

Experimental Setup and Design

Kjell Petersen

Computational Biology Unit



microarray.no

Biological generality

Replicates

Batch

Biological generality

Replicates

Batch

microarray.no

Overview

Compare the right samples for your hypothesis

Biological generality – aim for the right level

Factors influencing the outcome of the experiment

How to control for known and unknown factors

Time series experiments

Comparing the right things for your hypothesis

What is your hypothesis ?

Why are these groups interesting to investigate closer at the transcriptomics level ?

What do you expect to change in expression levels?
Particular systems related to the treatment/exposure/difference of the sample groups?

- Immune response ?
- Metabolic processes ?

Will need this input to decide **when** to sample and **what** to sample

There exists no such thing as a hypothesis free microarray experiment!

(How would you design it...?)

Comparing the right things

- Make sure there are no confounding effects
 - E.g. samples from patients vs control where all patients are old and all controls are young
- Biopsies should be taken from the same part of tissue
 - E.g. If one sample cuts through a major vein, differences seen could be due to different cell type compositions
- Choose the right model system
 - Do the cell types present express the genes of interest, e.g. not all treatments can be evaluated by blood samples, or in cell cultures

Biological generality – aim for the right level

The same hypothesis can be investigated at many different levels

Origin of cells

Comparing cell cultures started from cell lines

Comparing cell cultures started from different individuals

Comparing cells harvested directly from individuals

Replicates / reproducibility

Where to aim:

What is already known and published in your particular field?

Biological vs technical replicates

- Generally we use biological replicates to answer biological questions and technical replicates to answer technical questions
- What is a biological replicate?
 - Are cell cultures that originates from the same biological source biological or technical replicates?
- What biological generality of your results are you aiming for?
- What is your long term plan ? Except for getting things published...

How many replicates?

Do we have to pool RNA to get enough for hybridisation?

How many do you have ready now?

– Will you get more later?

What is the expected expression difference of targeted biology in these samples ?

Will “no change” be a desired significant result ?

Budget

- Prioritize some groups/contrasts, and run more biological replicates to improve generality of results?
- Time is money also in academia, remember analysis costs.

Factors influencing the outcome of the experiment

Avoid systematic errors

- Technical noise is added to the experiment every time you touch a sample at any step. It is important to control this, so that equal amounts of noise is added to all samples across the sample groups
- Common steps to control, include sampling, RNA extraction, labeling and hybridisation, but there may be others in your particular experiment
- Important to identify all steps that can add noise and that can be controlled.
- To control the steps you must also get to know your batches
 - Capacity per day
 - Persons involved
- The sample groups should generally balance across the batches
- Randomize the order of treatment within a batch

Other known factors to be careful about

Production of microarrays – use same batch

Sampling time for samples collected over time

Intrinsic properties of the samples

Sex

Age

Standardised treatment

Transcriptome in sync with respect to the hypothesis

How to order

Bad

A1
A2
A3
A4
A5
B1
B2
B3
B4
B5

Better

A1
B1
A2
B2
A3
B3
A4
B4
A5
B5

Best

A1
B4
A3
B2
A2
B3
A5
B1
A4
B5

General strategy

- For each step to control for systematic bias:
 - Distribute the biologic groups systematically in a balanced fashion: for two groups of equal size, every second from each group
 - Within each biological group, reshuffle the order of these samples
 - Divide it into roughly equal same size batches limited by your capacity for the step
- Tip: Colour code the sample names by biological group and the column next to it by batch

Experiment plan example

Biology	Sampling order	Extraction order
A1	A1	A4
A2	B4	B3
A3	A3	A5
A4	B2	B4
A5	A2	A1
B1	B3	B1
B2	A5	A3
B3	B1	B5
B4	A4	A2
B5	B5	B2

Exp plan with batches

Biology	Sampling order	Extraction order
A1	A1	A4
A2	B4	B3
A3	A3	A5
A4	B2	B4
A5	A2	A1
B1	B3	B1
B2	A5	A3
B3	B1	B5
B4	A4	A2
B5	B5	B2

microarray.no

Additional question

- Are the samples naturally paired ?
 - Before and after treatment
 - Two different samples from same individual
- Should then not be split in the design/experimental plan!

Keep them as pair to avoid added noise, since these samples are expected to be naturally in sync with each other (less noise/variance) compared to any other random pairing of samples.

Time series experiments

- Why consider doing a time series experiment:
 - Biology is dynamic
 - The genes can change their expression in different ways during time
 - Observe cascades and secondary effects

Experimental design (Ex.d)

- Often larger and complex
- What to consider in an experimental design of a time series study:
 - Time points
 - Distance between time points
 - Replicates
 - Control

Ex.d.: Time points and the distance between them

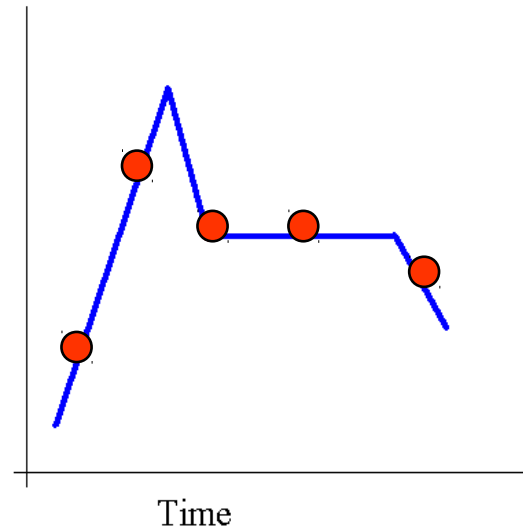
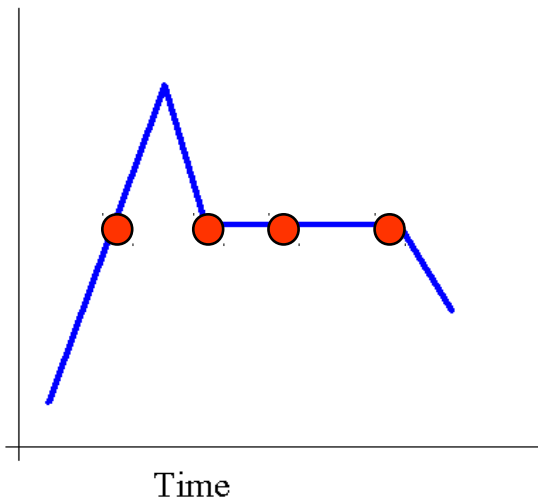
- How to choose:
 - What is already known about the system/genes
 - Theoretical and experimental knowledge
 - Early /late response
 - The direction of the response

Will changes in some genes influence others

- What is your field of interest

Ex.d.: Replicates

- How many replicates to use:
 - Many time points but few replicates



- Should be enough to do the analysis you want!

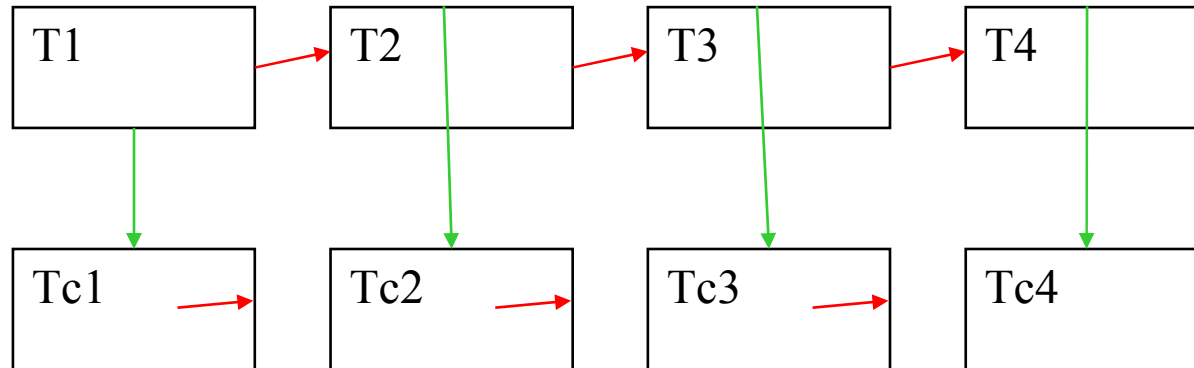
Ex.d.: Control

- What to use as a control and how many controls?
 - One control point or one time series control
 - Differences between treatments
 - The other time points as a control

Ex.d.: Control

- One control
 - Use one time point or a untreated sample as control
 - Ok if no other time effects are expected, such as growth, phase of expression
- A untreated time series as control
 - Same time points, technology, sample source and same handling except the treatment
 - Find genes that change due to treatment over time, filter out some of the effects due to handling and time

Ex.d.: Control



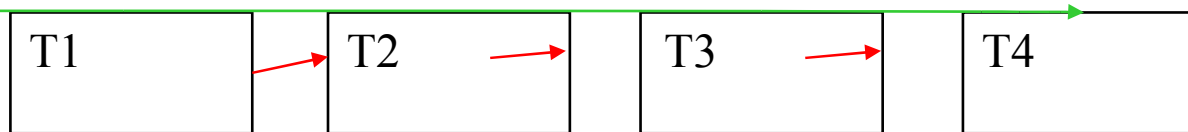
- **Vertically**: Differences due treatment, but you reduce the effect of the handling
- **Horizontally**: Differences due to treatment, time and some handling effect

Ex.d.: Control

- Differences between treatments - compare two or more time series
- Have the same:
 - Time points
 - Conditions
 - Sample source
 - Technology
 - Handling

Ex.d.: Control

- The other time points as a control:
 - Changes due to treatment, time but it does not exclude the effect of the handling
 - Less arrays used
 - See **short** time effects and **long** time effects'



- Ok if no other time effects are expected, such as growth, phase of expression

Additional slides

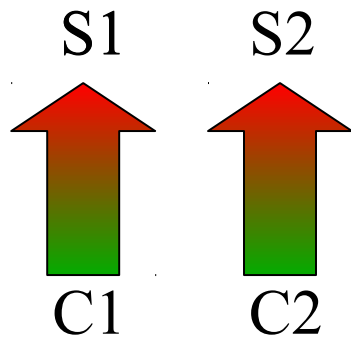
Experimental Setup and Design

- Biological aspects
 - What is your hypothesis or question?
 - What else is known beforehand on the topic ?
- Technical aspects
 - Compare the right things to each other
 - Choice of platform
 - Avoid systematic errors
 - Ideal: each step, one person, one protocol, one day
 - Plan biological replicates
 - statistical significance of findings
 - choose results to validate
- What is the budget
 - Experimental and analysis costs

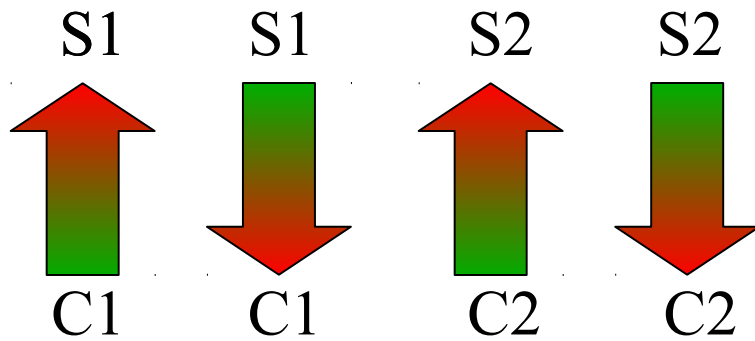
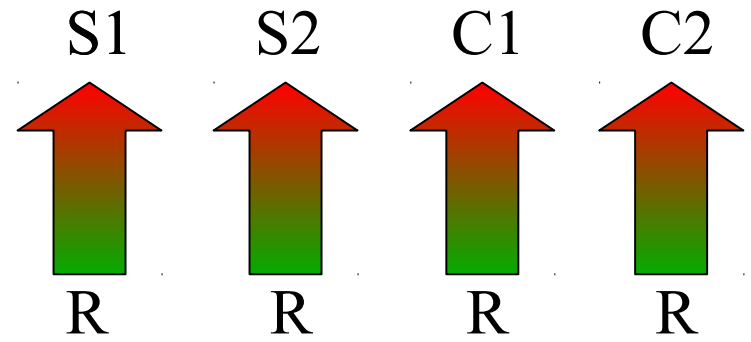
Additional questions for two channel platforms

- Are the samples naturally paired ?
 - Before and after treatment
- Direct comparison vs indirect ?
- Dye swaps ?
- Common reference design
 - Representable, independent RNA
 - Dye swaps strongly recommended
 - More noisy
 - More flexible

Direct



Indirect



Direct w/dye swap

