Modelling Biological Data:
Second RiboMed Summer School

Nuno Morais and Mariana Ferreira

Statistical modelling for splicing analyses

Tentative Program

23 June		
		Requirements
Session 1	Quantification of alternative sequence inclusion levels from RNA sequencing data	

10:00 - 11:00 (theoretical)	• Introduction of the speakers and their research interests
	• Previous lab work on alternative splicing regulation
	• Estimating alternative sequence inclusion levels from RNA-sequencing data:
	  ○ Sources of uncertainty in individual samples
	  ○ Similarities and differences from gene expression quantification
	  ○ Different approaches and their interpretations	

short break		

11:10 -12:10 (hands-on)	• The (mathematical) nature of the percent spliced-in (PSI) ratio
	  ○ Constrained range, intrinsically normalised, logit transformation
	  ○ Relationship with the beta distribution
	• Differential splicing with many samples (*psichomics*)
		`logit()`
		`rbeta()`
		`fread()`
		R packages:
		`car`
		`data.table`
		`psichomics`

lunch break		
Session 2	Differential splicing analyses	
13:10 - 14:10 (hands-on)	• Modelling the estimation uncertainty in individual samples while accounting for variability among replicates using beta distributions	Previously processed tables from GTEx (provided) R packages: *`data.table`* *`ggplot2`* *`psichomics`*
short break		
14:20 - 15:20 (hands-on)	• Other differential splicing approaches (e.g., non-parametric tests) • Sum-up and assessment of learning outcomes	

Overall learning goals:

- Identify and understand the sources of uncertainty when estimating the inclusion levels of alternative sequences from RNA sequencing data and how those impact differential splicing analyses;
- Estimate the effect size of alternative splicing changes between biological conditions and its statistical significance using beta distributions for the analysis and visualisation of inclusion levels of alternative sequence from an event-centered perspective.