

# Functional Significance Checking in Noisy Biological Networks

S. Akshay, Sukanya Basu, Supratik Chakraborty  
Rangapriya Sundarajan, Prasanna Venkataraman

Indian Institute of Technology Bombay  
Advanced Centre for Treatment, Research & Education in Cancer  
India

CP 2019

# Typical Molecular Cell Biology Experiment

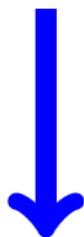
# Typical Molecular Cell Biology Experiment

Hypothesis formation



## Typical Molecular Cell Biology Experiment

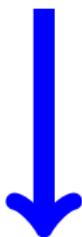
## Hypothesis formation



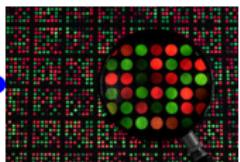
## Reagents, cells, ...

# Typical Molecular Cell Biology Experiment

Hypothesis formation



Reagents, cells, ...



$\mu$ array/RNA-Seq

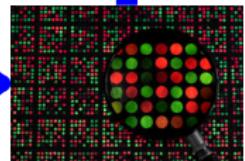
# Typical Molecular Cell Biology Experiment

Hypothesis formation



Reagents, cells, ...

Domain knowledge



$\mu$ array/RNA-Seq

# Typical Molecular Cell Biology Experiment

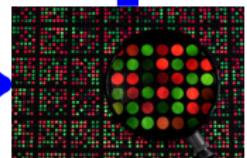
Hypothesis formation



Domain knowledge



Reagents, cells, ...



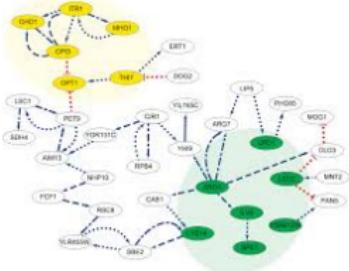
$\mu$ array/RNA-Seq

# Typical Molecular Cell Biology Experiment

Hypothesis formation



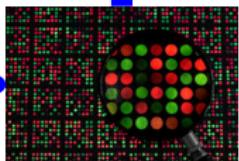
Domain knowledge



Explanation!



Reagents, cells, ...



$\mu$ array/RNA-Seq

# Typical Molecular Cell Biology Experiment

Hypothesis formation

Domain knowledge

- Wetlab costs time (hours, days ...) and money

Reagents, cells, ...

μarray / RNA-seq

# Typical Molecular Cell Biology Experiment

Hypothesis formation

Domain knowledge

- **Wetlab costs time (hours, days ...) and money**
- Can we provide computational help in designing focused wetlab experiments?

Reagents, cells, ...

Proteins / RNA-seq

# Typical Molecular Cell Biology Experiment

Hypothesis formation

Domain knowledge

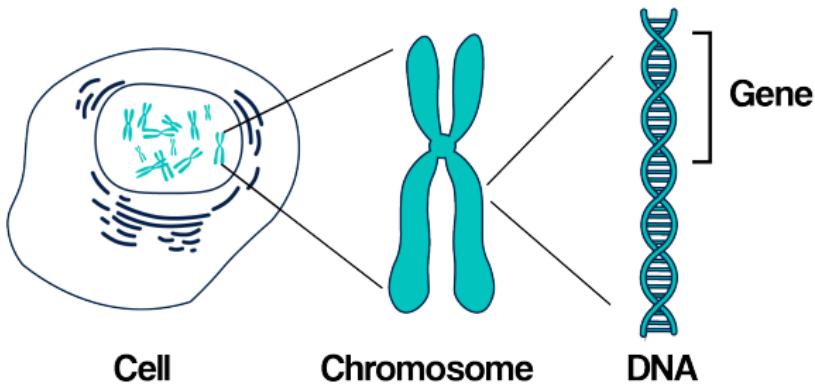
- **Wetlab costs time (hours, days ...) and money**
- Can we provide computational help in designing focused wetlab experiments?
- Can we shortlist agents playing crucial role in outcome of experiment?

Reagents, cells, ...

array, RNA-seq

# Molecular Biology: A Quick Primer

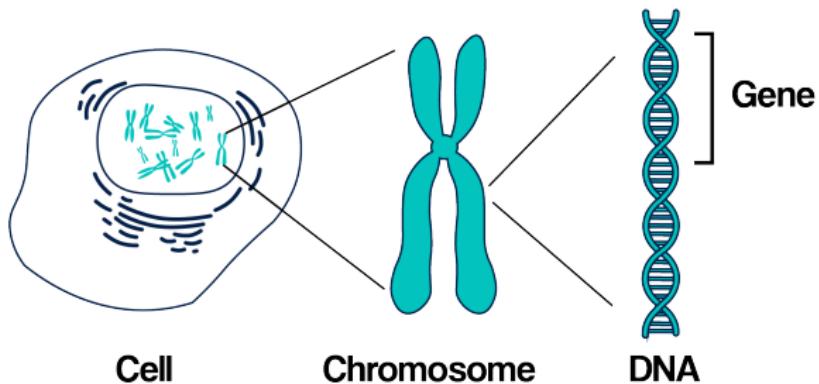
- Cell, nucleus, chromosome, DNA



Source: kintalk.org

# Molecular Biology: A Quick Primer

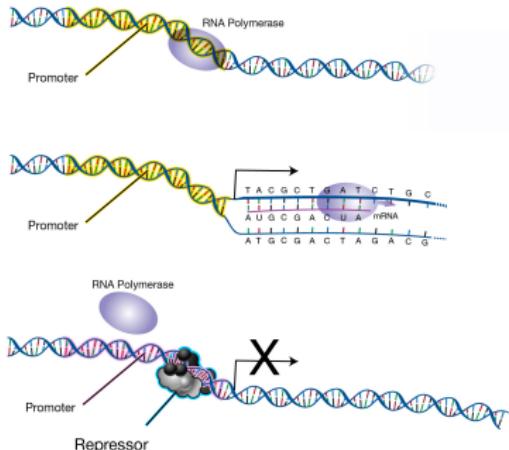
- Cell, nucleus, chromosome, DNA
- Gene: Part of DNA encoding a small set of functions



Source: kintalk.org

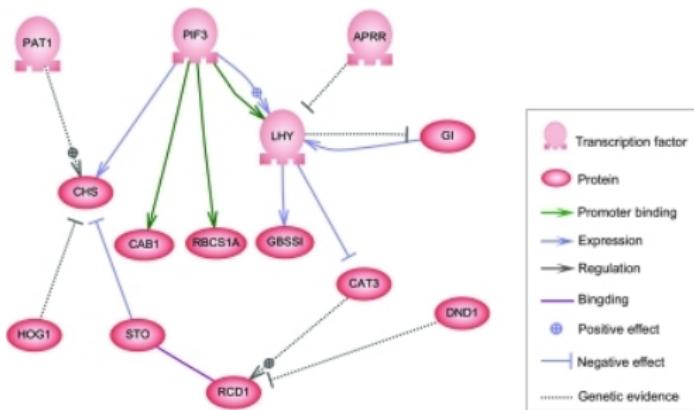
# Molecular Biology: A Quick Primer

- Cell, nucleus, chromosome, DNA
- Gene
- Gene expression, repression, ... interactions



# Molecular Biology: A Quick Primer

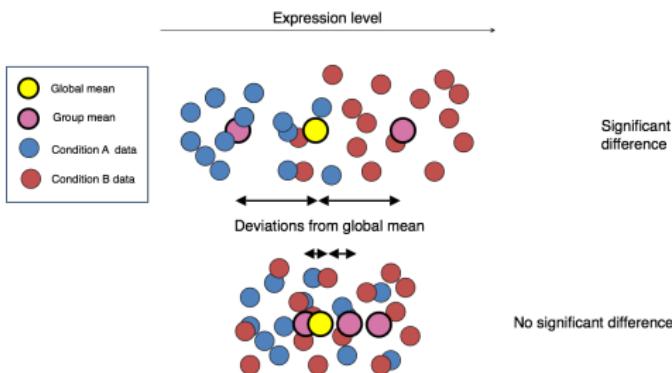
- Cell, nucleus, chromosome, DNA
- Gene
- Gene expression, repression, ... interactions
- Gene regulatory pathways: Encoding domain knowledge



Source: Wikipedia

# Molecular Biology: A Quick Primer

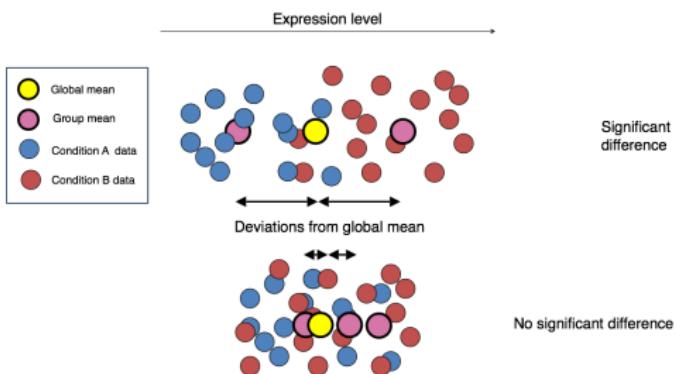
- Cell, nucleus, chromosome, DNA
- Gene
- Gene expression, repression, ... interactions
- Gene regulatory pathways
- Differential gene expression



# Molecular Biology: A Quick Primer

- Cell, nucleus, chromosome, DNA
- Gene
- Gene expression, repression, ... interactions
- Gene regulatory pathways
- Differential gene expression

Not differentially expressed  $\neq$  Not expressed/repressed



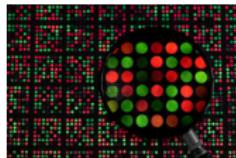
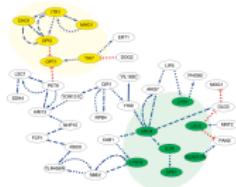
# A Biologist's Ask

Are observed diff. expressions consistent with domain knowledge?  
If so, can we find one/some/all explanations?

## A Biologist's Ask

Are observed diff. expressions consistent with domain knowledge?  
If so, can we find one/some/all explanations?

## Domain Knowledge as Network

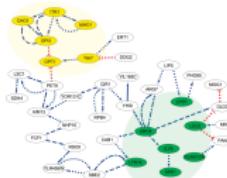


## Differential Expression Data

# A Biologist's Ask

Are observed diff. expressions consistent with domain knowledge?  
If so, can we find one/some/all explanations?

Domain Knowledge as Network



Algorithm ?

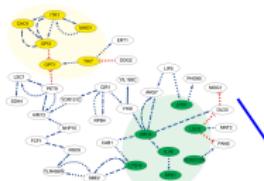


Differential Expression Data

# A Biologist's Ask

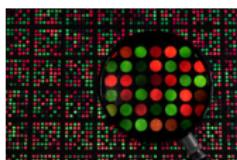
Are observed diff. expressions consistent with domain knowledge?  
If so, can we find one/some/all explanations?

Domain Knowledge as Network



No Explanation

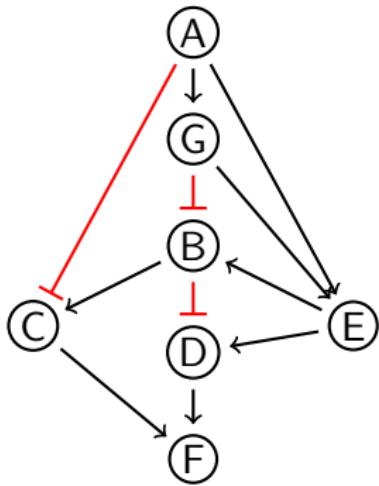
Algorithm ?



Explanation(s)

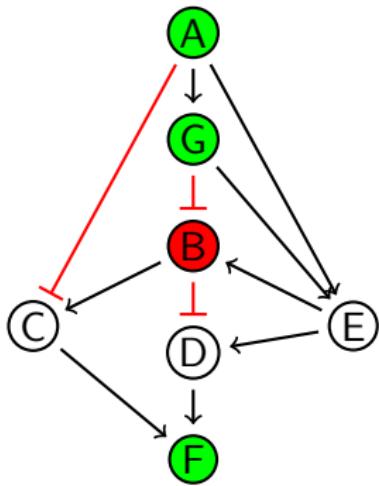
Differential Expression Data

# A Toy Example



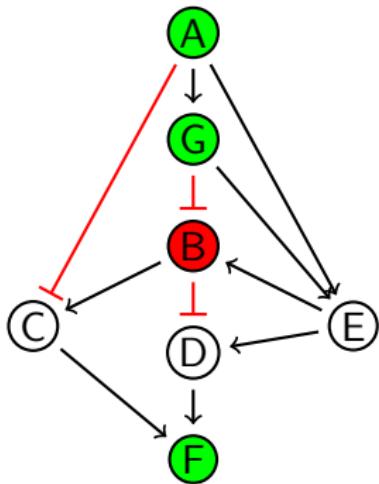
- 7 genes of interest: A, B, C, D, E, F, G

# A Toy Example



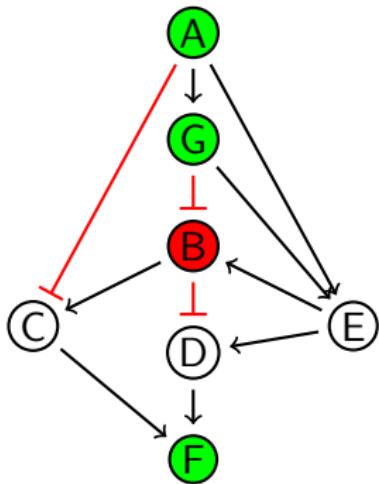
- 7 genes of interest: A, B, C, D, E, F, G
- Differential Expressions
  - A↑, G↑, F↑
  - B↓
  - C, D, E not differentially expressed
    - Expression not significantly different from control
    - Could have been ↑ or ↓

# A Toy Example



- 7 genes of interest: A, B, C, D, E, F, G
- Differential Expressions
  - A↑, G↑, F↑
  - B↓
  - C, D, E not differentially expressed
    - Expression not significantly different from control
    - Could have been ↑ or ↓
- Assumption: Overexpressed genes activated, repressed genes inhibited

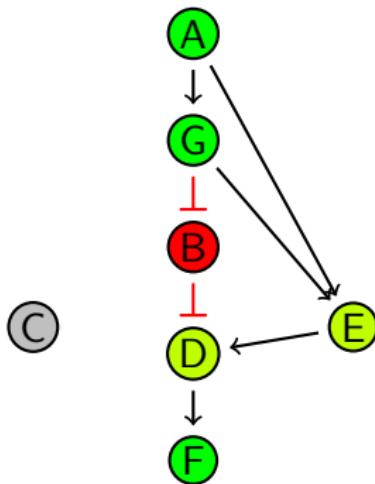
# A Toy Example



- 7 genes of interest: A, B, C, D, E, F, G
- Differential Expressions
  - $A \uparrow, G \uparrow, F \uparrow$
  - $B \downarrow$
  - C, D, E not differentially expressed
    - Expression not significantly different from control
    - Could have been  $\uparrow$  or  $\downarrow$
- Assumption: Overexpressed genes activated, repressed genes inhibited

Is there a network-consistent explanation of diff. expressions?

# A Toy Example



- 7 genes of interest: A, B, C, D, E, F, G
- Differential Expressions
  - A↑, G↑, F↑
  - B↓
  - C, D, E not differentially expressed
    - Expression not significantly different from control
    - Could have been ↑ or ↓

Is there a network-consistent explanation of diff. expressions?

**Yes**, there is an **explanation subgraph**

# Explanation Subgraph Problem

**Given:**

- Domain-knowledge network  $G = (V, E, \mu)$ , where  
 $\mu : E \rightarrow \{\text{activation, inhibition}\}$

# Explanation Subgraph Problem

**Given:**

- Domain-knowledge network  $G = (V, E, \mu)$ , where  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$

# Explanation Subgraph Problem

**Given:**

- Domain-knowledge network  $G = (V, E, \mu)$ , where  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$

# Explanation Subgraph Problem

**Given:**

- Domain-knowledge network  $G = (V, E, \mu)$ , where  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$

**Find** a subgraph  $G'$  of  $G$  and  $\lambda' : V \rightarrow \{\uparrow, \downarrow\}$  such that

# Explanation Subgraph Problem

**Given:**

- Domain-knowledge network  $G = (V, E, \mu)$ , where  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$

**Find** a subgraph  $G'$  of  $G$  and  $\lambda' : V \rightarrow \{\uparrow, \downarrow\}$  such that

- $s, t$  in  $G'$

# Explanation Subgraph Problem

**Given:**

- Domain-knowledge network  $G = (V, E, \mu)$ , where  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$

**Find** a subgraph  $G'$  of  $G$  and  $\lambda' : V \rightarrow \{\uparrow, \downarrow\}$  such that

- $s, t$  in  $G'$
- $\lambda'$  consistent with  $\lambda$

# Explanation Subgraph Problem

**Given:**

- Domain-knowledge network  $G = (V, E, \mu)$ , where  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$

**Find** a subgraph  $G'$  of  $G$  and  $\lambda' : V \rightarrow \{\uparrow, \downarrow\}$  such that

- $s, t$  in  $G'$
- $\lambda'$  consistent with  $\lambda$
- Every vertex in  $G'$  reachable from  $s$ .

# Explanation Subgraph Problem

**Given:**

- Domain-knowledge network  $G = (V, E, \mu)$ , where  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$

**Find** a subgraph  $G'$  of  $G$  and  $\lambda' : V \rightarrow \{\uparrow, \downarrow\}$  such that

- $s, t$  in  $G'$
- $\lambda'$  consistent with  $\lambda$
- Every vertex in  $G'$  reachable from  $s$ .
- Every  $s-t$  path in  $G'$  passes through some  $v \notin \{s, t\}$  with  $\lambda(v) = \uparrow$ .

# Explanation Subgraph Problem

**Given:**

- Domain-knowledge network  $G = (V, E, \mu)$ , where  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$

**Find** a subgraph  $G'$  of  $G$  and  $\lambda' : V \rightarrow \{\uparrow, \downarrow\}$  such that

- $s, t$  in  $G'$
- $\lambda'$  consistent with  $\lambda$
- Every vertex in  $G'$  reachable from  $s$ .
- Every  $s-t$  path in  $G'$  passes through some  $v \notin \{s, t\}$  with  $\lambda(v) = \uparrow$ .
- Every  $v$  in  $G'$  with  $\lambda(v) = \uparrow$  lies on some  $s-t$  path.

# Explanation Subgraph Problem

**Given:**

- Domain-knowledge network  $G = (V, E, \mu)$ , where  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$

**Find** a subgraph  $G'$  of  $G$  and  $\lambda' : V \rightarrow \{\uparrow, \downarrow\}$  such that

- $s, t$  in  $G'$
- $\lambda'$  consistent with  $\lambda$
- Every vertex in  $G'$  reachable from  $s$ .
- Every  $s-t$  path in  $G'$  passes through some  $v \notin \{s, t\}$  with  $\lambda(v) = \uparrow$ .
- Every  $v$  in  $G'$  with  $\lambda(v) = \uparrow$  lies on some  $s-t$  path.

$G'$  is called an **explanation subgraph** for  $G$  and  $\lambda$ .

# Explanation Subgraph Problem

**Given:**

- Domain-knowledge network  $G = (V, E, \mu)$ , where  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$

**Find** a subgraph  $G'$  of  $G$  and  $\lambda' : V \rightarrow \{\uparrow, \downarrow\}$  such that

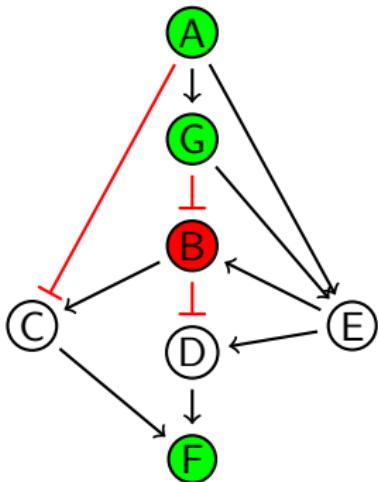
- $s, t$  in  $G'$
- $\lambda'$  consistent with  $\lambda$
- Every vertex in  $G'$  reachable from  $s$ .
- Every  $s-t$  path in  $G'$  passes through some  $v \notin \{s, t\}$  with  $\lambda(v) = \uparrow$ .
- Every  $v$  in  $G'$  with  $\lambda(v) = \uparrow$  lies on some  $s-t$  path.

$G'$  is called an **explanation subgraph** for  $G$  and  $\lambda$ .

**Theorem**

Checking existence of explanation subgraph is NP-complete.

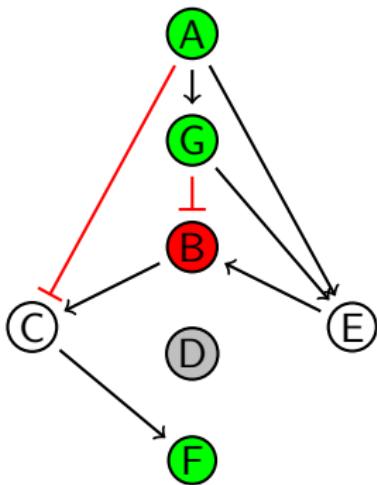
# Toy Example Revisited



- 7 genes of interest: A, B, C, D, E, F, G
- Differential Expressions
  - A↑, G↑, F↑
  - B↓
  - C, D, E not differentially expressed
    - Expression not significantly different from control
    - Could have been ↑ or ↓

There is an **explanation subgraph**.

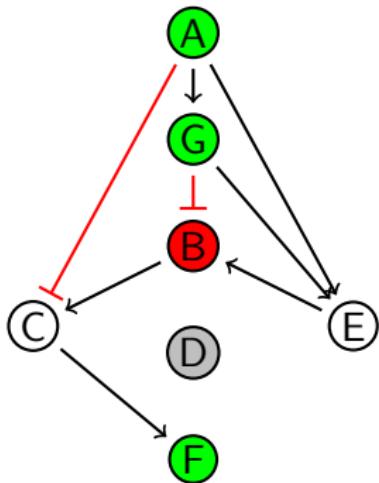
# Toy Example Revisited



- 7 genes of interest: A, B, C, D, E, F, G
- Differential Expressions
  - A↑, G↑, F↑
  - B↓
  - C, D, E not differentially expressed
    - Expression not significantly different from control
    - Could have been ↑ or ↓

What if D didn't play any role?

# Toy Example Revisited

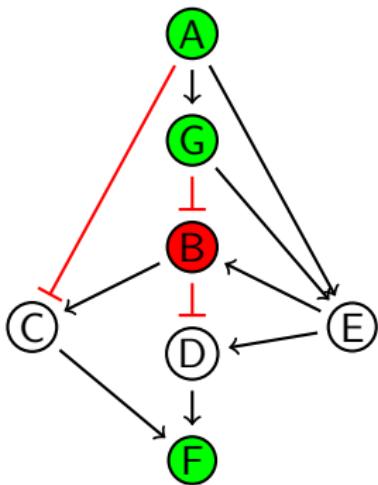


- 7 genes of interest: A, B, C, D, E, F, G
- Differential Expressions
  - A↑, G↑, F↑
  - B↓
  - C, D, E not differentially expressed
    - Expression not significantly different from control
    - Could have been ↑ or ↓

What if D didn't play any role?

No, there is no **explanation subgraph**  
Despite topological paths connecting A, G, B, F

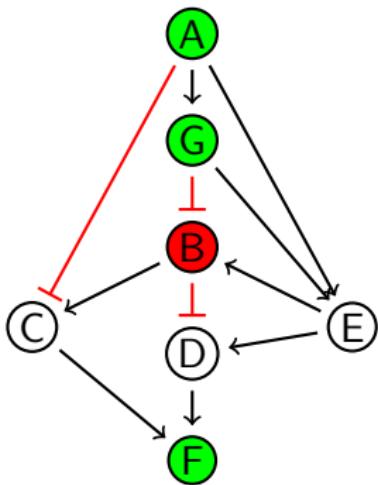
# Toy Example Revisited



- 7 genes of interest: A, B, C, D, E, F, G
- Differential Expressions
  - A $\uparrow$ , G $\uparrow$ , F $\uparrow$
  - B $\downarrow$
  - C, D, E not differentially expressed
    - Expression not significantly different from control
    - Could have been  $\uparrow$  or  $\downarrow$

If all diff. expressions & domain-knowledge encoding correct

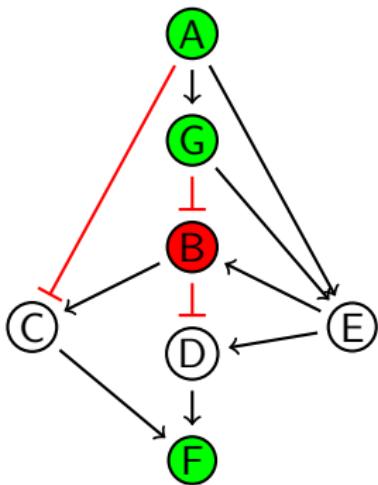
# Toy Example Revisited



- 7 genes of interest: A, B, C, D, E, F, G
- Differential Expressions
  - A↑, G↑, F↑
  - B↓
  - C, D, E not differentially expressed
    - Expression not significantly different from control
    - Could have been ↑ or ↓

If all diff. expressions & domain-knowledge encoding correct  
D is **functionally significant** in explaining diff. expressions.

# Toy Example Revisited

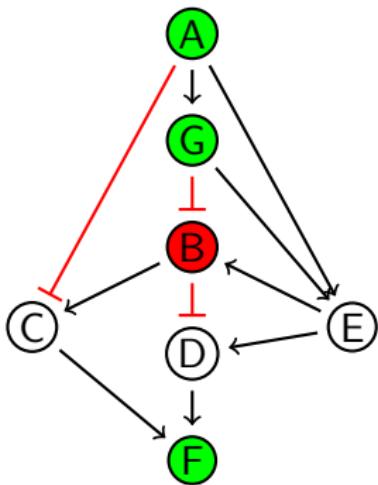


- 7 genes of interest: A, B, C, D, E, F, G
- Differential Expressions
  - A↑, G↑, F↑
  - B↓
  - C, D, E not differentially expressed
    - Expression not significantly different from control
    - Could have been ↑ or ↓

What if

- At most one diff. expression measurement is in error?

# Toy Example Revisited

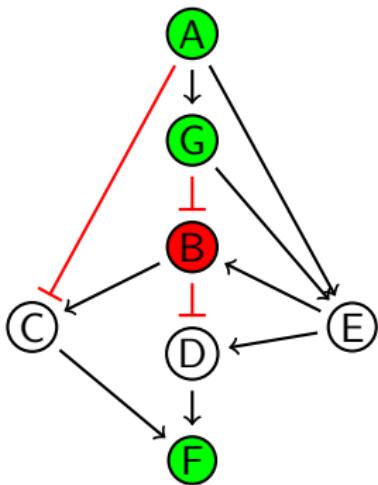


- 7 genes of interest: A, B, C, D, E, F, G
- Differential Expressions
  - A↑, G↑, F↑
  - B↓
  - C, D, E not differentially expressed
    - Expression not significantly different from control
    - Could have been ↑ or ↓

What if

- At most one diff. expression measurement is in error?
- At most one edge in domain knowledge network is in error?

# Toy Example Revisited



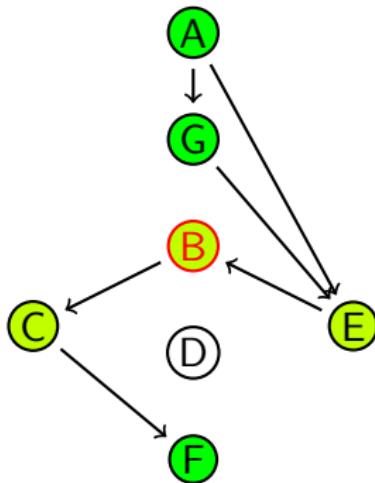
- 7 genes of interest: A, B, C, D, E, F, G
- Differential Expressions
  - A↑, G↑, F↑
  - B↓
  - C, D, E not differentially expressed
    - Expression not significantly different from control
    - Could have been ↑ or ↓

What if

- At most one diff. expression measurement is in error?
- At most one edge in domain knowledge network is in error?

**Is D still necessary to find an explanation?**

# Toy Example Revisited

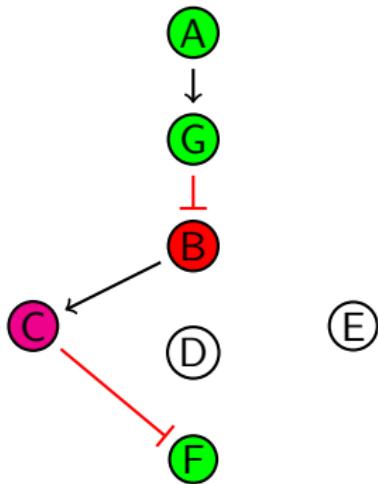


- 7 genes of interest: A, B, C, D, E, F, G
- Differential Expressions
  - A↑, G↑, F↑
  - B↓
  - C, D, E not differentially expressed
    - Expression not significantly different from control
    - Could have been ↑ or ↓

At most one diff. expression measurement in error

D is **no longer functionally significant**

# Toy Example Revisited



- 7 genes of interest: A, B, C, D, E, F, G
- Differential Expressions
  - A↑, G↑, F↑
  - B↓
  - C, D, E not differentially expressed
    - Expression not significantly different from control
    - Could have been ↑ or ↓

At most one edge in domain knowledge network in error

D is **no longer functionally significant**

# The Biologist Asks for More

## Domain Knowledge as Network



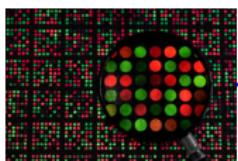
Algorithm



Gene1

Gene2

Gene3

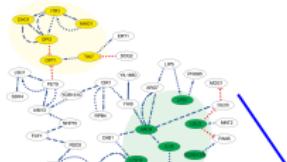


Differential Expression Data

“Functionally significant”

# The Biologist Asks for More

Domain Knowledge as Network



Gene1

**Too difficult to solve in practice**



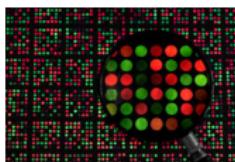
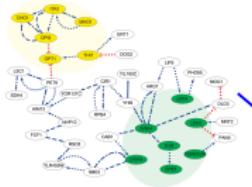
Gene3

Differential Expression Data

“Functionally significant”

# Simplified Biologist's Ask

Domain Knowledge as Network



Differential Expression Data

Gene id  
Candidate func. sig. gene

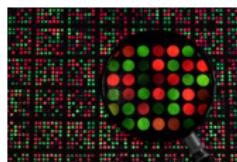
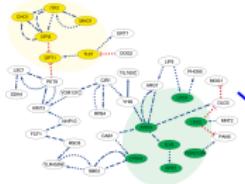
Algorithm



Is Gene func. sig.?  
Yes/No

# Simplified Biologist's Ask

Domain Knowledge as Network



Differential Expression Data

Gene id  
Candidate func. sig. gene

Algorithm



Is Gene func. sig.?  
Yes/No



“Noise” Bound

# Relaxed Explanation Subgraph

**Given:**

- Domain knowledge network  $G = (V, E, \mu)$ , with  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$

# Relaxed Explanation Subgraph

**Given:**

- Domain knowledge network  $G = (V, E, \mu)$ , with  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$
- Bound  $n$  on diff. expr. measurement error, and  $e$  on edge label error in domain-knowledge network

# Relaxed Explanation Subgraph

**Given:**

- Domain knowledge network  $G = (V, E, \mu)$ , with  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$
- Bound  $n$  on diff. expr. measurement error, and  $e$  on edge label error in domain-knowledge network
- Relaxation weight  $R : V \cup E \rightarrow \mathbb{N}$  measuring cost of admitting error in a diff. expr. or edge label

# Relaxed Explanation Subgraph

**Given:**

- Domain knowledge network  $G = (V, E, \mu)$ , with  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$
- Bound  $n$  on diff. expr. measurement error, and  $e$  on edge label error in domain-knowledge network
- Relaxation weight  $R : V \cup E \rightarrow \mathbb{N}$  measuring cost of admitting error in a diff. expr. or edge label

**Find** an explanation subgraph  $G'$  for  $G$  and  $\lambda$  such that

# Relaxed Explanation Subgraph

**Given:**

- Domain knowledge network  $G = (V, E, \mu)$ , with  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$
- Bound  $n$  on diff. expr. measurement error, and  $e$  on edge label error in domain-knowledge network
- Relaxation weight  $R : V \cup E \rightarrow \mathbb{N}$  measuring cost of admitting error in a diff. expr. or edge label

**Find** an explanation subgraph  $G'$  for  $G$  and  $\lambda$  such that

- $\sum_{v \in V'} | \text{diff. expr. error in } v | R(v) \leq n$
- $\sum_{e \in E'} | \text{edge label error in } e | R(e) \leq e$

# Relaxed Explanation Subgraph

**Given:**

- Domain knowledge network  $G = (V, E, \mu)$ , with  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$
- Bound  $n$  on diff. expr. measurement error, and  $e$  on edge label error in domain-knowledge network
- Relaxation weight  $R : V \cup E \rightarrow \mathbb{N}$  measuring cost of admitting error in a diff. expr. or edge label

**Find** an explanation subgraph  $G'$  for  $G$  and  $\lambda$  such that

- $\sum_{v \in V'} |\text{diff. expr. error in } v| R(v) \leq n$
- $\sum_{e \in E'} |\text{edge label error in } e| R(e) \leq e$

$G'$  is an  $(n, e)$  **relaxed explanation subgraph** for  $G, \lambda$  and  $R$ .

# Formalizing Functional Significance

**Given:**

- Domain knowledge network  $G = (V, E, \mu)$ , with  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$
- Bounds  $n$  on diff. expr. measurement error and  $e$  on error edge label error in domain-knowledge network
- $R : V \cup E \rightarrow \mathbb{N}$

# Formalizing Functional Significance

**Given:**

- Domain knowledge network  $G = (V, E, \mu)$ , with  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$
- Bounds  $n$  on diff. expr. measurement error and  $e$  on error edge label error in domain-knowledge network
- $R : V \cup E \rightarrow \mathbb{N}$
- Candidate functionally significant vertex  $c$

# Formalizing Functional Significance

**Given:**

- Domain knowledge network  $G = (V, E, \mu)$ , with  $\mu : E \rightarrow \{\text{activation, inhibition}\}$
- Differential expressions  $\lambda : V \rightarrow \{\uparrow, \downarrow, ?\}$
- Stimulus  $s$  and target  $t$  vertices with  $\lambda(s) = \uparrow, \lambda(t) \neq ?$
- Bounds  $n$  on diff. expr. measurement error and  $e$  on error edge label error in domain-knowledge network
- $R : V \cup E \rightarrow \mathbb{N}$
- Candidate functionally significant vertex  $c$

**Does removing  $c$  force us to admit more errors in finding an explanation subgraph?**

# Central Result

## Pareto Dominance

For every  $(n', e')$ -relaxed explanation subgraph without  $c$ , where  $(0, 0) \preceq (n', e') \preceq (n, e)$ , does there exist an  $(n'', e'')$ -relaxed explanation subgraph with  $c$ , where  $(n'', e'') \prec (n', e')$ ?

## Pareto Dominance

For every  $(n', e')$ -relaxed explanation subgraph without  $c$ , where  $(0, 0) \preceq (n', e') \preceq (n, e)$ , does there exist an  $(n'', e'')$ -relaxed explanation subgraph with  $c$ , where  $(n'', e'') \prec (n', e')$ ?

$i'$  is said to be functionally significant for  $G, \lambda, R$ , and  $(n, e)$  if above Pareto dominance holds.

# Central Result

## Pareto Dominance

For every  $(n', e')$ -relaxed explanation subgraph without  $c$ , where  $(0, 0) \preceq (n', e') \preceq (n, e)$ , does there exist an  $(n'', e'')$ -relaxed explanation subgraph with  $c$ , where  $(n'', e'') \prec (n', e')$ ?

$i'$  is said to be functionally significant for  $G, \lambda, R$ , and  $(n, e)$  if above Pareto dominance holds.

Checking functional significance of a candidate node is in  $\Pi_2^P$  and is  $\Sigma_2^P$ -hard.

# Implementation

- Iterative algorithm that finds the Pareto curve for  $(n', e')$  with and without  $c$
- Binary search over the  $(n, e)$ -space to speed up Pareto curve finding
- For each candidate  $(n', e')$ , a SAT-encoding is used to find existence of solution.

# Implementation

- Iterative algorithm that finds the Pareto curve for  $(n', e')$  with and without  $c$
- Binary search over the  $(n, e)$ -space to speed up Pareto curve finding
- For each candidate  $(n', e')$ , a SAT-encoding is used to find existence of solution.

## Observations

- With no “noise”, no solutions (explanation graphs) obtained on real data.
- After introducing small “noise” bounds (viz.  $(10, 15)$ ), hundreds of thousands of solutions obtained.
- Examining each solution practically impossible
  - Hence, functional significance checking is useful in practice.
  - ERK and AKT experiments were validated by wet-lab experiments.

# Some Results

$(n, e) = (30, 30)$

Source-Target pair	F.S. Cand.	Pareto shift	SAT Calls	Time (in hrs)
Synthetic1-5var-W(5, 5)	x	Y	5	.035
Synthetic2-15var-W(5, 5)	x	Y	6	.35
Synthetic3-45var-W(0, 0)	x	Y	2	.004
TNF $\alpha$ -IkBa (Expr/Act merged)	None	-	62	5
TNF $\alpha$ -IkBa (Expr/Act merged)	p38	Y	72	5
TNF $\alpha$ -IkBa (Expr/Act merged)	ERK	N	62	2.6
TNF $\alpha$ -IkBa (Expr/Act merged)	PIK3CA	Y	71	1.5
TNF $\alpha$ -IkBa (Expr/Act merged)	AKT	Y	42	11
TNF $\alpha$ -IkBa (Expr only)	None	-	63	9
TNF $\alpha$ -IkBa (Expr only)	p38	Y	63	15
TNF $\alpha$ -IkBa (Expr only)	ERK	Y	63	15
TNF $\alpha$ -IkBa (Expr only)	PIK3CA	N	68	14
TNF $\alpha$ -IkBa (Expr only)	AKT	N	68	18.4
TNF $\alpha$ -IkBa (Act only)	None	-	64	15.6
TNF $\alpha$ -IkBa (Act only)	p38	Y	64	37
TNF $\alpha$ -IkBa (Act only)	ERK	N	64	25.8
TNF $\alpha$ -IkBa (Act only)	PIK3CA	Y	64	18.5
TNF $\alpha$ -IkBa (Act only)	AKT	Y	54	44
TNF $\alpha$ -A20	None	-	56	0.3
TNF $\alpha$ -A20	ERK	Y	57	0.7
TNF $\alpha$ -A20	AKT	N	52	0.3
TNF $\alpha$ -A20	p38	N	54	0.3

# Conclusion

- End-to-end analysis of functional significance using constraint solving
- Identified wet-lab post-validated functionally significant genes
- Running time still a concern, but better than wet-lab experiments.
- Implementation using 2QBF solver under way.