Estimating causal effects from observational data

Marloes Maathuis
ETH Zurich

Causal questions

- Causal questions are about the mechanism behind the data or about predictions after some outside intervention

Causal questions

- Causal questions are about the mechanism behind the data or about predictions after some outside intervention

- Example questions about the mechanism behind the data:
    - Does smoking cause lung cancer?
    - What are major causes of global warming?
    - What is the gene regulatory network of yeast?

Causal questions

- Causal questions are about the mechanism behind the data or about predictions after some outside intervention

- Example questions about the mechanism behind the data:
  - Does smoking cause lung cancer?
  - What are major causes of global warming?
  - What is the gene regulatory network of yeast?

- Examples for predictions in changed systems:
  - How is the stock market going to react to some new policy interventions?
  - What is the average value of a phenotype after certain gene knock-outs?
  - What are predicted sales after a new advertising campaign?

Randomized controlled experiments

- Causal questions are best answered by randomized controlled experiments:
  - Groups are equal except for the treatment conditions
    $\Rightarrow$ any difference in outcome must be caused by the treatment
  - Example: clinical trials to test new drugs

Randomized controlled experiments

- Causal questions are best answered by randomized controlled experiments:
  - Groups are equal except for the treatment conditions
    $\Rightarrow$ any difference in outcome must be caused by the treatment
  - Example: clinical trials to test new drugs

- But sometimes such experiments are impossible, as they may be:
  - infeasible (global warming, smoking)
  - unethical (smoking)
  - expensive / time consuming (gene knock-outs)

Research question

- Can we learn causal effects from observational data in high-dimensional systems?

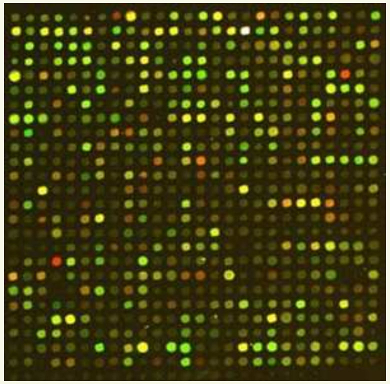
Research question

- Can we learn causal effects from observational data in high-dimensional systems?

- Example: gene regulatory network of yeast:
  - identify pairs of genes between which there is a large effect

  from observational data
  - gene expression levels of wild-type yeast

  with many more variables than observations
  - > 5000 genes
  - 63 yeast organisms

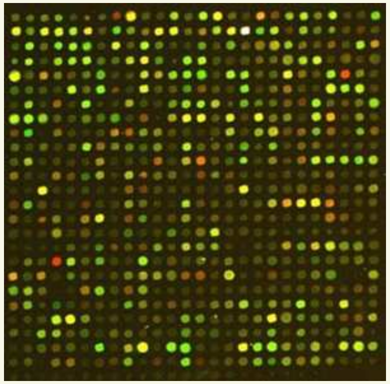


Research question

- Can we learn causal effects from observational data in high-dimensional systems?
- Example: gene regulatory network of yeast:
  - identify pairs of genes between which there is a large effect

  from observational data
  - gene expression levels of wild-type yeast

  with many more variables than observations
  - > 5000 genes
  - 63 yeast organisms



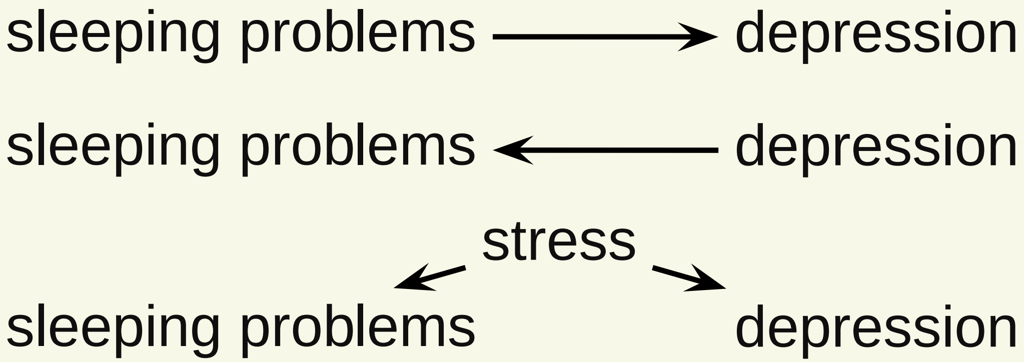
- Focus on developing scalable algorithms with proven statistical properties and validations on real data

Estimating causal effects from observational data

- It is impossible to estimate causal effects from observational data without making additional assumptions

Estimating causal effects from observational data

- It is impossible to estimate causal effects from observational data without making additional assumptions
- Common approach (e.g., Pearl, 2000; Robins et al, 2000):
  - assume that causal relations are known qualitatively and can be represented by a directed acyclic graph (DAG)



sleeping problems → depression

sleeping problems ← depression

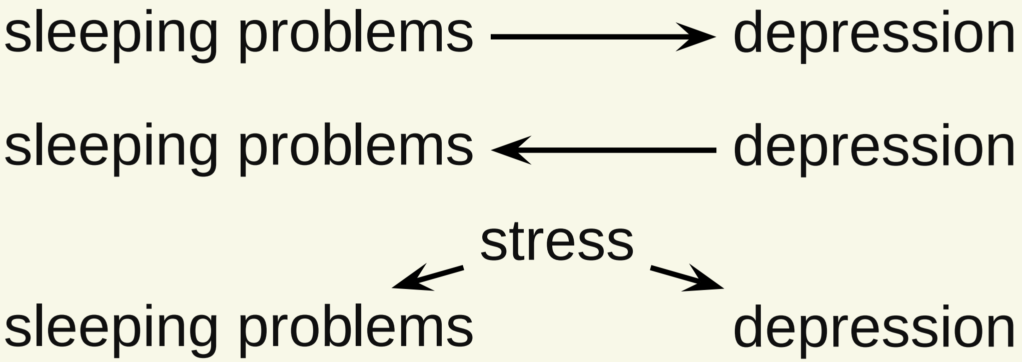
stress

sleeping problems depression

Estimating causal effects from observational data

- It is impossible to estimate causal effects from observational data without making additional assumptions

- Common approach (e.g., Pearl, 2000; Robins et al, 2000):
  - assume that causal relations are known qualitatively and can be represented by a directed acyclic graph (DAG)



sleeping problems → depression

sleeping problems ← depression

stress

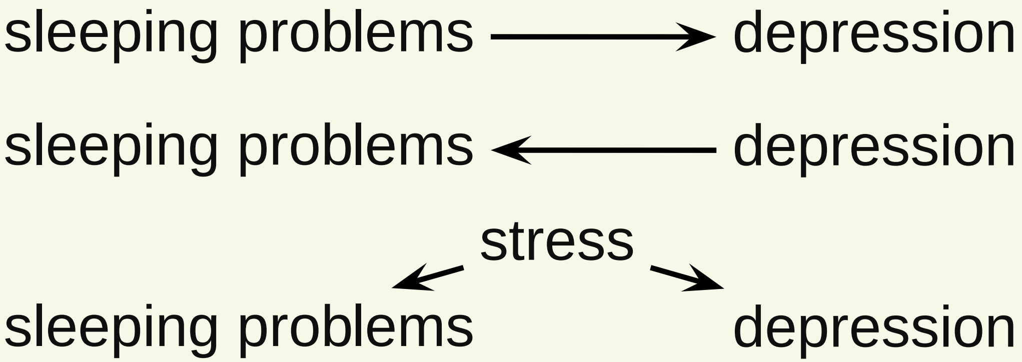
sleeping problems depression

  - then the sizes of the causal effects can be estimated from observational data (e.g., covariate adjustment)

Estimating causal effects from observational data

- It is impossible to estimate causal effects from observational data without making additional assumptions

- Common approach (e.g., Pearl, 2000; Robins et al, 2000):
    - assume that causal relations are known qualitatively and can be represented by a directed acyclic graph (DAG)



sleeping problems → depression

sleeping problems ← depression

stress

sleeping problems depression

-
    - then the sizes of the causal effects can be estimated from observational data (e.g., covariate adjustment)

- But knowing the graph structure is unrealistic in high-dimensional settings...

What can we do when the DAG is unknown?

- Assume the data come from an unknown DAG

What can we do when the DAG is unknown?

- Assume the data come from an unknown DAG
- A DAG encodes conditional independence relationships. Example: $X_1 \to X_2 \to X_3$ implies $X_1 \perp\!\!\!\perp X_3 | X_2$.

What can we do when the DAG is unknown?

- Assume the data come from an unknown DAG
- A DAG encodes conditional independence relationships. Example: $X_1 \to X_2 \to X_3$ implies $X_1 \perp\!\!\!\perp X_3 | X_2$.
- So given all conditional independence relationships in the observational distribution, can we infer the DAG?

What can we do when the DAG is unknown?

- Almost... several DAGs can encode the same conditional independence relationships. They are Markov equivalent.

What can we do when the DAG is unknown?

- Almost... several DAGs can encode the same conditional independence relationships. They are Markov equivalent.
- Example:

	$X_1 \perp\!\!\!\perp X_3$	$X_1 \perp\!\!\!\perp X_3 \vert X_2$
$X_1 \longrightarrow X_2 \longrightarrow X_3$	F	T
$X_1 \longleftarrow X_2 \longleftarrow X_3$	F	T
$X_1 \longleftarrow X_2 \longrightarrow X_3$	F	T
$X_1 \longrightarrow X_2 \longleftarrow X_3$	T	F

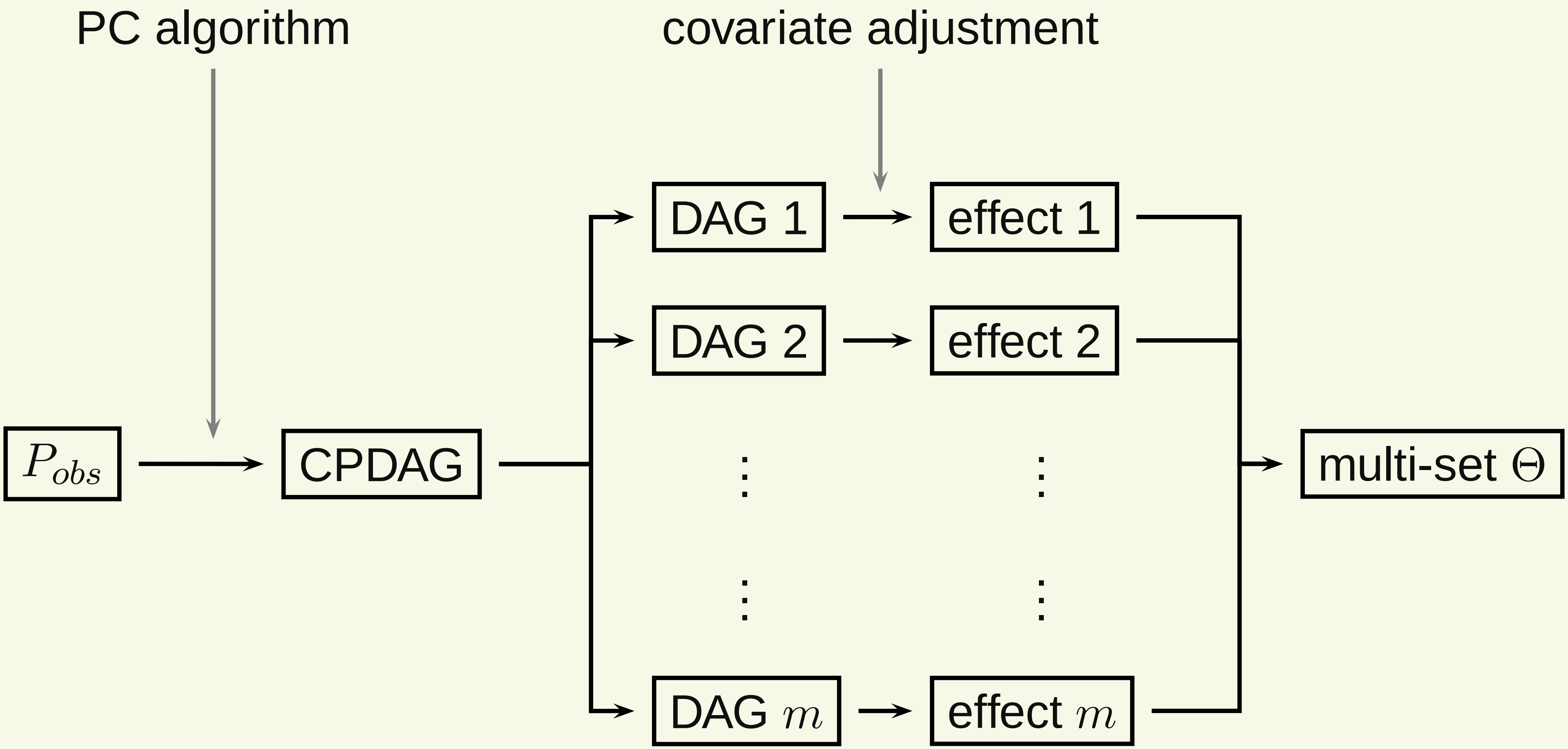
What can we do when the DAG is unknown?

- Almost... several DAGs can encode the same conditional independence relationships. They are Markov equivalent.
- Example:

	$X_1 \perp\!\!\!\perp X_3$	$X_1 \perp\!\!\!\perp X_3 \mid X_2$
$X_1 \longrightarrow X_2 \longrightarrow X_3$	F	T
$X_1 \longleftarrow X_2 \longleftarrow X_3$	F	T
$X_1 \longleftarrow X_2 \longrightarrow X_3$	F	T
$X_1 \longrightarrow X_2 \longleftarrow X_3$	T	F

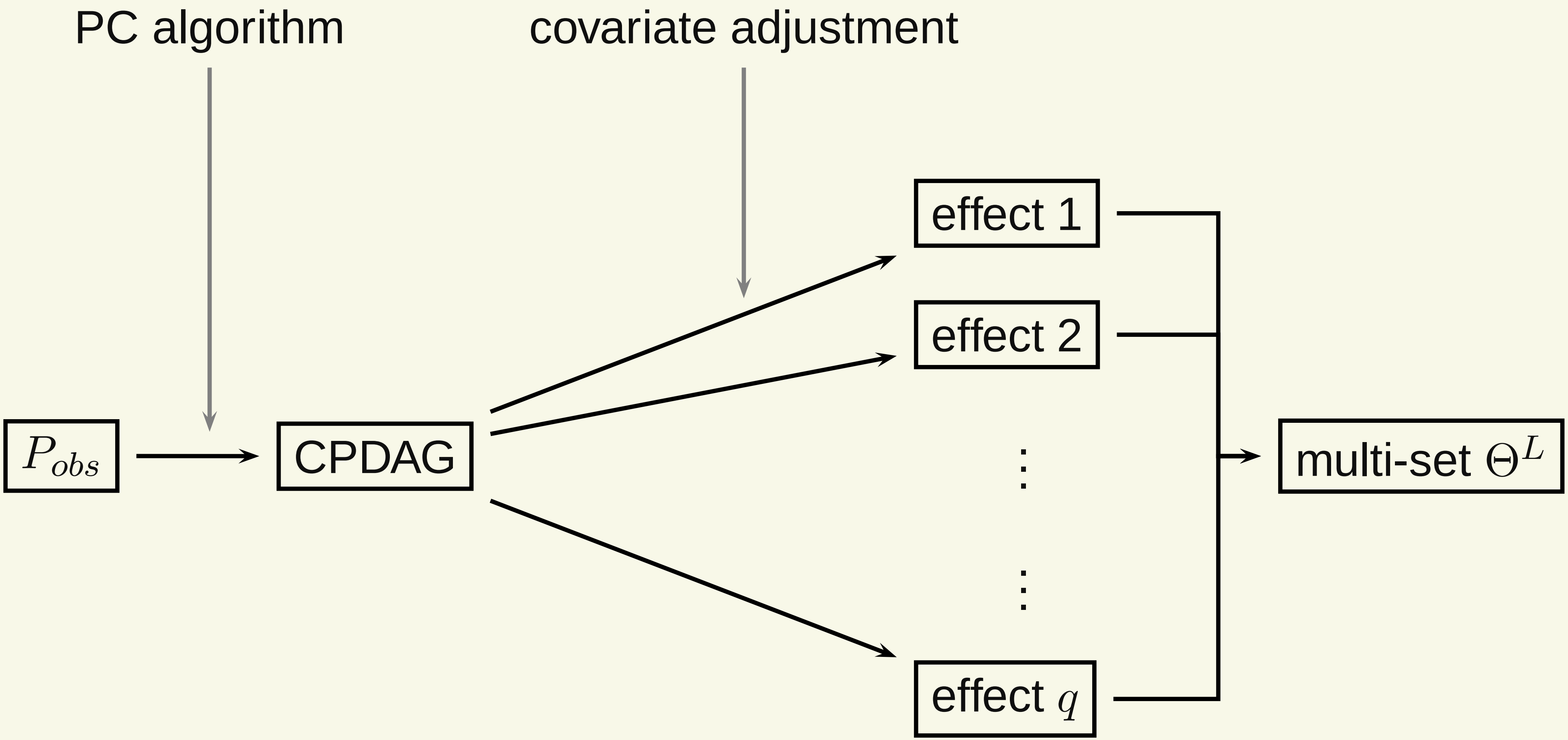
- A Markov equivalence class of graphs can be uniquely represented by a CPDAG. These can be learned by, e.g., the PC algorithm (Spirtes et al, 2000)

IDA algorithm: oracle version



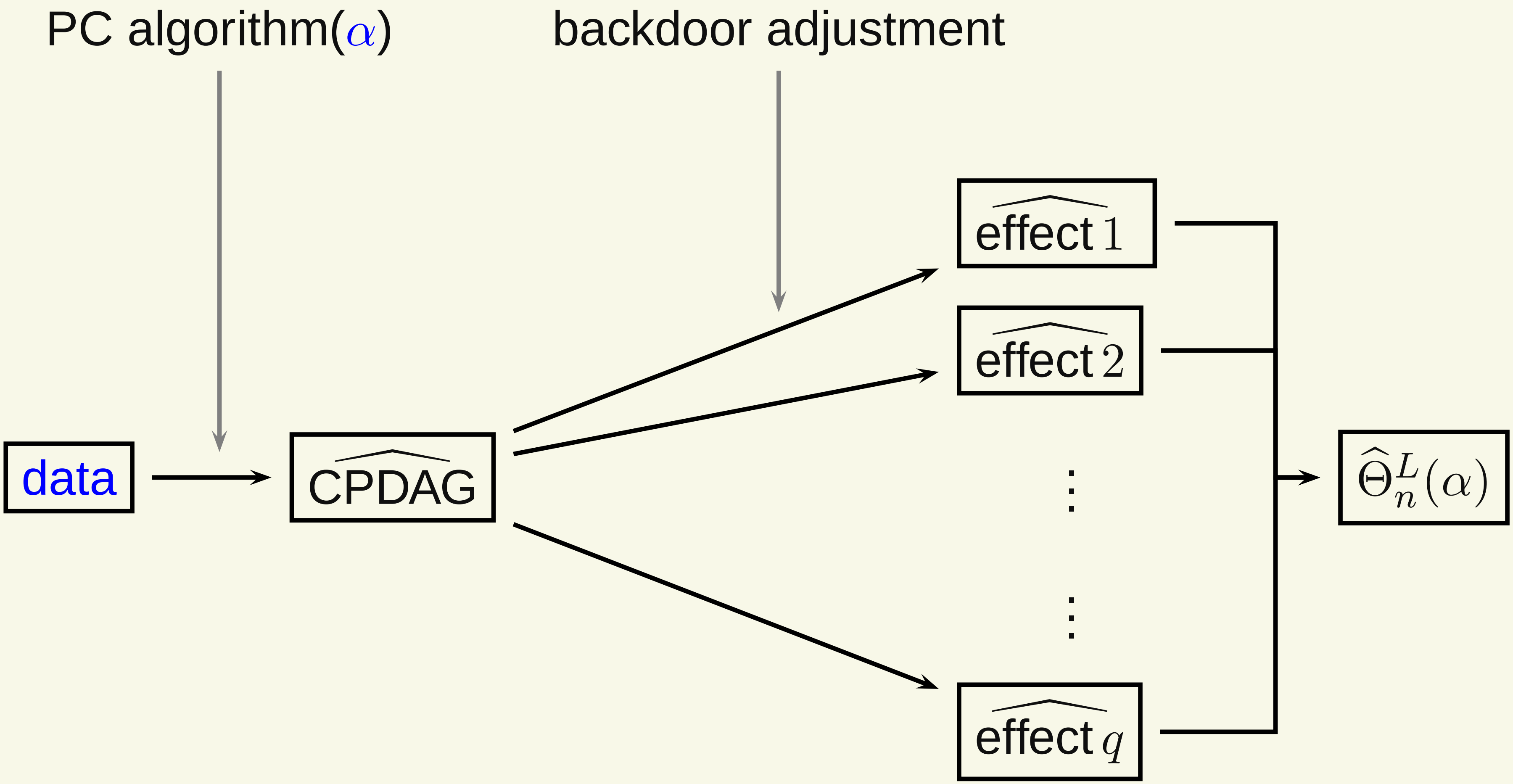
The true causal effect is in Θ.
We can obtain bounds on the size of the causal effect.

IDA algorithm: local oracle version



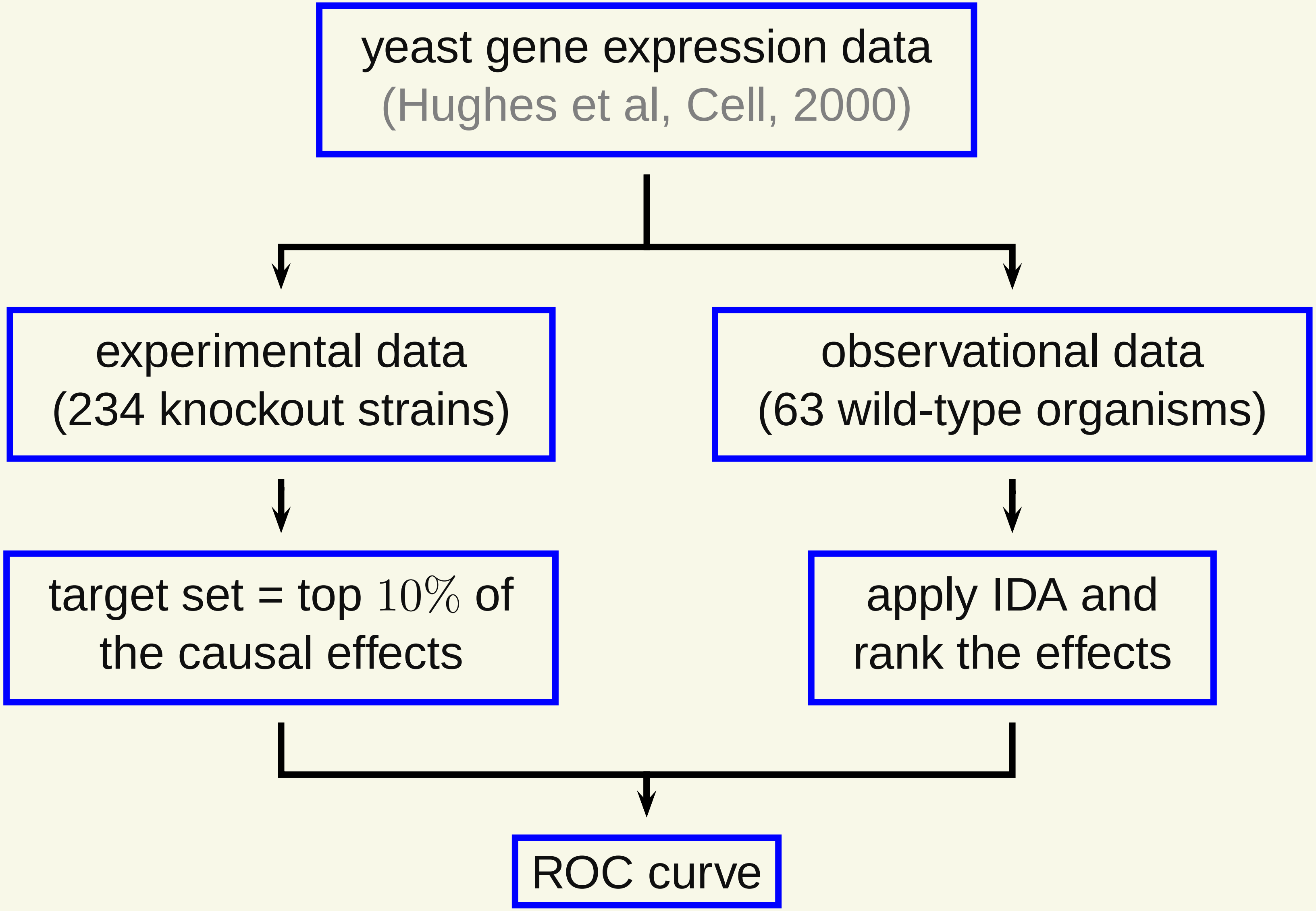
Bounds based on Θ^L are identical to bounds based on Θ.
Proof uses graph theoretic properties of the CPDAG.

IDA algorithm: local sample version


PC algorithm(α)
backdoor adjustment
data
$\widehat{\text{CPDAG}}$
$\widehat{\text{effect } 1}$
$\widehat{\text{effect } 2}$
$\vdots$
$\vdots$
$\widehat{\text{effect } q}$
$\widehat{\Theta}_n^L(\alpha)$

The estimates are consistent in certain sparse high-dimensional settings

Validation: overview



yeast gene expression data
(Hughes et al, Cell, 2000)

experimental data
(234 knockout strains)

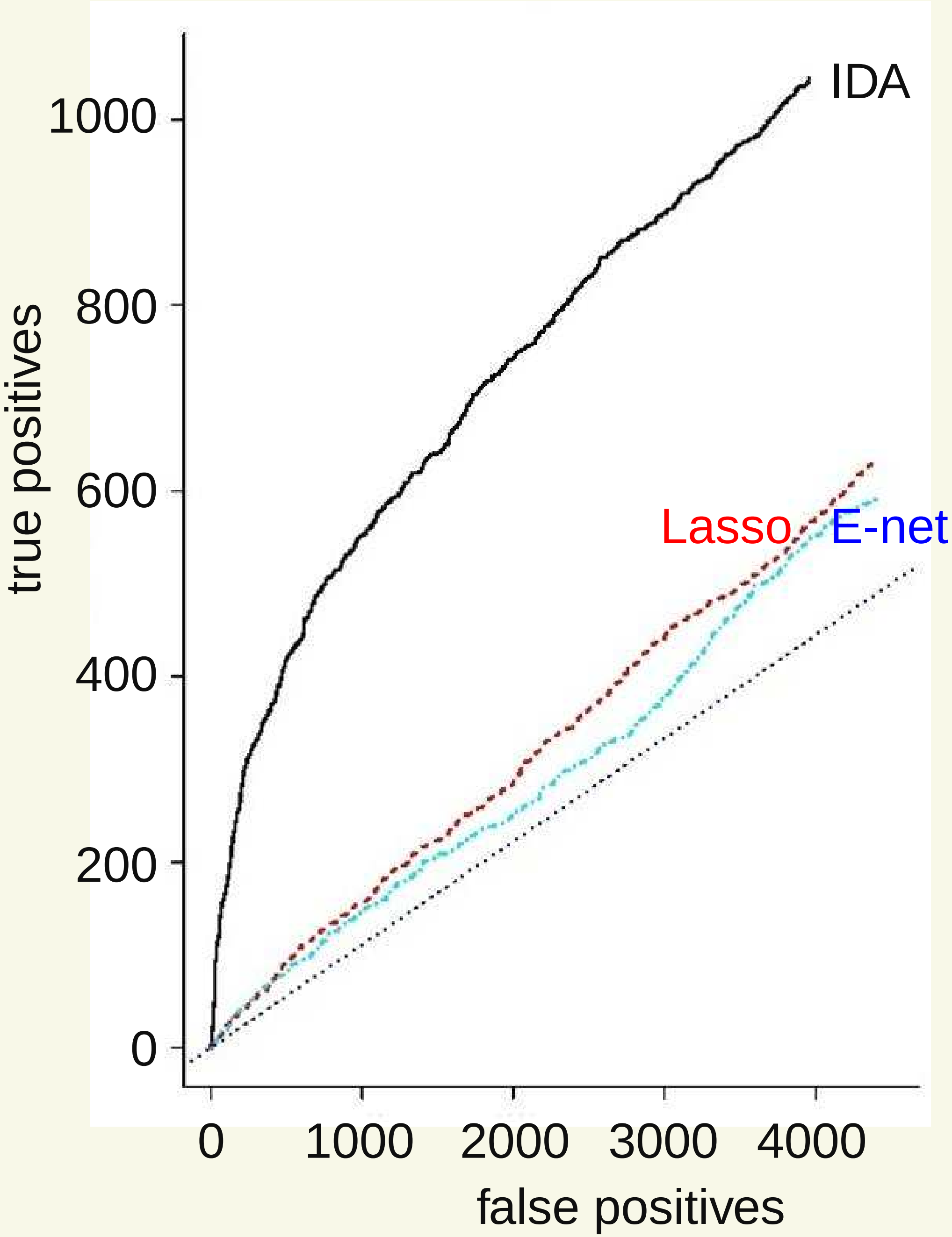
observational data
(63 wild-type organisms)

target set = top 10% of
the causal effects

apply IDA and
rank the effects

ROC curve

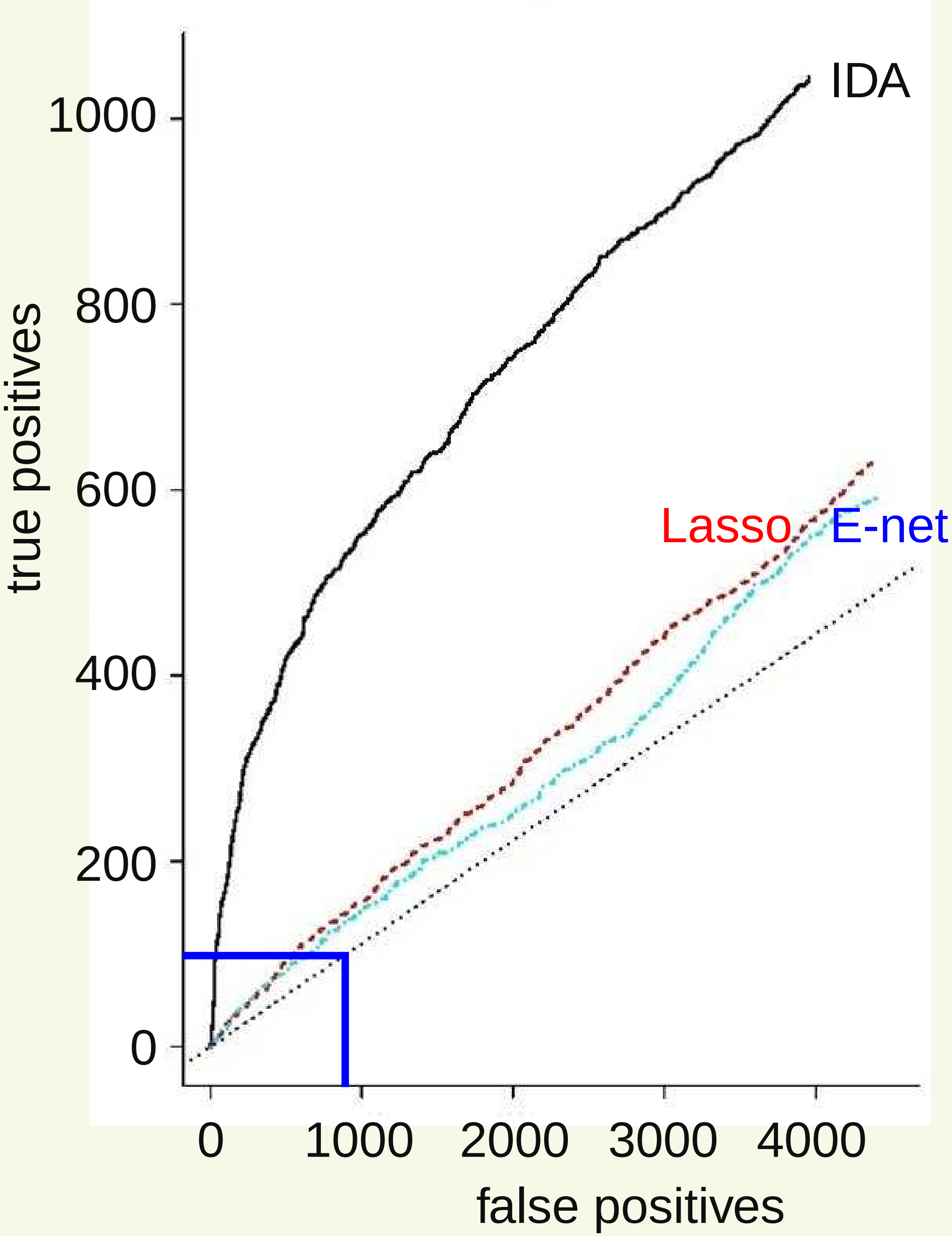
Validation of IDA on yeast gene expression data



Target set: top 10% of effects from experimental data

Source: Nature Methods, 2010

Validation of IDA on yeast gene expression data



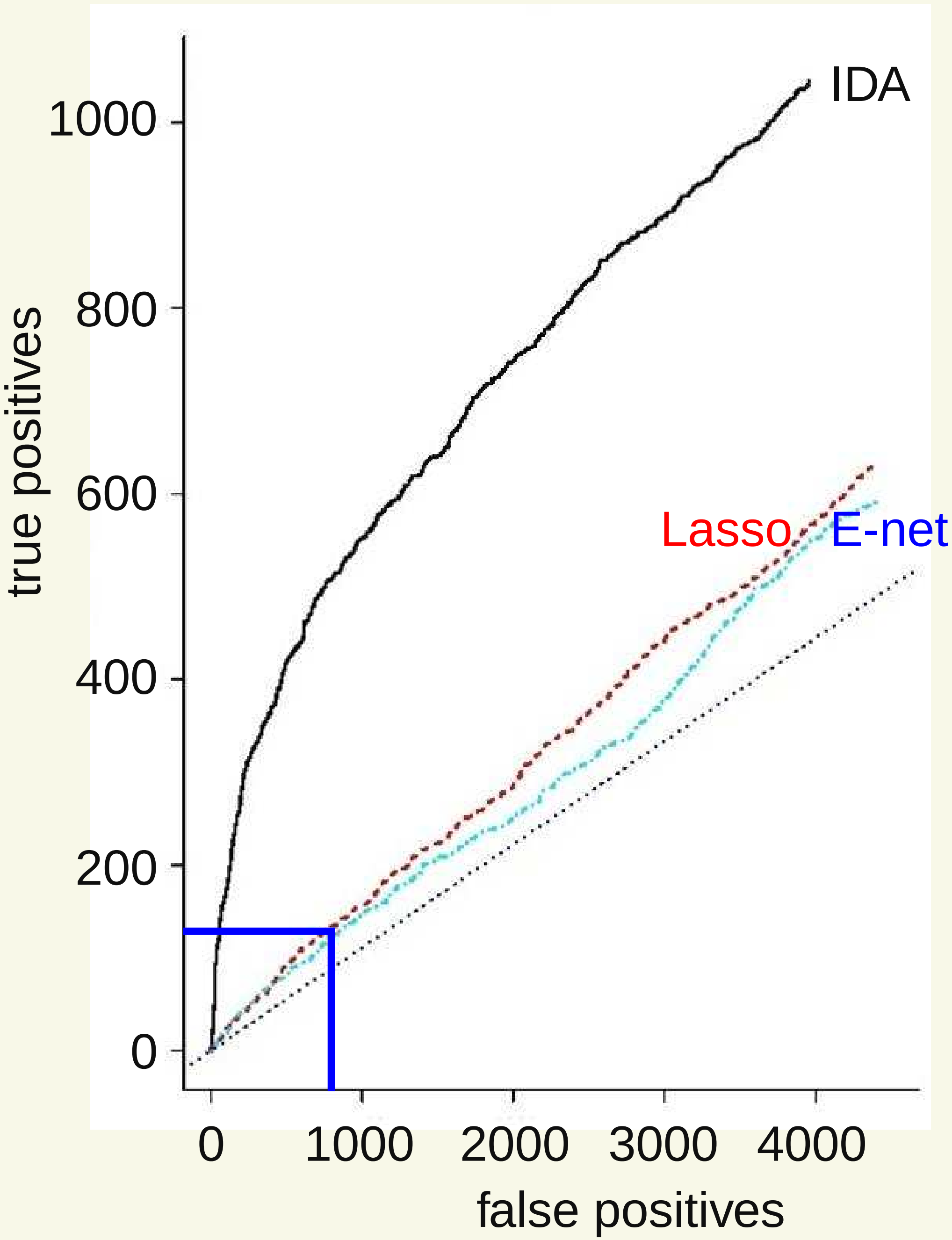
Target set: top 10% of effects from experimental data

Consider top $q = 1000$ effects

	TP	FP
Random guessing	100	900

Source: Nature Methods, 2010

Validation of IDA on yeast gene expression data



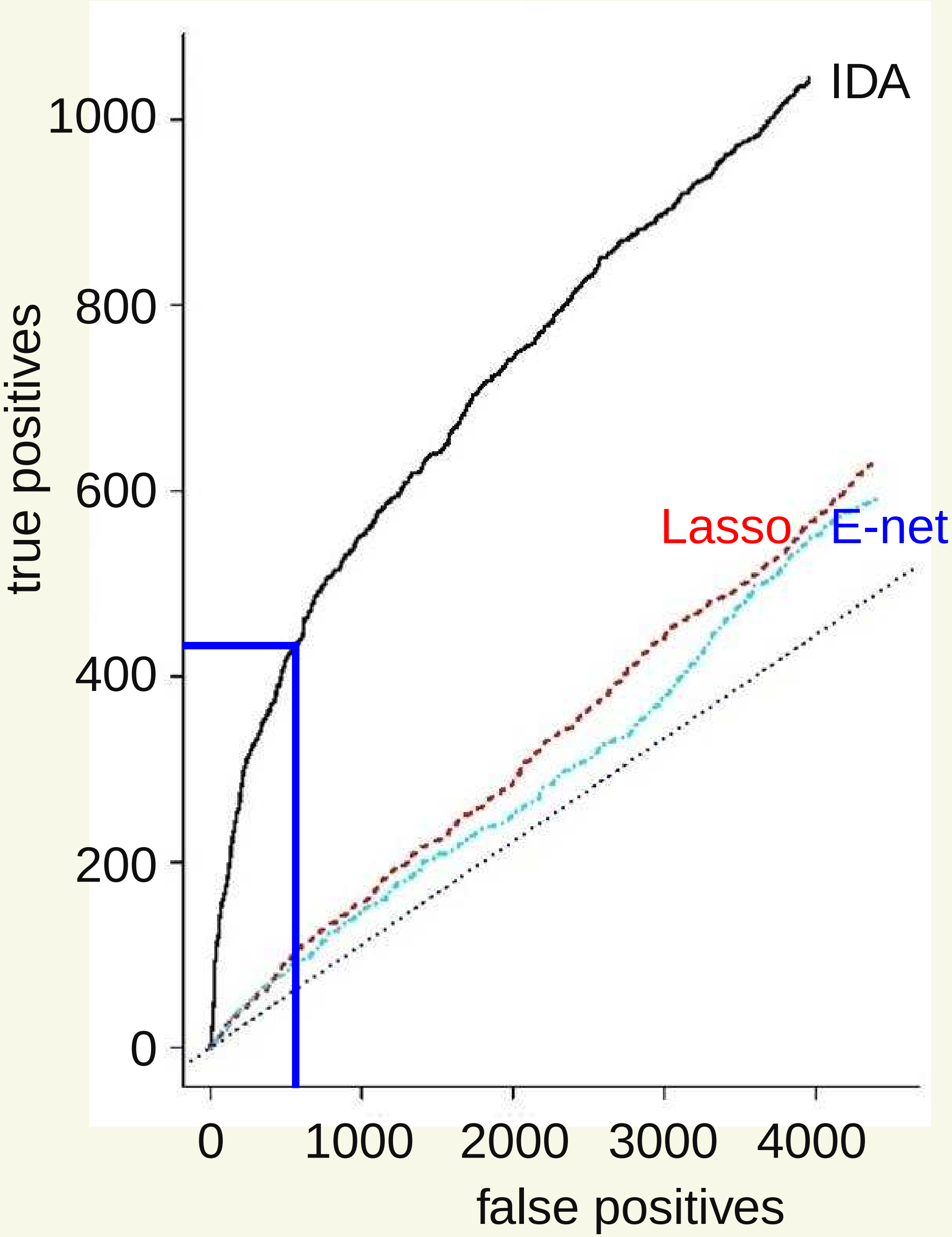
Target set: top 10% of effects from experimental data

Consider top $q = 1000$ effects

	TP	FP
Random guessing	100	900
Lasso / E-net	130	870

Source: Nature Methods, 2010

Validation of IDA on yeast gene expression data



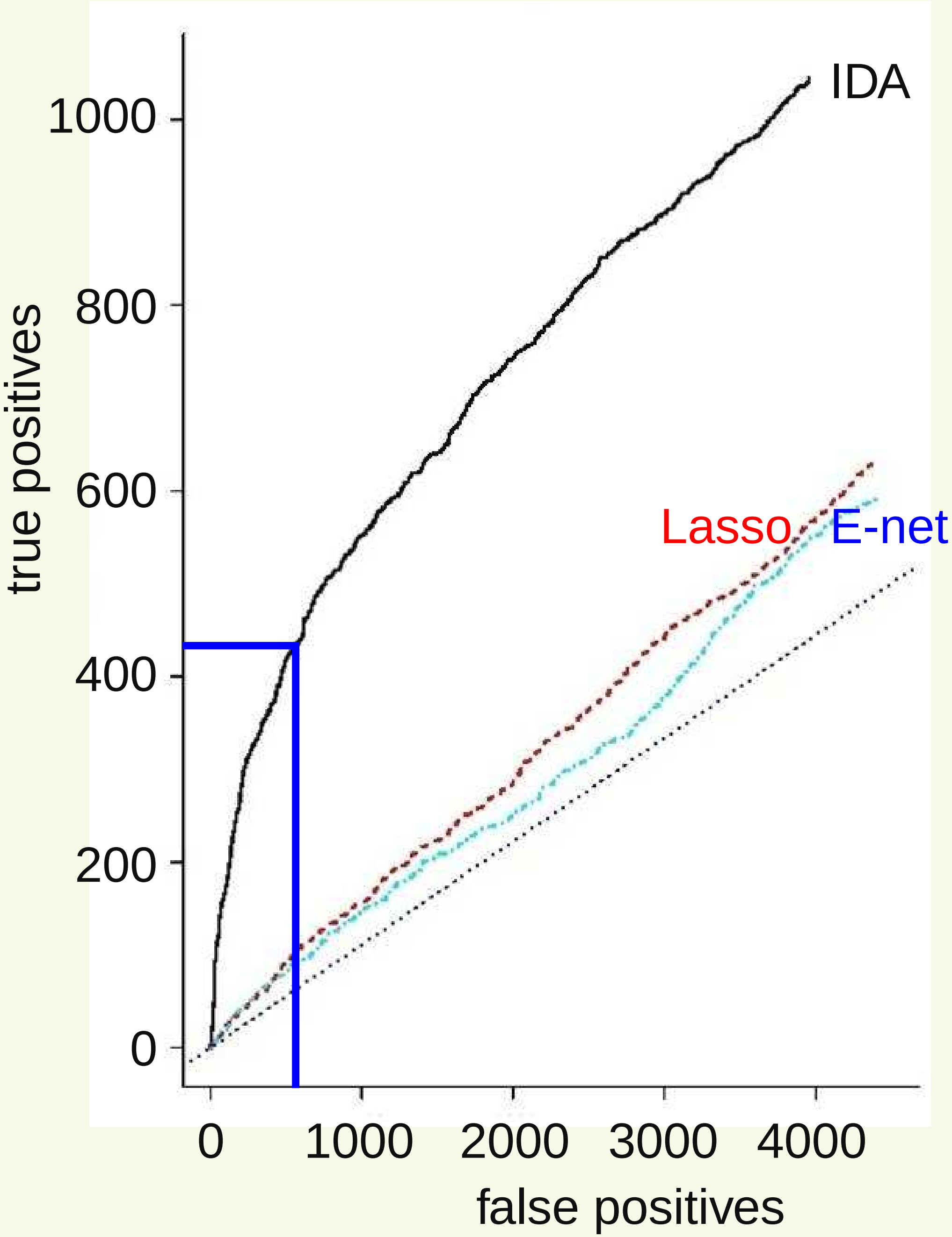
Target set: top 10% of effects from experimental data

Consider top $q = 1000$ effects

	TP	FP
Random guessing	100	900
Lasso / E-net	130	870
IDA	425	575

Source: Nature Methods, 2010

Validation of IDA on yeast gene expression data



Target set: top 10% of effects from experimental data

Consider top $q = 1000$ effects

	TP	FP
Random guessing	100	900
Lasso / E-net	130	870
IDA	425	575

Possible use:
design of experiments

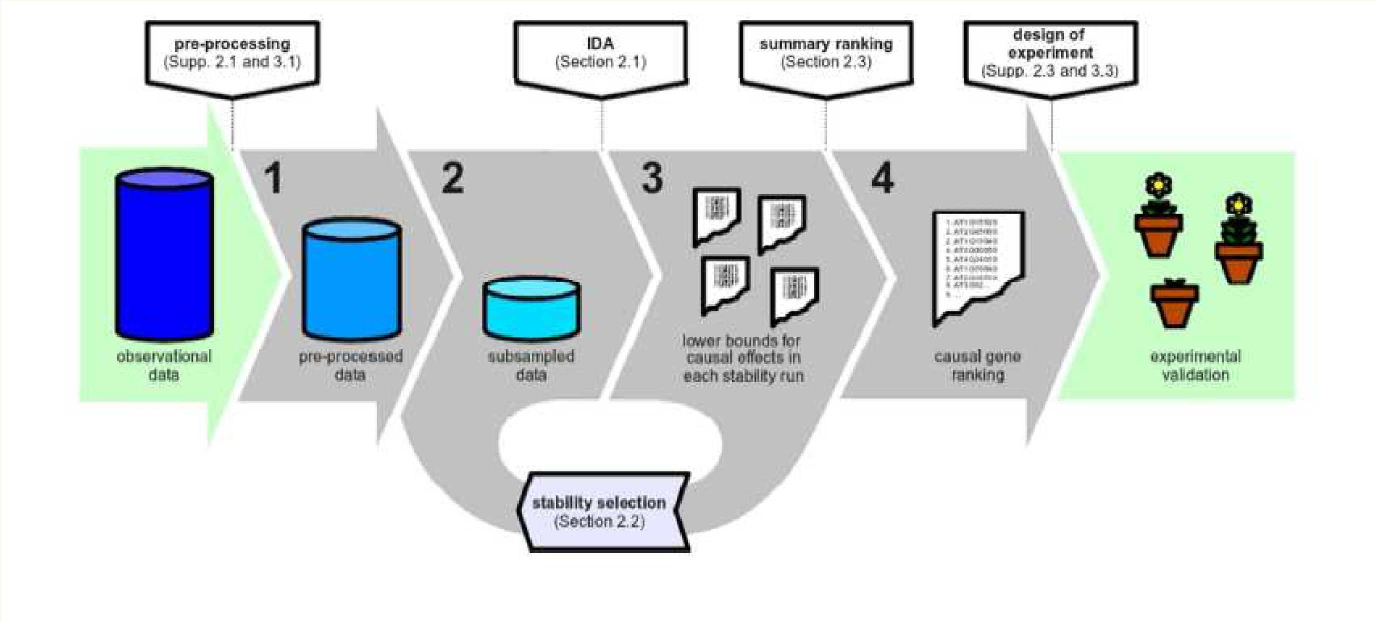
Source: Nature Methods, 2010

Further work and extensions

- R-package pcalg
  (Kalisch et al 2012, J. Stat. Softw.)

Further work and extensions

- R-package pcalg
  (Kalisch et al 2012, J. Stat. Softw.)
- Improved performance in combination with sub-sampling
  (Stekhoven et al 2012, Bioinformatics)



Further work and extensions

- R-package pcalg

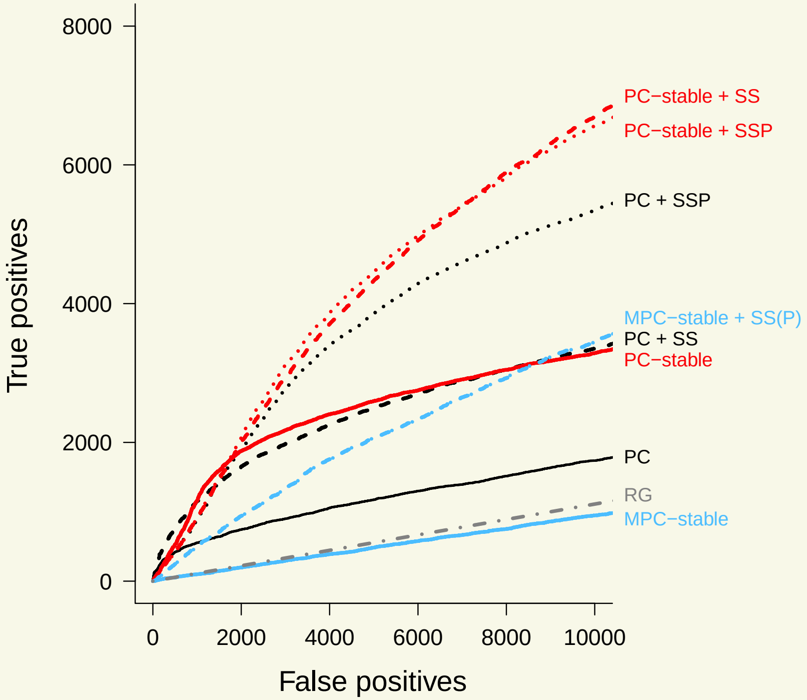
  (Kalisch et al 2012, J. Stat. Softw.)

- Improved performance in combination with sub-sampling

  (Stekhoven et al 2012, Bioinformatics)

- Resolving order-dependence in the PC algorithm

  (Colombo & Maathuis 2014, JMLR)



Further work and extensions

- R-package pcalg
  (Kalisch et al 2012, J. Stat. Softw.)
- Improved performance in combination with sub-sampling
  (Stekhoven et al 2012, Bioinformatics)
- Resolving order-dependence in the PC algorithm
  (Colombo & Maathuis 2014, JMLR)
- Improved performance with other causal structure learning methods
  (Nandy et al, arXiv:1507.02608)

Further work and extensions

- R-package pcalg
  (Kalisch et al 2012, J. Stat. Softw.)
- Improved performance in combination with sub-sampling
  (Stekhoven et al 2012, Bioinformatics)
- Resolving order-dependence in the PC algorithm
  (Colombo & Maathuis 2014, JMLR)
- Improved performance with other causal structure learning methods
  (Nandy et al, arXiv:1507.02608)
- joint-IDA: allowing for multiple simultaneous interventions
  (Nandy et al 2017, Ann. Statist.)

Further work and extensions

- R-package pcalg
  (Kalisch et al 2012, J. Stat. Softw.)
- Improved performance in combination with sub-sampling
  (Stekhoven et al 2012, Bioinformatics)
- Resolving order-dependence in the PC algorithm
  (Colombo & Maathuis 2014, JMLR)
- Improved performance with other causal structure learning methods
  (Nandy et al, arXiv:1507.02608)
- joint-IDA: allowing for multiple simultaneous interventions
  (Nandy et al 2017, Ann. Statist.)
- LV-IDA: allowing for latent variables
  (Malinsky & Spirtes 2016, PGM)

Further work and extensions

- R-package pcalg
  (Kalisch et al 2012, J. Stat. Softw.)
- Improved performance in combination with sub-sampling
  (Stekhoven et al 2012, Bioinformatics)
- Resolving order-dependence in the PC algorithm
  (Colombo & Maathuis 2014, JMLR)
- Improved performance with other causal structure learning methods
  (Nandy et al, arXiv:1507.02608)
- joint-IDA: allowing for multiple simultaneous interventions
  (Nandy et al 2017, Ann. Statist.)
- LV-IDA: allowing for latent variables
  (Malinsky & Spirtes 2016, PGM)
- Complete graphical criteria for covariate adjustment
  (Perković et al 2015, UAI; Perković et al 2016, JMLR)

Summary

- There is a need for causal methods for observational data

Summary

- There is a need for causal methods for observational data
- Such methods cannot replace randomized controlled experiments. But they can be very valuable as exploratory method:
  - hypothesis generation
  - prioritization of experiments

Summary

- There is a need for causal methods for observational data
- Such methods cannot replace randomized controlled experiments. But they can be very valuable as exploratory method:
  - hypothesis generation
  - prioritization of experiments
- IDA estimates bounds on causal effects from observational data, assuming the data come from an unknown DAG:
  - computationally feasible for large sparse systems
  - statistical properties (consistency)
  - validations in biological systems
  - various extensions available

Thank you!

maathuis@stat.math.ethz.ch