



BioMANTA: Modeling and Network Analysis of Biological Networks II - Knowledge Discovery and Network Analysis

-{ B y }-

Muhammad Shoaib Sehgal

ARC Centre of Excellence in Bioinformatics
Institute for Molecular Bioscience
University of Queensland Australia

Project outline



BioMANTA

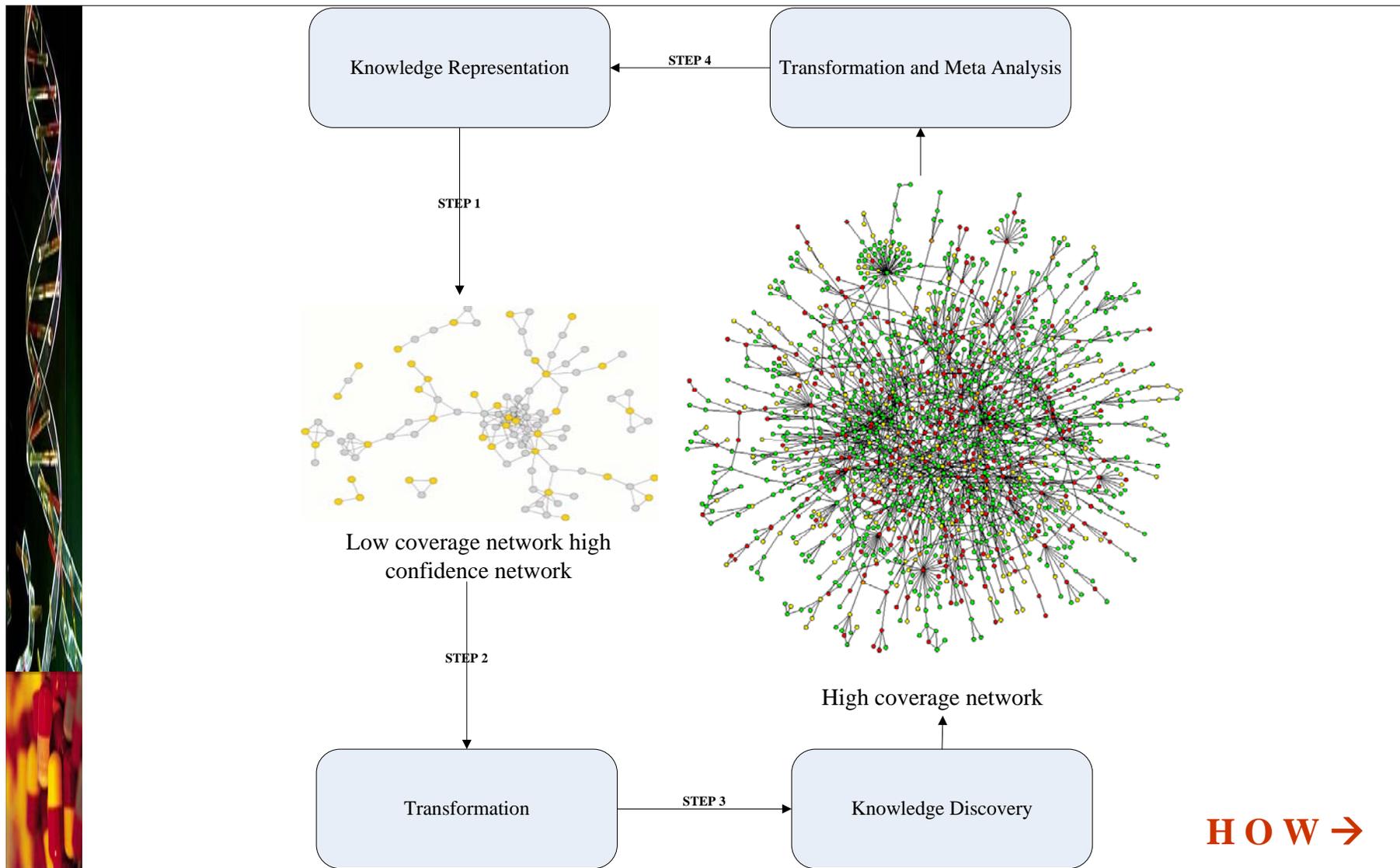
Modeling
and
Analysis of
Biological
Network
Activity

- ✍ Collaboration between Pfizer, RTC at Cambridge, Boston and IMB at UQ.
- ✍ Development of novel biological network analysis methods and infrastructure for querying biological data in a semantically-enabled format.
- ✍ Key Components
 - ✍ Knowledge Representation
 - ✍ Knowledge Discovery
 - ✍ Visualization

Problem overview

- ✗ Proteins operate as a part of **highly interconnected cellular networks** referred to as interactome networks to perform their functionality.
- ✗ Only a small portion of these interaction networks is known.
- ✗ The **objective** is to infer high coverage networks
 - ✗ To analyse interrelationships between drug targets and disease-gene products (**bipartite graph**) to gain insights of drug targets in the context of cellular and disease networks
 - ✗ To **reconstruct a network of genes** (and their products) linked to particular disease for instance, Alzheimer's
 - ✗ To elucidate signaling pathways and biological functions related to genes/proteins networks **to understand cellular functionality**

BioMANTA: Our proposed framework

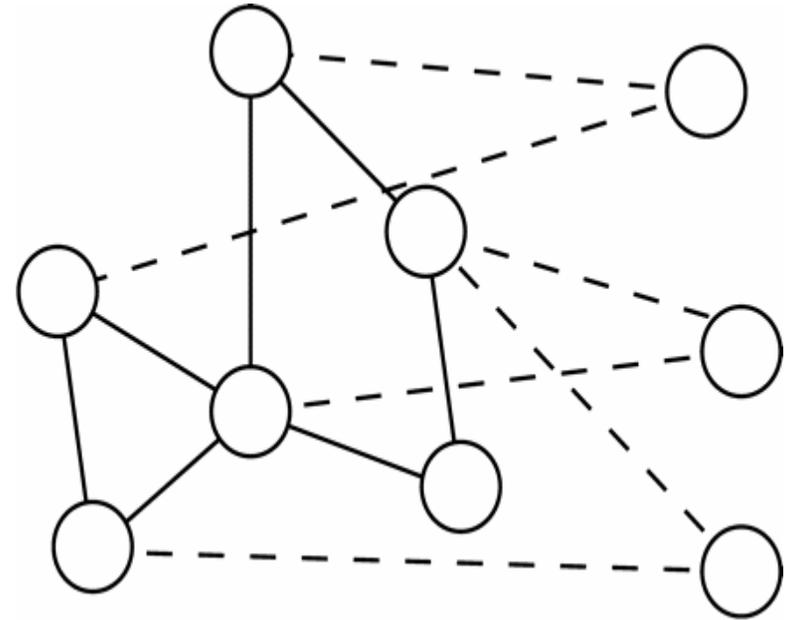


HOW →

Network inference techniques

Unsupervised methods

- ✗ Bayesian networks
- ✗ Differential equations
- ✗ Similarity based methods
- ✗ Hybrid methods
- ✗ Advantages/Disadvantages



K. Bleakley *et al.*, *Bioinformatics*, 2007.

Network inference techniques (cont ...)

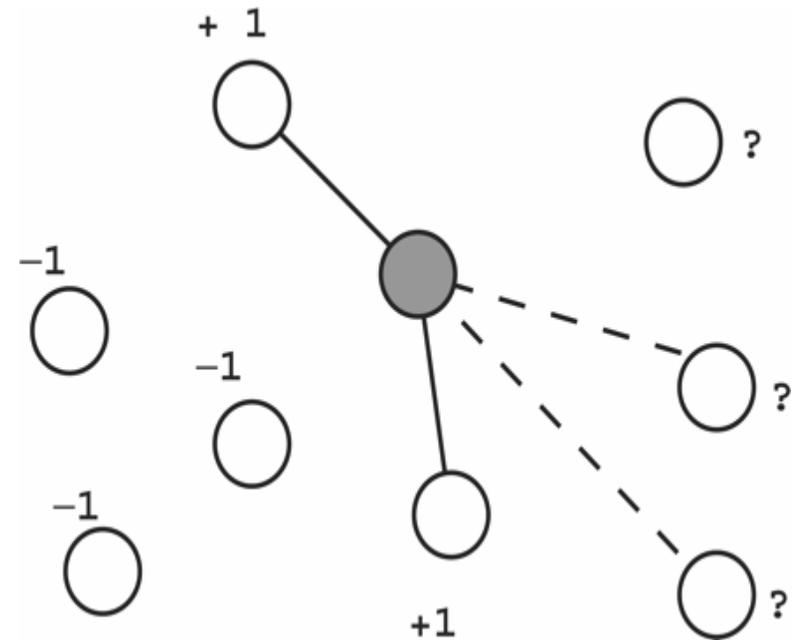
Supervised methods

~~✍~~ SVM

~~✍~~ Stochastic methods

~~✍~~ NN

~~✍~~ Pros/Cons



K. Bleakley *et al.*, *Bioinformatics*, 2007.

Proposed model

Semi-Supervised Model

- ✗ To **infer new links** while concomitantly **keeping FDR low**.
- ✗ **Integration** of networks **using heterogenous data sources**.
- ✗ Inference using:
 - Expression** data (N_{exp}) [$n * m$]
 - Protein localization** data (N_{loc}) [$n * 22$]
 - Phylogenetic** profiles (N_{phy}) [$n * 145$]

$$N = \sum_i w_i N_i$$

Proposed model (cont ...)

Scoring high confidence interactions

- Each PPI interaction is assigned a score based on experimental method and the number of instances an interaction is reported in the literature.
- The weights are computed using:

$$C_i = \sum_{j=1}^M \omega_j \cdot \beta_e$$

where

- C is a score assigned to each interaction
- M are the number of observations reporting i^{th} interaction
- ω is the throughput weight assigned to each observation
- β is the confidence weight assigned to experimental method

Proposed model (cont ...)

Data sets

✗ Yeast

✗ 157 different time stamps

✗ Homo sapiens

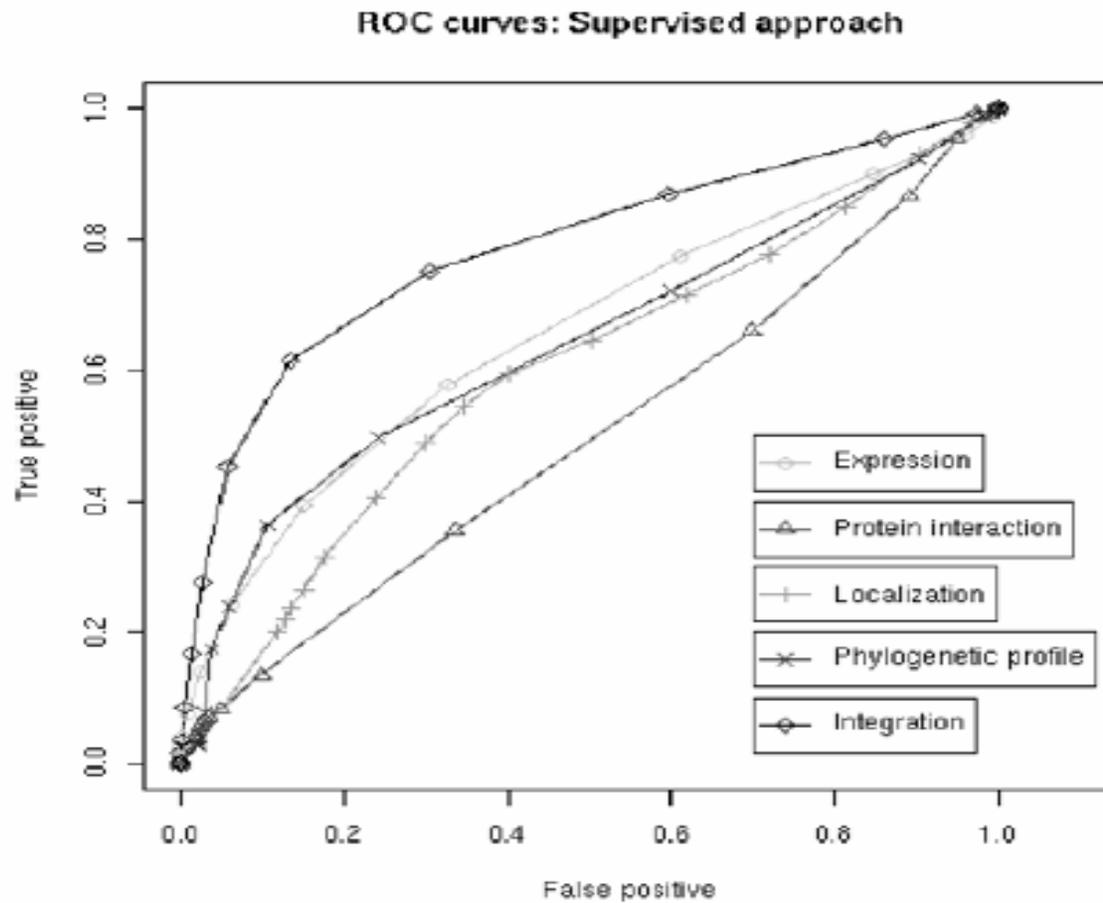
✗ 13 different types of cancer/non-cancerous tissue data including:

breast, pancreas, colon, brain, bladder, ovary, uterus, kidney, liver, lung, lymphoma, stomach, and prostate tissues

✗ Mouse

Idiopathic heart failure and normal heart data

Proposed model (cont ...)



Y. Yamanishi *et al.*, *Bioinformatics*, 2005.

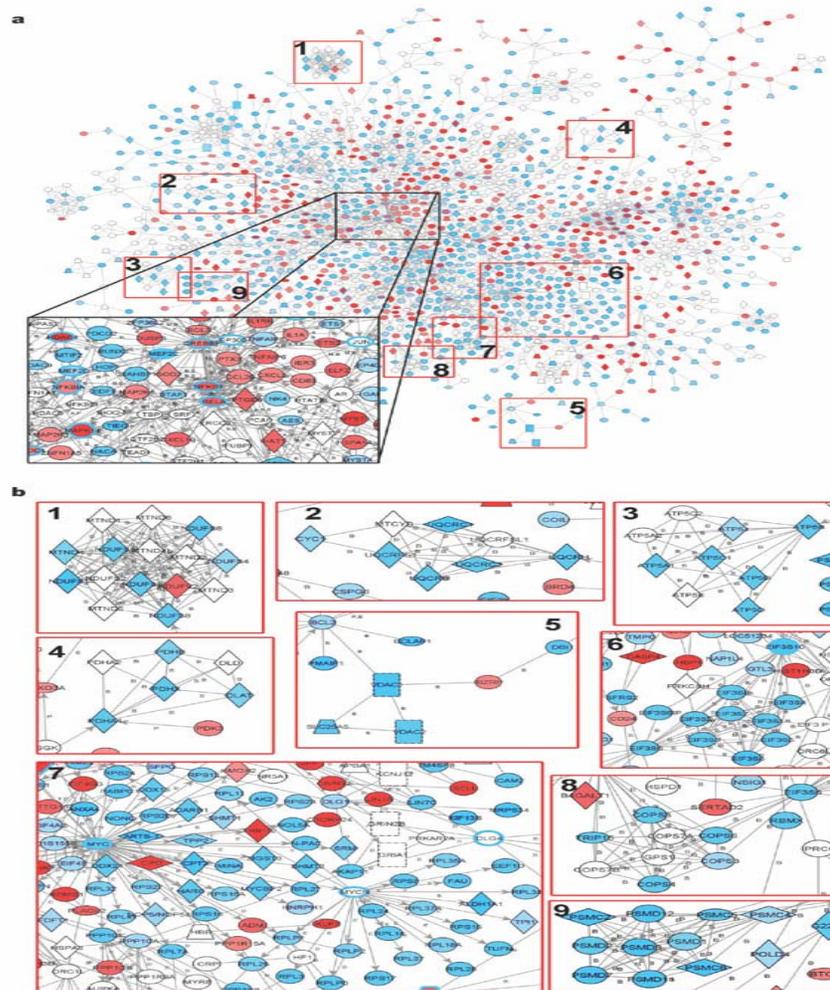
Meta analysis

- ✎ After network inference, **meta analysis** methods are applied to search systematic properties of complex disease genes with respect to their connectedness.

Methods:

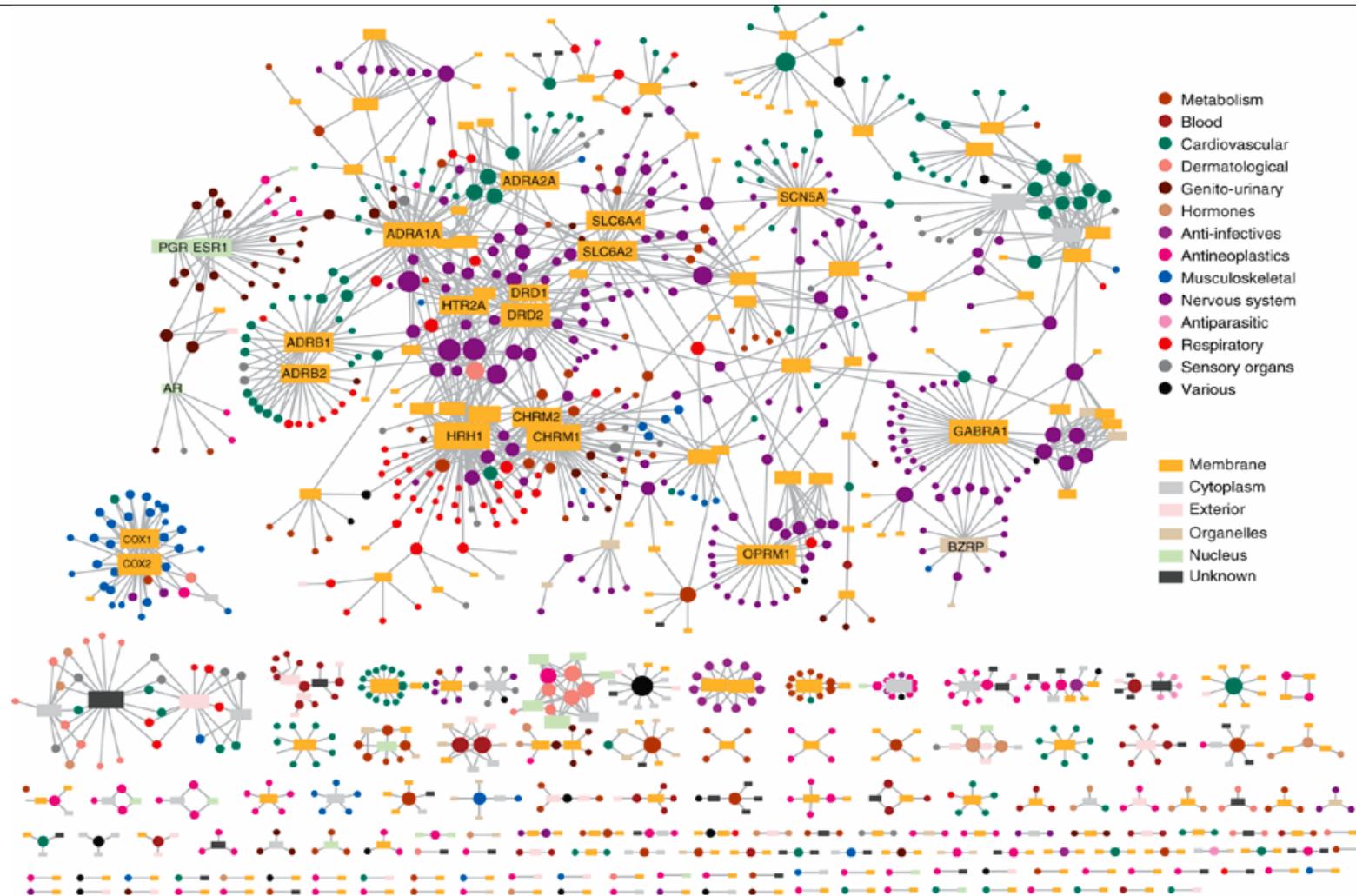
- ✎ Probabilistic Flow Analysis
- ✎ Markov Clustering

Meta analysis (cont ...)



S. E. Calvano *et al.*, *Nature Letter*, 2007.

Meta analysis (cont ...)



Y. Muhammed *et al.*, *Nature Biotechnology*, 2007.

Conclusions

- ✍ Inference of **high coverage biochemical networks** are important to find the interrelationships between drug targets and disease-gene products.
- ✍ Once the networks are inferred **meta analysis methods** are applied to infer important pathways (hub genes/protein networks).
- ✍ The interaction strengths are fed back to knowledge representation model to increase the likelihood for the selection of high quality interactions.

Acknowledgements

Victor Farutin

Chris Bouton

Mike Schaffer

Fred Jerva



Pfizer Research and Technology Centre,
Cambridge, Massachusetts, US



Kevin Burrage

Mark Ragan

Melissa Davis

IMB Institute for Molecular Bioscience



Australian Government
Australian Research Council

ARC Centre of Excellence in Bioinformatics

Visible Cell Team

Members of Group Ragan at IMB

Jane Hunter

Andrew Newman

Imran Khan

Yuan-Fang Li



Q & A

