treatment)$$

- Non linear mixed effects modeling for the richness (family size in terms of OTUs activity).
- Modeling intervention.

The Lachnospiraceae Family





3.2: correlation among OTUs

- Modeling interaction between OTUs

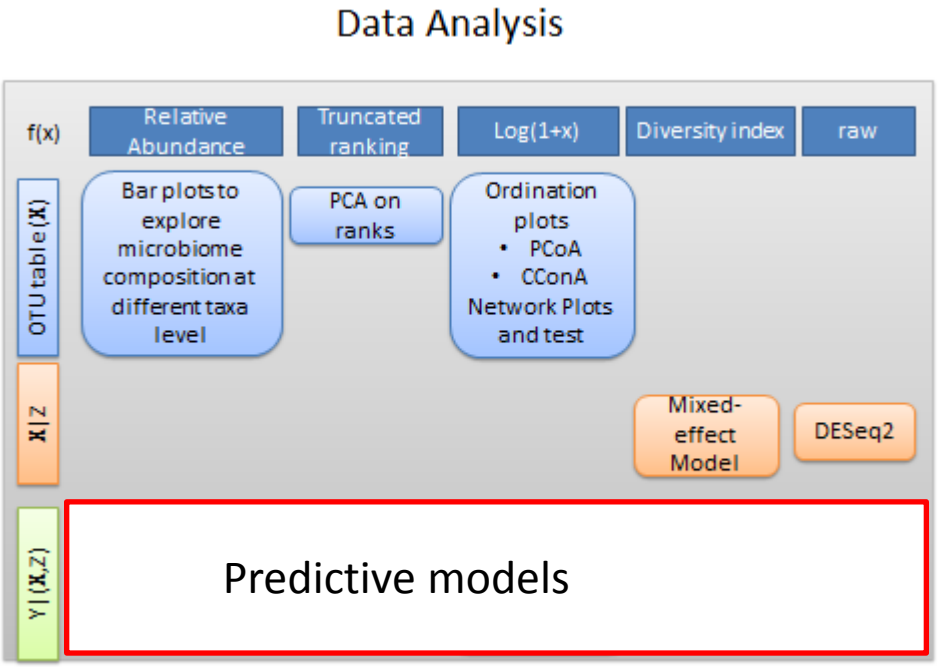
$$\Delta Y_i(t) = \log\left(Y_i(t)\right) - \log\left(Y_i(t-1)\right)$$

$$\tilde{Y}_i(t) = \frac{\Delta Y_i(t)}{\Delta t}$$

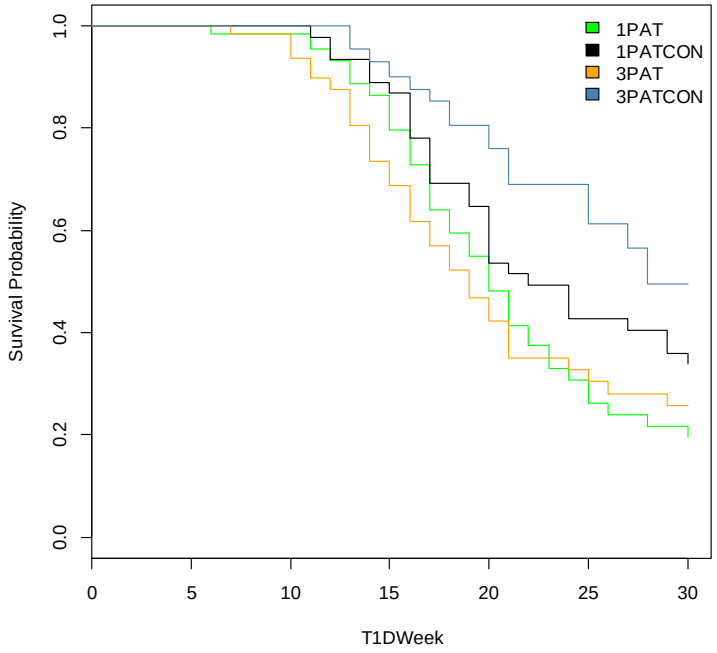
$$\tilde{Y}_i(t) = \beta_0 + \sum_{k=1}^{K} \alpha_{ik} Y_k + \varepsilon_{it}$$

- LASSO/EN models.

3.4: Predictive models for diabetes based on microbiome data



Clinical response: time to event data



Can we predict the time to event using microbiome data ?

3.5: Risk score for diabetes

- Modeling high/low risk for diabetes using microbiome risk score.

$$R_i = f(OTU_1, OTU_2, ..., OTU_K)$$

Risk score to develop diabetes

$$\text{Risk group} = \begin{cases} low & R_i \le \theta \\ high & R_i > \theta \end{cases}$$

$$S(t) = f(RG_i)$$

Can we find a significant different between the risk group in terms the time to developing diabetes ?