Biomarkers, Experimental Designs, Linear Models, Omics Data and Diagnostic measures and Validation

One approach to meet them all

Alex Sanchez-Pla
Departament de Genètica, Microbiología y Estadística.
Universitat de Barcelona (UB)
2024-02-19

A typical scenario

A common situation in biomedicine and other fields:

- An individual can be in one of two states: healthy or unhealthy
- Goal: Detect unhealthy state as son as posible so that preventive actions can be taken.
    - Unhealthyness is not always obvious, particularly at early stages.

This situation appears in many contexts, not only nutrition:

- In nutritional studies it is necessary to find out if people have eaten a given nutrient or if they adhere to a certain type of diet, even if there are no nutritional questionaire available.
    - In this case it is common to try to find it out from the analysis of metabolomic data.

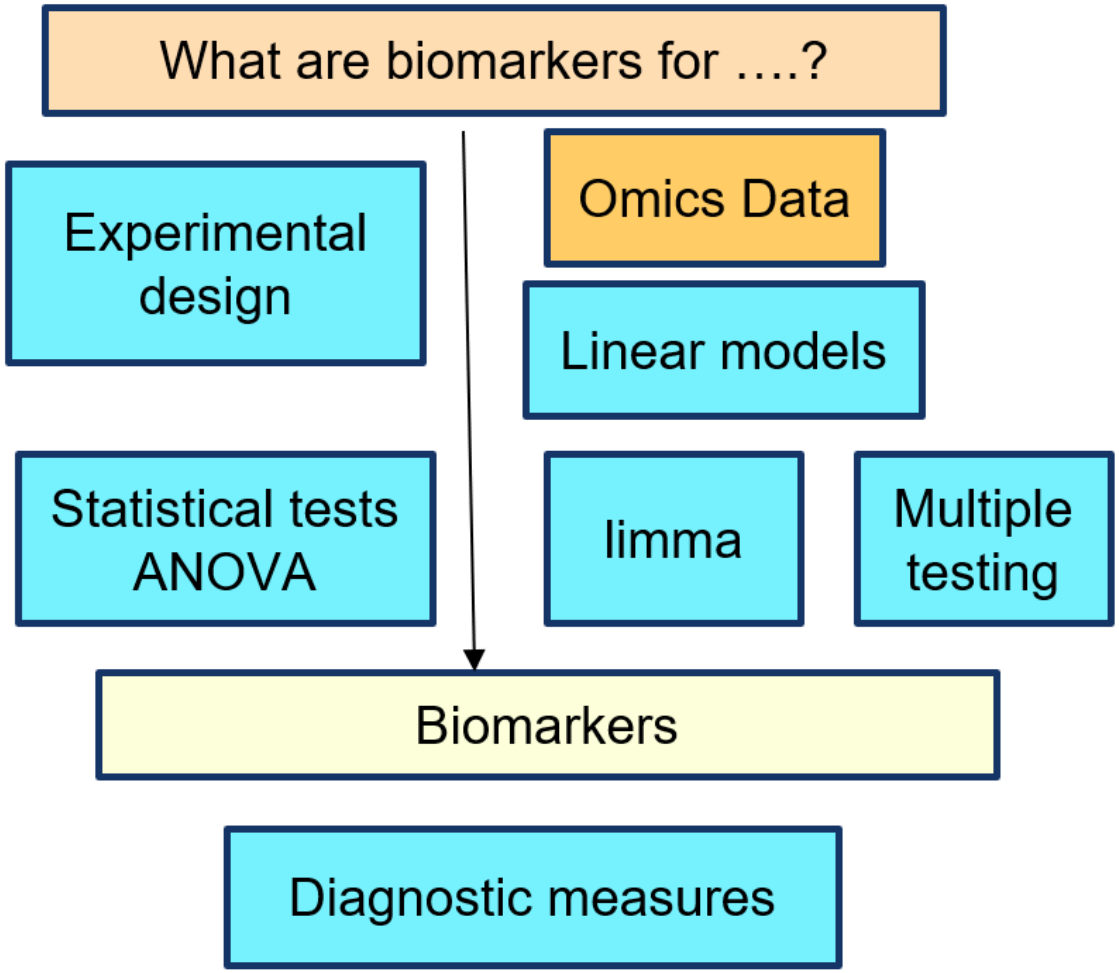
Building biomarkers

- A (bio)markers for the *unhealthy state*, is *anything*, whose values are different among healthy and unhealthy.
    - Healthy/Unhealthy is only one possibility!
- Not anything that differs is a useful biomarker: The difference between conditions must be *consistent enough* to be used to distinguish each other group, not only on the samples used to build the biomarker but in (any) other independent datasets.
- It should happen whith *error rates as small as possible* in both
    - False positive errors: Declaring unhealthiness in healthy cases
    - False negative errors : Declaring health in unhealthy cases

The road to biomarkers

- Given a hypothesis about how a certain condition or disease works,
- An experiment must be designed to identify its potential biomarkers.
- Statistical analysis of the experiment results can provide a list of candidate biomarkers. This may require
  - Different statistical methods: Each experimental design may have an ANOVA model suitable for its analysis.
  - If *omics* data are used, high dimensionality and, often small sample size has to be accounted for.
- The candidate biomarkers list must be validate against an independent dataset to determine their suitability as biomarkers

The road to biomarkers



Let's go for it!