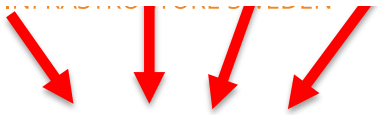
Towards different biological conditions - what to think about

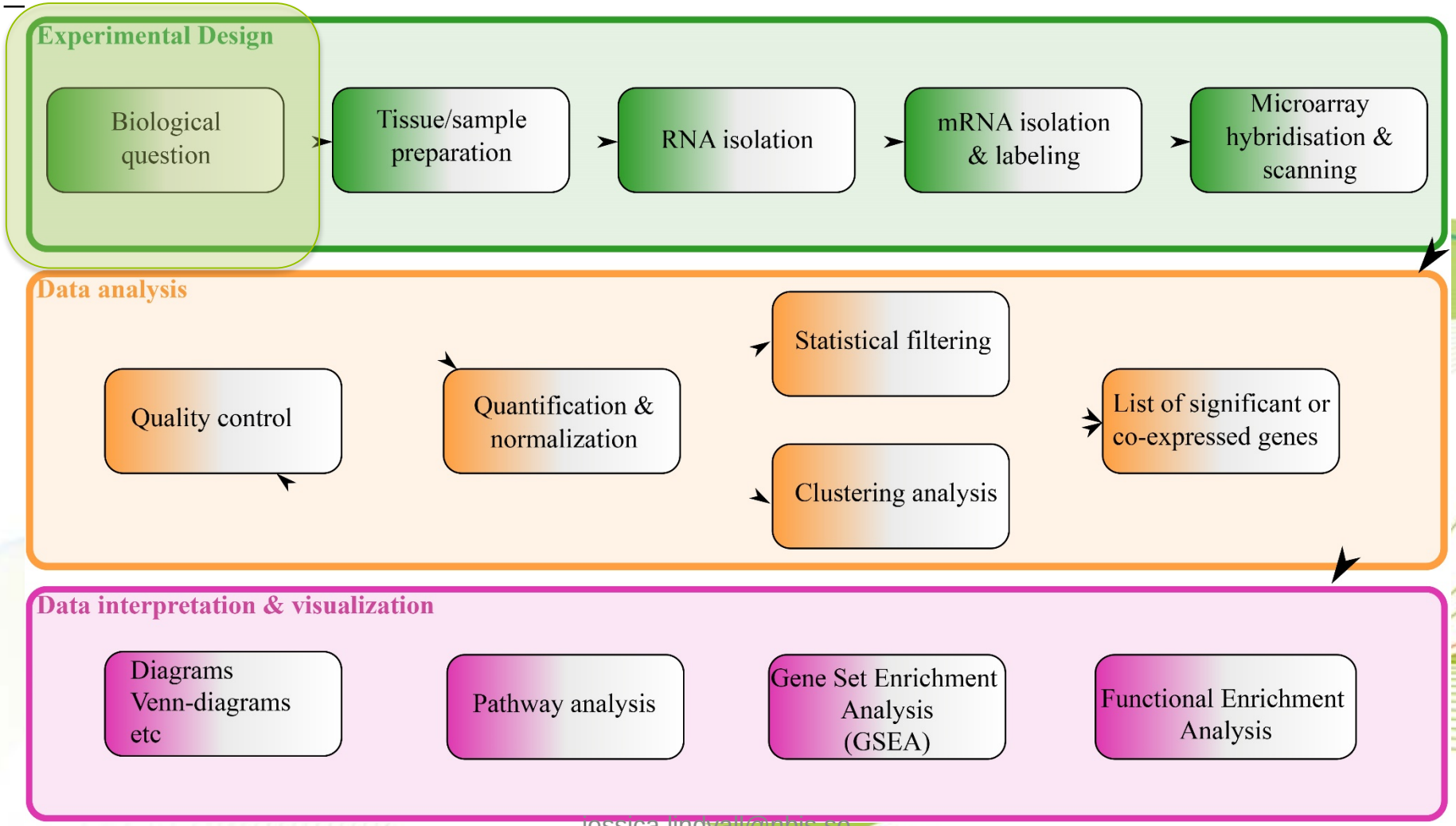
Jessica M. Lindvall, PhD

NBIS Bioinformatics expert, Training coordinator
jessica.lindvall@nbis.se

DESIGN
Statistics+Biology

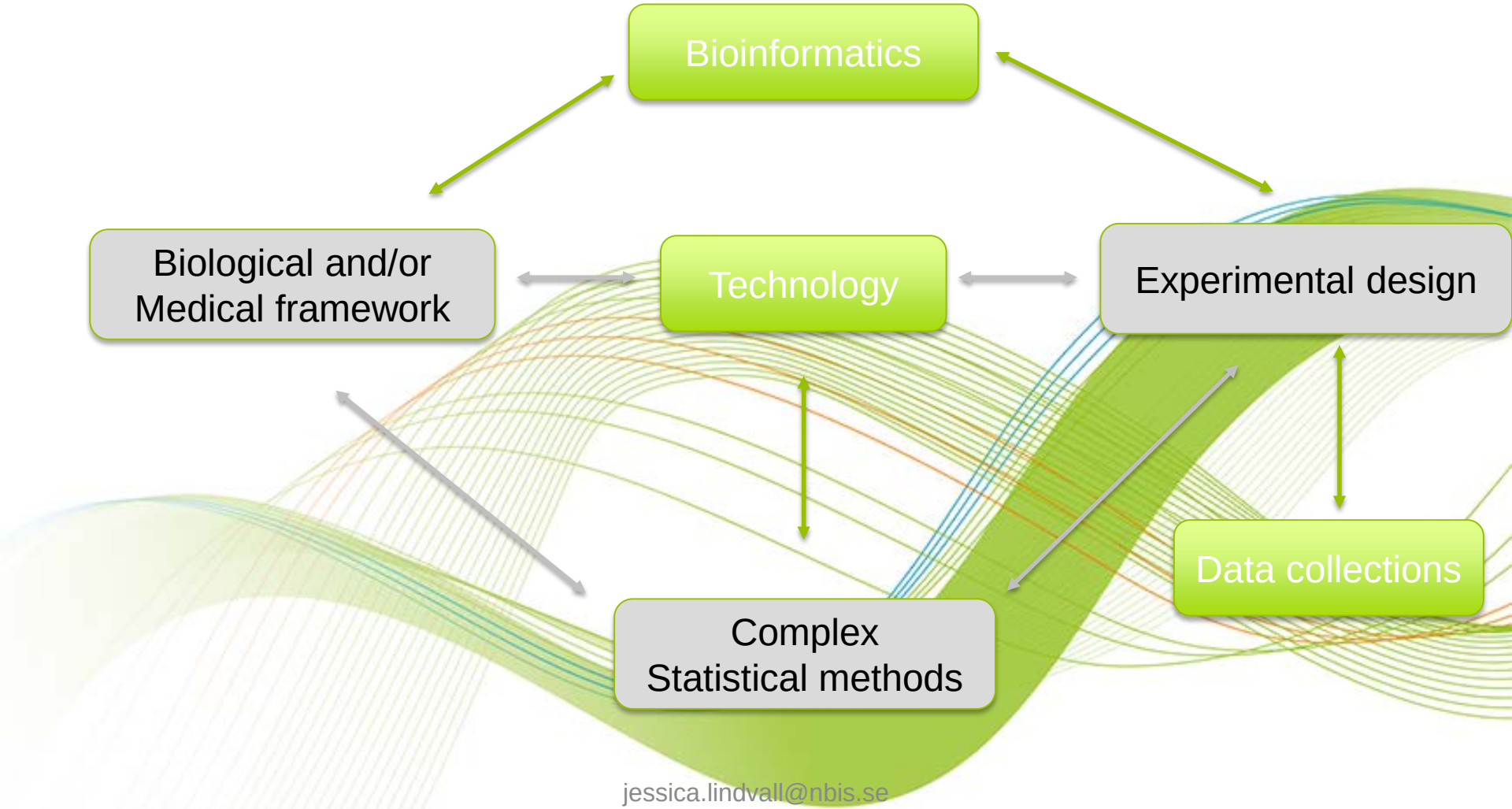


Microarray Data Analysis Workflow

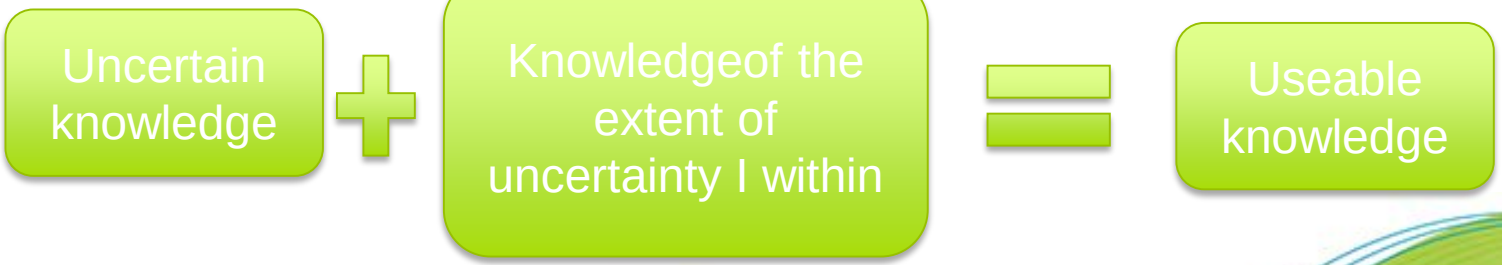


jessica.lindvall@nbis.se

It all goes together!



jessica.lindvall@nbis.se



Measurement Model

m= ≠+e

m – measurement model

≠ - True, however UNKNOWN value

What is the *meaning* of e?

What is the *variance* of e?

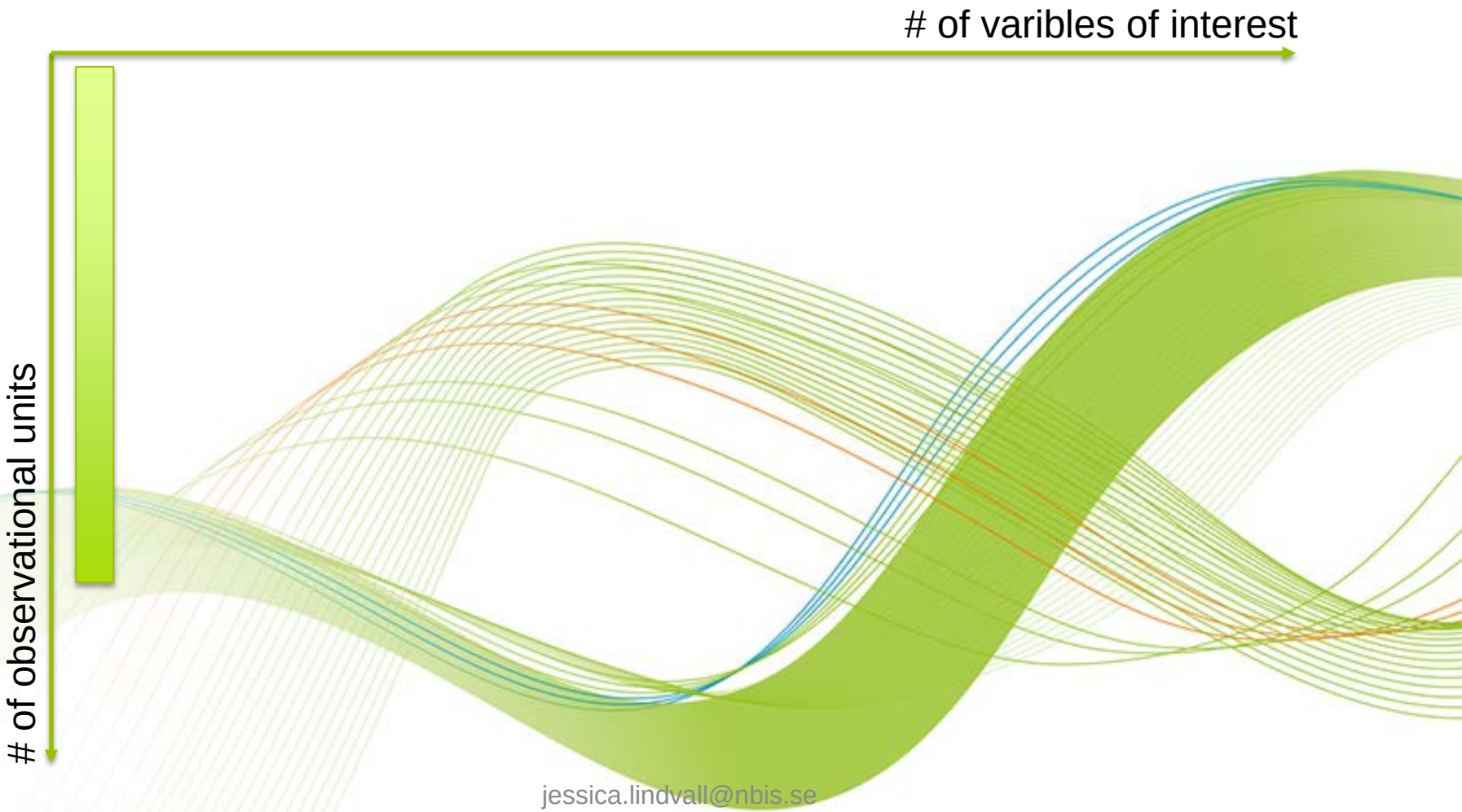
Dependence between e and ≠

Distribution of e and ≠

Decisions on the Experimental design influence the measurement model

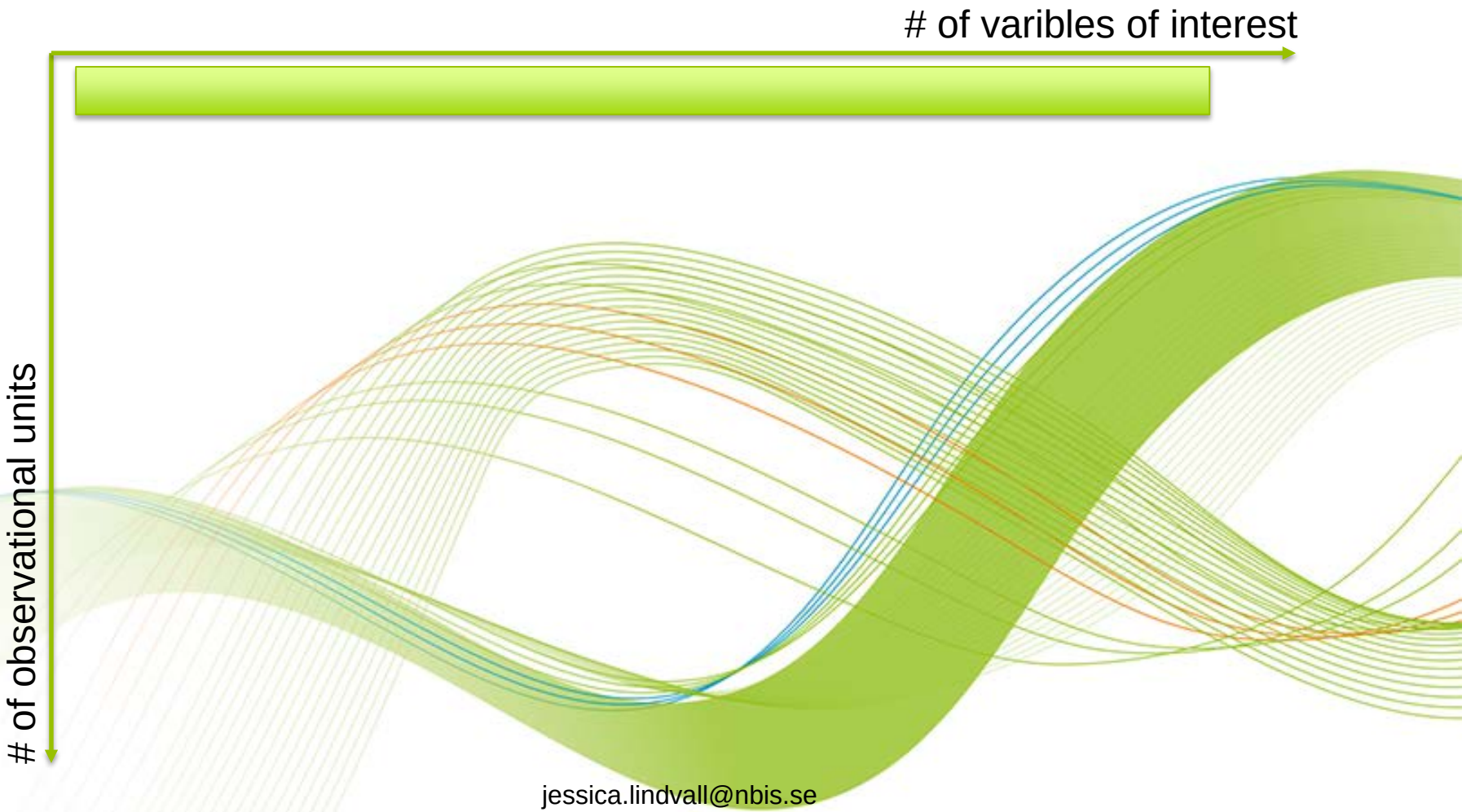
jessica.lindvall@nbis.se

Information dilemma – too many or too few?



jessica.lindvall@nbis.se

Information dilemma – too many or too few?



jessica.lindvall@nbis.se

The experimental set-up – a general take!

- Scientists deal mostly with experiments of the following form:
  - A number of alternative conditions / treatments,
  - one of which is applied to each experimental unit,
  - an observation (or several observations) then being made on each unit.
- The objective is:
  - Separate out differences between the conditions / treatments from the uncontrolled variation that is assumed to be present,
  - Take appropriate statistical steps towards understanding the phenomena under investigation.

Experimental design – From a Biologist perspective

- *know the pitfalls and assumptions* of particular design,
- be able to identify the type of *model appropriate for the sampling design and type of data* that will be collected,
  - Species
  - Tissue/Blood
  - Primary cells
  - Cell lines
- be able to biologically *interpret the output* of analyses using the statistical model,
- be able to *design experiments optimally*, i.e. with the best possible use of our *limited time and resources*.

jessica.lindvall@nbis.se

Experimental design – From a Statistician perspective

- Description and detection of the pattern MUST BE done in a rigours way,
  - Detect gradients in space and time -> models explaining the pattern,
  - High confidence in estimations of the parameters in the statistical model
- Hypothesis testing of the design and model of the experiment is crucial.

jessica.lindvall@nbis.se

Statisticians and Biologists – the common language

Statistician to the Biologist

- frame our questions in such a way as to get a sensible answer (statistical solidity),

Biologist to the Statistician

- be aware of biological considerations that may cause statistical problems
- understand the advice or analyses that we receive, and be able to translate that back into biology

jessica.lindvall@nbis.se

Main requirements for an experimental setup

Once the *conditions / treatments*, *experimental units*, and the *nature of the observations* have been fixed, the main requirements are:

- Experimental units receiving different treatments should differ in no systematic way from one another – Assumptions that certain sources of variation are absent or negligible should, as far as practical, be avoided;
- Random errors of estimation should be suitably small, and this should be achieved with as few experimental units as possible; ·
- The conclusions of the experiment should have a wide range of validity;
- The experiment should be simple in design and analysis;
- A proper statistical analysis of the results should be possible without making artificial assumptions.

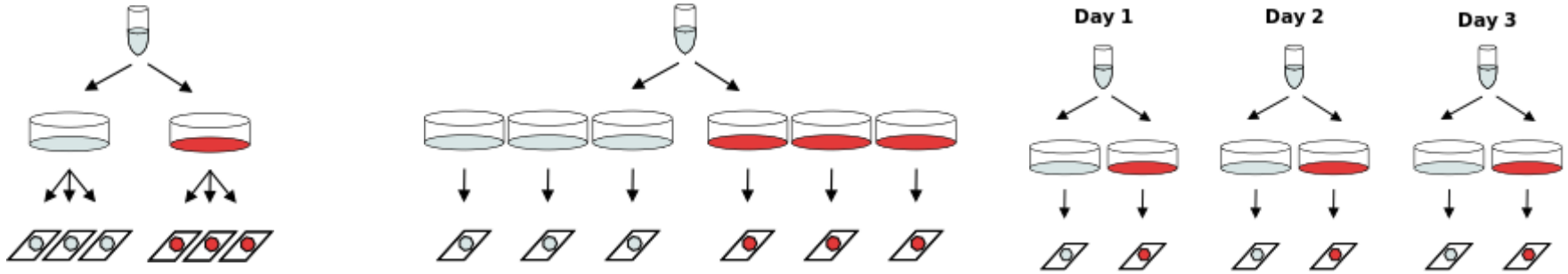
Taken from Cox DR (1958) Planning of experiments, Wiley & Sons, New York (page 13)

Raise awareness – what to take into consideration

- General thoughts – always applicable!
- 3 different examples – different (biological) ways of thinking!
  - Established (immortalized) cell lines
  - Human primary cells
  - Animal models e.g. mice or flies

jessica.lindvall@nbis.se

Aim for “true” Biological replicates



BAD **BETTER** **GOOD ENOUGH**

- The ***ideal*** design would have ***biological replicates (i.e. cells from multiple people or animals)***

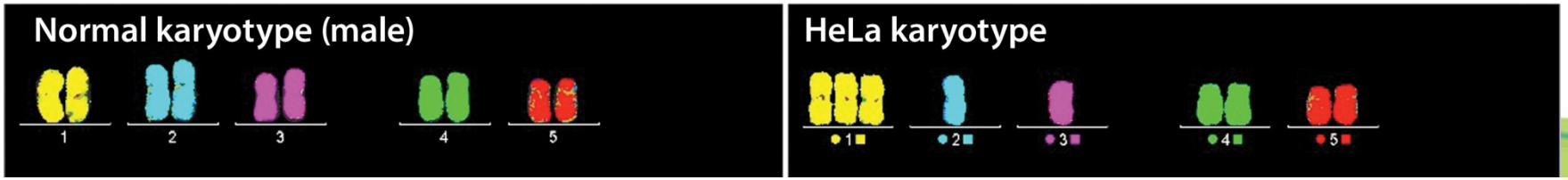
jessica.lindvall@nbis.se

Established cell-lines

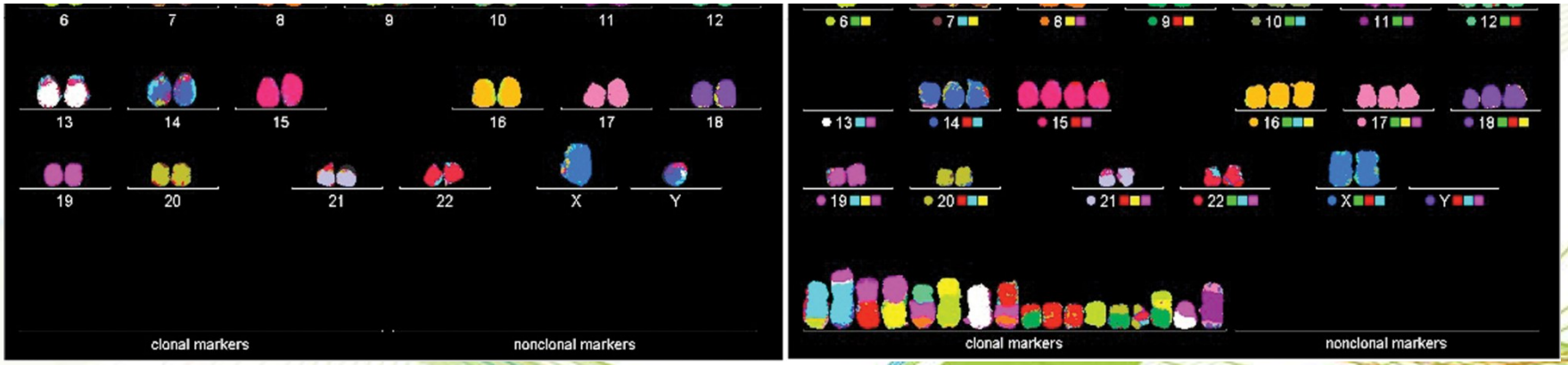
- HeLa cells
  - Transformed/immortalized cell line -> strongest clone “wins”!
  - Number of passages different from different labs -> strongest clone “wins”!
  - (Not the same “winning” clone in the different labs (or not even within the same lab maybe))

Established cell-lines

- HeLa cells



Hence, established cell-lines (generally) DO NOT reflect life



*Do a Karyotype test **before** your experiment to be aware of what you have!*

Human primary cells

- Potentially there is a mixture* of cells e.g. blood/PBMC
- Better if one can purify specific cell types e.g. B-cells from blood etc.
  - ** there are tools to handle this*
- “Cleaner” – no messy karyotypes- reflecting “true” biology
- Harder to collect (maybe) Need higher number of replicate samples per group – to mirror the true population

Animal models

- More controlled compared to Human primary cells/tissues
- Strain background *might* (IS) make a difference in outcome
- Less replicates are needed (due to more control) compared to human primary cells/tissues
- Number of animals might increase – Sufficient number of primary cells hard to obtain?!?

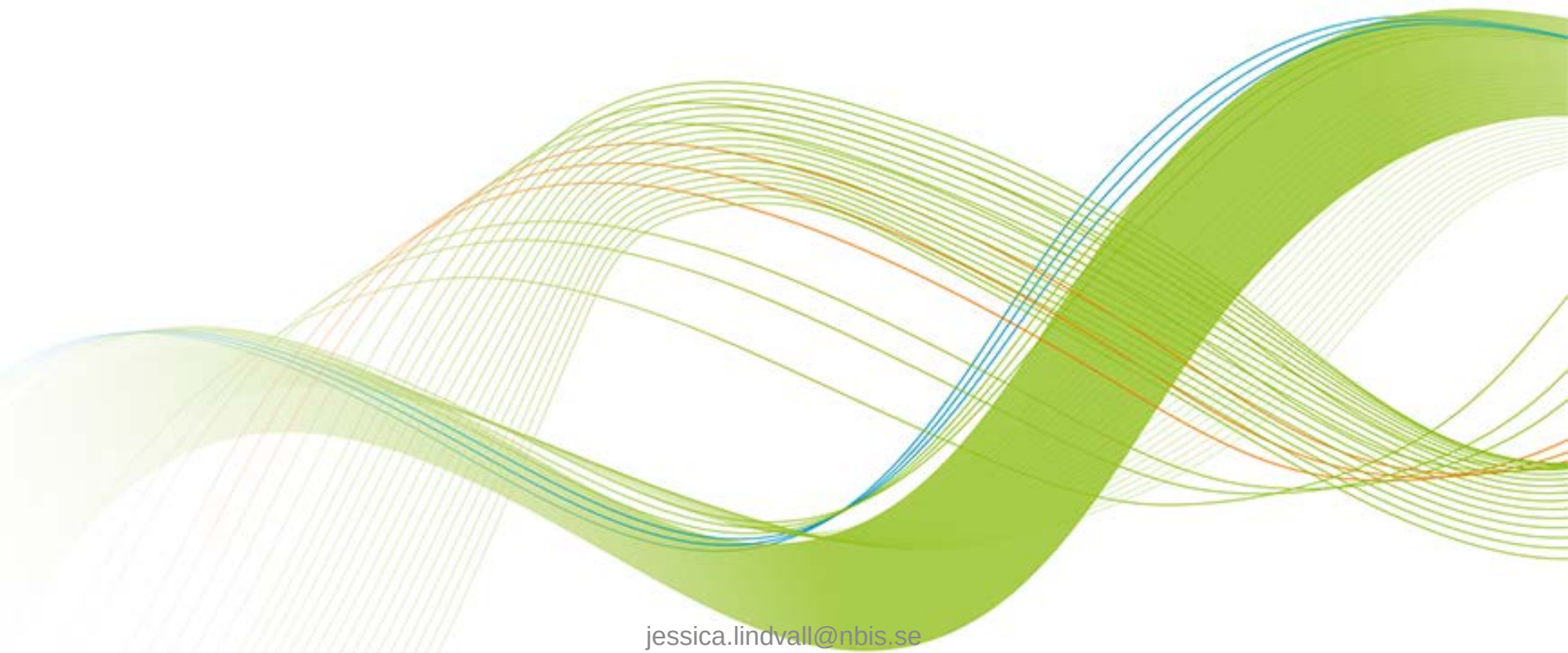
Biological data usually messy!

- Use statistics as a tool to clarify the biological question
- Design of the project important **BEFORE** starting the experiment
  - Starting material (factors influencing the data output)
    - Mixture of cell types
    - Transformed cell-lines (messy karyotypes)
    - Gender, Age…
    - Batch effects (experimental date, lab effects, technology)
  - Awareness of your data takes you a long way!

One last thing to keep in mind – The picture!

jessica.lindvall@nbis.se

Questions?



jessica.lindvall@nbis.se