A 2x2 factorial microarray experiment

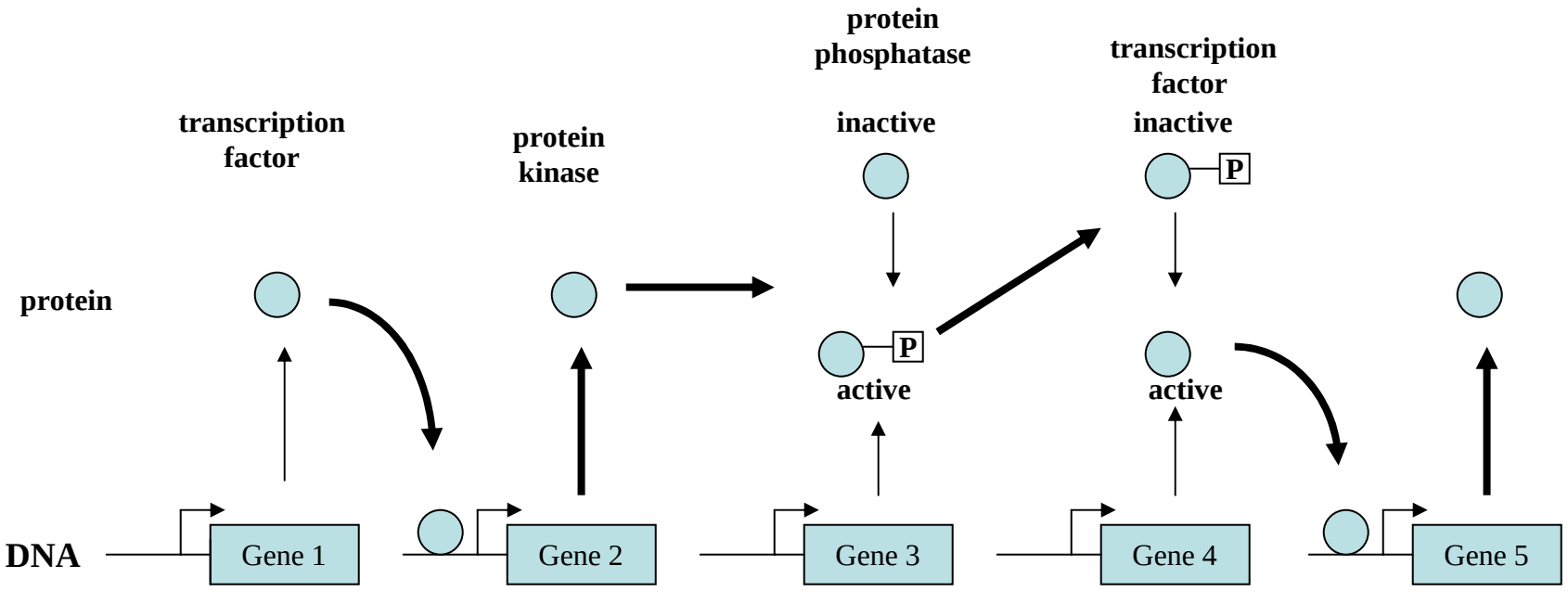
Robert Gentleman, Denise Scholtens

Arden Miller, Sandrine Dudoit

Complexity of genomic data

- The functioning of cells is a complex and highly structured process.
- In the next slide we show a stylized biochemical pathway (adapted from Wagner, 2001).
- There are transcription factors, protein kinase and protein phosphatase reactions.
- Tools are being developed that allow us to explore this functioning in a multitude of different ways.

An example of the interactions between some genes (adapted from Wagner 2001)



Overview

- Wagner (2001) suggests that the holy grail of functional genomics is the reconstruction of genetic networks.
- In this tutorial we examine some methods for doing this in factorial genome wide RNA expression experiments.
- Such experiments are easy to carry out and are becoming widespread. Tools for analyzing them are badly needed.

Gene effects

- A factor can either inhibit or enhance the production of mRNA for any gene.

- The inhibition or enhancement of mRNA production for any given gene can affect transcription for other genes either through inhibition or enhancement.

Targets

- We define a target of a factor to be a gene whose expression of mRNA is altered by the presence of the factor.
- A primary target is a target that is directly affected by the factor.
- A secondary target is a target whose transcription is altered only via the effects of some other genes, i.e., can be traced back to one or more primary targets.

CX experiment

- There are two factors
  - Estrogen, E: known to affect transcription of various genes (some known, some unknown).
  - Cyclohexamide, CX: known to stop all translation (with very few exceptions).
- The design is a classical 2x2 factorial design, with two replicates.
- We are interested in the main effects and interactions for E and CX.

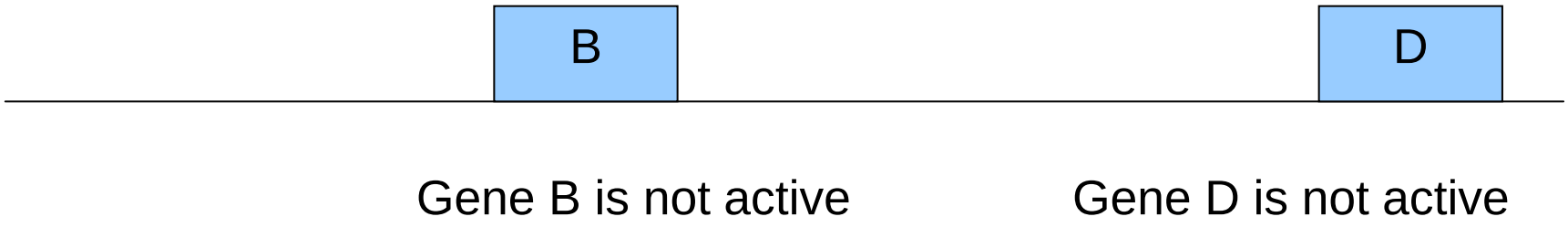
CX experiment

- We identify as targets all genes whose expression of mRNA is affected by the application of E.
- A target can be either primary or secondary
  - primary if E directly affects expression of mRNA.
  - secondary if mRNA production is affected by some other gene and can be traced back to a primary target.

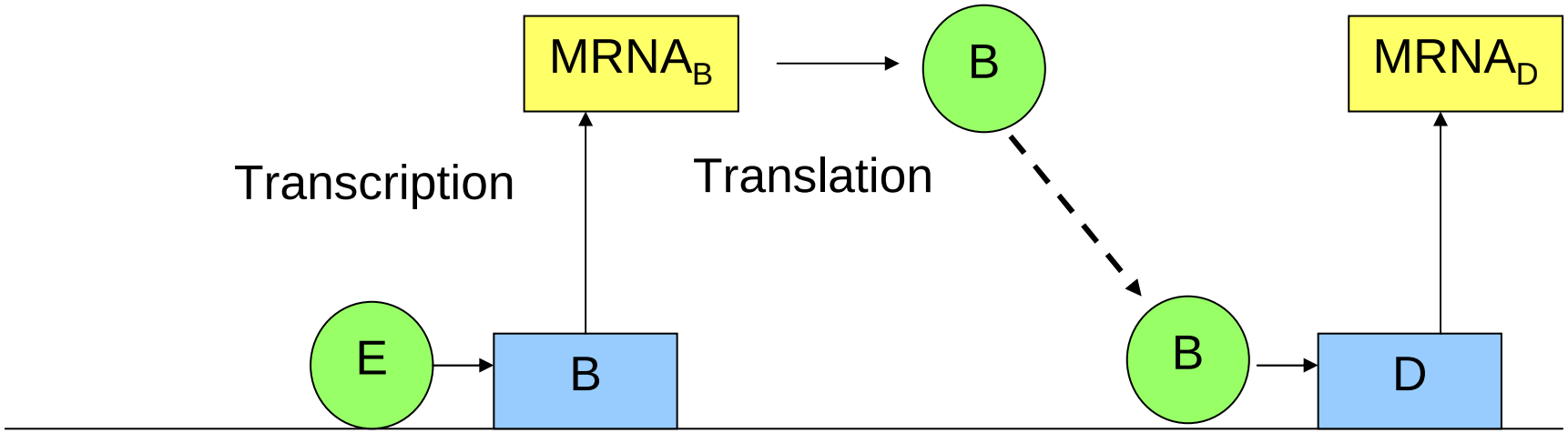
Scenario 1

- Assume that there are two related genes, B and D, where
  - B is a primary target of E,
  - D is a secondary target only via B.
- Neither is expressed initially.
- E causes B to be expressed and this in turn causes D to be expressed.
- The addition of CX by itself may not affect expression of either B or D.

No factors applied



E only



B is a Primary Target of E

Production of mRNA$_B$ is enhanced by E

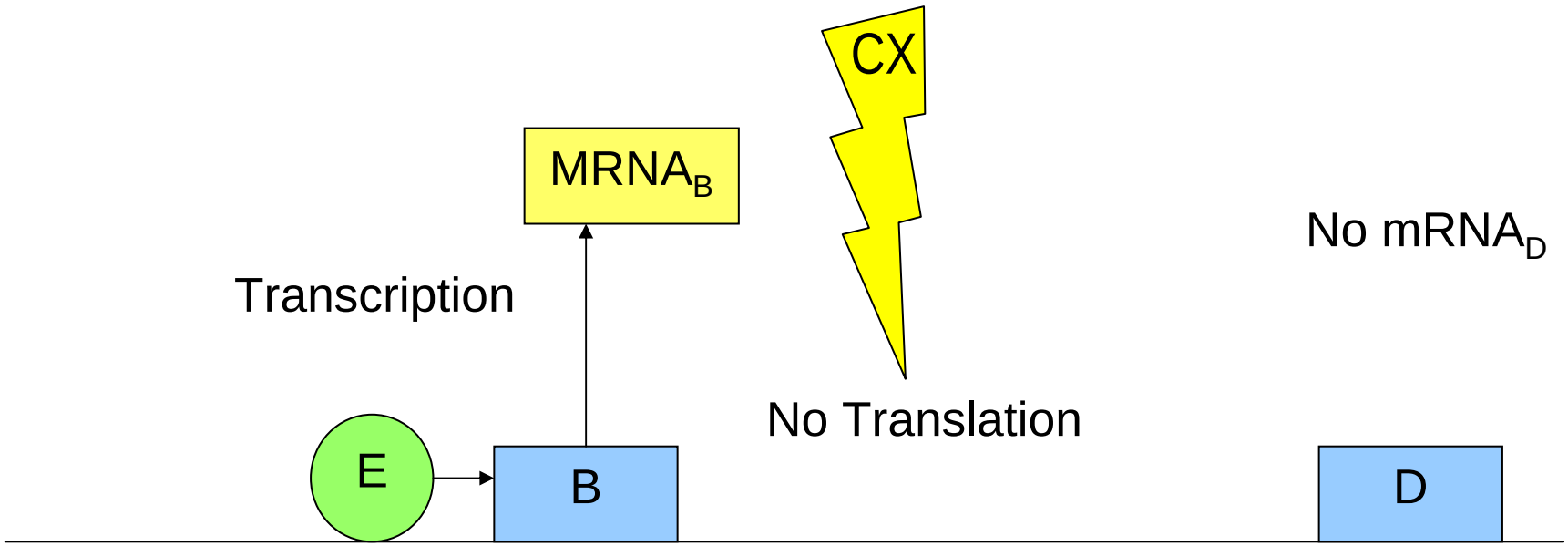
D is a Secondary Target of E

Production of mRNA$_D$ is enhanced by B

Scenario 1

- In the presence of both CX and E we see increased expression of $mRNA_B$ but not of $mRNA_D$.
- CX stops translation of B and hence transcription of D.
- This will be one of the principles we can use to differentiate between primary targets of E (such as B) and secondary targets of E (such as D).

E and CX both present



B is a Primary Target

Production of mRNA$_B$ is enhanced by E

Production of mRNA$_D$ is decreased (prevented)

Interpretation: Scenario 1

	$mRNA_B$	$mRNA_D$
Nothing	Low	Low
E	High	High
CX	Low(?)	Low (?)
E and CX	High	Low

Scenario 1

- Note that while we show a direct relationship between the expression of B and of D we cannot detect such a relationship from these data (the purpose of this scenario is purely pedagogical).
- Other scenarios include
  - Suppression of D by B, enhancement of B by E.
  - Enhancement of D by B, and suppression of B by E.

CX experiment

- Assume the following linear model for the observed expression response (possibly on transformed data) of any given gene

$$y_{ig} = \mu_g + \beta_{Eg} x_{1i} + \beta_{CXg} x_{2i} + \beta_{E:CX,g} x_{1i} x_{2i} + \varepsilon_{ig}$$

- *i* indexes chips and *g* indexes genes.
- x_1 indicates the presence of E and x_2 indicates the presence of CX.

Inference

- The 2x2 CX microarray experiment measures the expression response of each gene under each of the four factor combinations.

- But there is a difference, B is a primary target of E, while D is a secondary target of E.

Inference

- If gene X is any target for E, the level of $mRNA_X$ might not change when E is added.
- $mRNA_X$ might already be being made as fast as possible, so addition of E has no effect.
- Production of $mRNA_X$ might already be suppressed by some other compound.
- A true baseline would help in resolving these situations.

Inference

- The introduction of CX provides a form of baseline.

- Since (among other things) CX halts translation we should be able to use the presence or absence of CX to find out about primary versus secondary targets.

Inference

- For any gene we can interpret the coefficients in the linear model as follows.
- The parameter β_E can be interpreted as the main effect of E.
- Genes for which β_E is different from zero are potential targets.
- As noted previously, not all targets will have β_E different from zero.

Inference

- The parameter β_{CX} can be interpreted as the main effect of CX.
- If β_{CX} is different from zero, this suggests that production of mRNA is translationally regulated.
- The interpretation of the interaction $\beta_{E:CX}$ is more difficult.

Primary targets

- Consider the case where we have only CX and CX+E.
- Since CX halts all translation, then any differences between the condition where CX alone is present and CX+E is present should indicate primary targets of E.
- This is equivalent to testing the hypothesis

H_0: $\mu+\beta_E+\beta_{CX}+\beta_{E:CX} = \mu+\beta_{CX}$, i.e.,

H_0: $\beta_E+\beta_{E:CX} = 0$

Primary targets

- Genes for which the hypothesis

  H_0: $\mu+\beta_E+\beta_{CX}+\beta_{E:CX} = \mu+\beta_{CX}$

  is rejected are candidates for primary targets.
- Those with β_E different from zero, but for which we do not reject H_0,are secondary targets.
- It seems likely that some inference may be drawn from the relationship between β_E and $\beta_{E:CX}$, their signs and their significance levels.

Scenario 1

	Primary	Secondary
β_E	> 0	> 0
β_{Cx}	= 0	= 0
$\beta_{E:CX}$	= 0	- β_E

Limitations

- While we may identify genes that are potentially primary targets and those that are potentially secondary targets we cannot identify gene—gene interactions, or feedback loops.
- We can observe the effects but not attribute them.
- The use of relevant metadata, biological and publication, seems pertinent and could help resolve some of the interactions.