

# Accelerating Innovation in Multidisciplinary Scientific Teams through Visual Analytics

Suresh K. Bhavnani PhD<sup>1</sup>, Shyam Visweswaran MD PhD<sup>2</sup>, Rohit Divekar MBBS PhD<sup>3</sup>

<sup>1</sup>Institute for Translational Sciences, University of Texas Medical Branch; <sup>2</sup>Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA; <sup>3</sup>Division of Allergic Diseases, Mayo Clinic, Rochester, MN

# Overview

- **Motivation**
  - Studying the Role of Visual Analytics in Enabling Innovation in Multidisciplinary Scientific Teams
- **Method**
  - Team Selection
  - Data Collection
  - Analytical Method
- **Results**
  - Evolution of the Visual Analytical Representation
  - Team Dynamics Associated with the Evolution
  - Team Outcomes
- **Conclusions and Future Research**

# Motivation

- **Explosion of Multi-Omics Biomedical Data**

- *All of Us* initiative is collecting de-identified data for 1,000,000 patients (Kaiser, 2015; Collins, 2015)
- Such data will contain multi-omics information for researchers
  - Molecular:** Genomic, proteomic, metabolomic, ...
  - Clinical:** Demographics, ER, ...
  - Environmental:** Pollutants, climate change, ...
  - Personal:** Health monitors, ...
  - Social:** Online support groups, ...

- **Need for Multidisciplinary Scientific Teams**

- Biologists
- Clinicians
- Bioinformaticians
- Statisticians
- Patient Stakeholders



- Identify disease mechanisms based on an integrated understanding of multi-omics data
- Translate discoveries into interventions

- **Well-Known Hurdles in Integrating Diverse Knowledge in Teams**

- Tension between disciplinary **heterogeneity** and disciplinary **collaboration** (Star and Briesemer, 1989)
- Difficulty in integration of disciplinary knowledge towards common goal (Salazar et al., 2012)

# Interventions Proposed to Integrate Diverse Knowledge

## Integrative Capacity Framework (Salazar et al., 2012)

- Multidisciplinary scientific teams have many barriers in integrating their knowledge
- Teams should progressively develop “integrative capacity” leading to shared identity and trust
- Interventions to develop integrative capacity:
  - **All Team Members:** Develop conflict and affect management skills
  - **Cross-Disciplinary Members:** Help bridge disciplinary differences
  - **Team Leaders:**
    - Minimize power and status differences
    - Provide shared goals, and common problem conceptualization
    - Use “**boundary objects**” which facilitate a collective understanding of diverse disciplinary knowledge while also being meaningful to each discipline

# Role of Boundary Objects in Teams

- **Boundary Object**

- An object that lives in multiple social worlds, and which has different identities in each (Star and Griesemer, 1989)
- Acts as anchors or bridges
- Example: Organization chart which shows the big picture, and how each is related to the rest

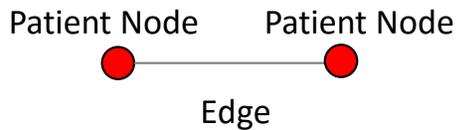


# Role of Visual Analytics in Integrating Diverse Data

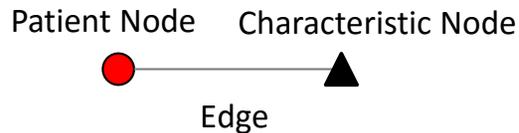
- **Visual Analytics**
  - The science of analytical reasoning facilitated by interactive visual interfaces (Thomas & Cook, 2004)
- **Classification** (Heer et al., 2010)
  - Maps (map of a country)
  - Hierarchies (organization chart)
  - Statistical Distributions and Charts (bar graph)
  - Time Series (longitudinal graphs)
  - **Networks** (network of chemicals and symptoms)

# Networks for Representing Multi-Omics Data

A network is a set of nodes connected in pairs by edges



## A. Unipartite Network

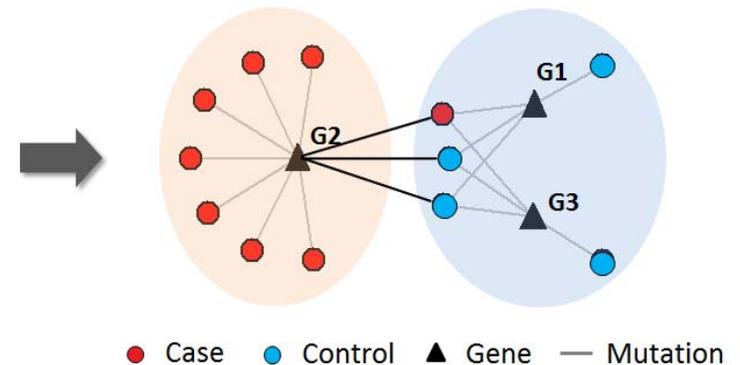


## B. Bipartite Network

Characteristics Outcomes

Patients	G1 G2 G3			Phenotype
P1		X		Case
P2	X	X	X	Control
P3		X		Case
P4		X		Case
P5		X		Case
P6	X			Control
P7	X	X	X	Control
P8		X		Case
P9		X		Case
P10			X	Control
P11	X	X	X	Control
P12		X		Case

Patient-Characteristic-Outcome Network



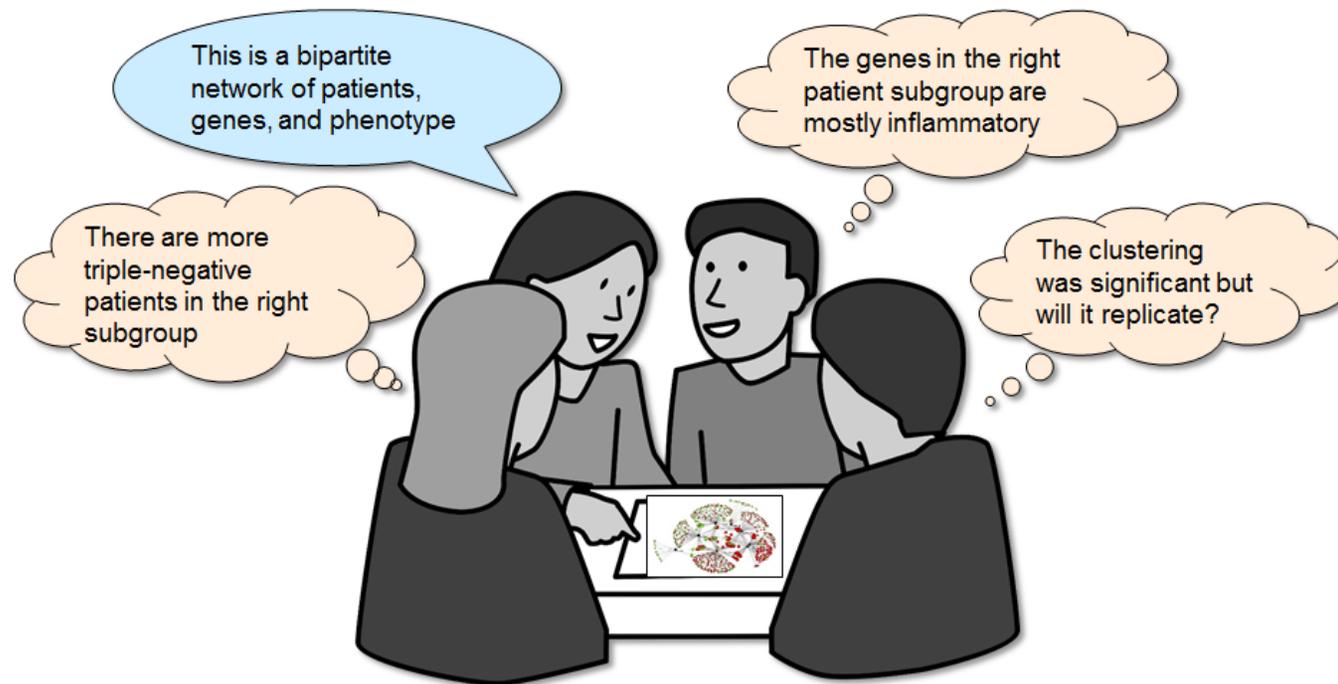
## Intra- and Inter-cluster Associations

**Translational Science:** Integrated understanding of molecular, and clinical data

**Precision Medicine:** Identification of patient subgroups and disease mechanisms leading to the design of targeted interventions

# Research Question

- What is the role of visual analytics in enabling multidisciplinary teams make translational discoveries in clinical and molecular data?



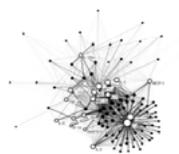
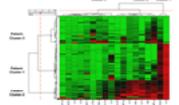
# Method

- **Criteria for Team Selection**
  - Possession of a dataset with clinical and molecular information
  - Goal of identifying biological mechanisms in disease and designing targeted interventions
  - Prior publications of analyzing the data
- **Intervention**
  - Visual analytical method introduced by a biomedical informatician in the team to enable identification of biological mechanisms in data
- **Data Collection and Analysis**
  - Changes in data representation (how, why, who)
  - Team Dynamics (roles)
  - Outcomes (novel discoveries, and publications)

# Results: Team Selection

- Asthma Multidisciplinary Scientific Team
  - **Members:** Physician, Molecular Biologist, Statistician, Programmer, Biomedical Informatician
  - **Data:** 108 Asthma Patients, 18 Cytokines, 6 Clinical Variables
  - **Goal:** Identify biological mechanisms that integrate molecular and clinical information
  - **Prior Publications:** Use of Conventional Biostatistics

# Results: 4 Stages of Representational Shifts

	Data Representation			Team Dynamics		Outcomes																																																									
	Motivation for Change	Method of Change	State	Primary Role	Secondary Role	Biomedical Insights	Methodological Insights	Publications & Grants																																																							
Data Preparation	A Bipartite networks need a uniform quantitative and interpretive range for all values to enable comparisons and clustering	Used min-max range normalization method of raw cytokine values	<table border="1"> <thead> <tr> <th></th> <th>IL-1Ra</th> <th>IL-2</th> <th>IL-2R</th> <th>IL-4</th> </tr> </thead> <tbody> <tr> <td>00017</td> <td>0.537</td> <td>0.004</td> <td>0.081</td> <td>0.062</td> </tr> <tr> <td>00007</td> <td>0.503</td> <td>0.185</td> <td>0.012</td> <td>0.048</td> </tr> <tr> <td>00010</td> <td>0.817</td> <td>0.000</td> <td>0.011</td> <td>0.009</td> </tr> <tr> <td>00026</td> <td>0.221</td> <td>0.967</td> <td>0.026</td> <td>0.967</td> </tr> <tr> <td>00019</td> <td>0.362</td> <td>0.009</td> <td>0.000</td> <td>0.066</td> </tr> <tr> <td>00005</td> <td>0.409</td> <td>0.084</td> <td>0.950</td> <td>0.009</td> </tr> <tr> <td>00077</td> <td>0.722</td> <td>0.199</td> <td>0.062</td> <td>0.116</td> </tr> <tr> <td>00084</td> <td>0.268</td> <td>0.209</td> <td>0.009</td> <td>0.205</td> </tr> <tr> <td>00066</td> <td>0.761</td> <td>0.242</td> <td>0.111</td> <td>0.162</td> </tr> <tr> <td>00100</td> <td>0.778</td> <td>0.017</td> <td>0.195</td> <td>0.048</td> </tr> </tbody> </table>		IL-1Ra	IL-2	IL-2R	IL-4	00017	0.537	0.004	0.081	0.062	00007	0.503	0.185	0.012	0.048	00010	0.817	0.000	0.011	0.009	00026	0.221	0.967	0.026	0.967	00019	0.362	0.009	0.000	0.066	00005	0.409	0.084	0.950	0.009	00077	0.722	0.199	0.062	0.116	00084	0.268	0.209	0.009	0.205	00066	0.761	0.242	0.111	0.162	00100	0.778	0.017	0.195	0.048	Visual Analyst	Programmer			
	IL-1Ra	IL-2	IL-2R	IL-4																																																											
00017	0.537	0.004	0.081	0.062																																																											
00007	0.503	0.185	0.012	0.048																																																											
00010	0.817	0.000	0.011	0.009																																																											
00026	0.221	0.967	0.026	0.967																																																											
00019	0.362	0.009	0.000	0.066																																																											
00005	0.409	0.084	0.950	0.009																																																											
00077	0.722	0.199	0.062	0.116																																																											
00084	0.268	0.209	0.009	0.205																																																											
00066	0.761	0.242	0.111	0.162																																																											
00100	0.778	0.017	0.195	0.048																																																											
Visual Representation and Discovery	B Map symbolic data to visual representation	Used Pajek network visualization application to map patients and cytokines to nodes, and normalized cytokine expression to edges through a bipartite network representation. Node size represented sum of cytokine expression, and edge thickness represented degree of cytokine expression		Visual Analyst	Physician, Biologist		Circular layout of nodes does not make salient similarity and differences among nodes																																																								
	C Lay out the network to reveal similarity and differences among nodes	Used force-directed layout algorithm to lay out nodes in a 2D plane		Visual Analyst, Statistician	Physician, Biologist	Existence of patient subgroups based on cytokine profiles	Statistical questions min-max normalization method, but finds other normalization methods produce comparable results. States that while min-max is useful for networks, it is vulnerable to new data if it changes the minimum and maximum values; visual analyst agrees																																																								
Quantitative Verification	D Quantitatively identify number and boundary of clusters to enable comprehension of relationships	Used hierarchical clustering and heatmaps of patients and cytokines to identify cluster boundaries; compared clustering with 1000 random permutations to determine statistical significance of patient and cytokine clusterings		Visual Analyst, Programmer	Physician, Biologist		Programmer questions linkage method used for hierarchical clustering based on past experience, but agrees to proceed as there are no guidelines on selecting the best linkage method																																																								
	E Understand how patient and cytokine cluster boundaries relate to network layout	Used Pajek application to color cytokines nodes based on cluster membership, and shapes in PowerPoint to encircle patient clusters		Visual Analyst, Physician, Biologist		Cytokine clusters are meaningful, but to fully understand the biological mechanisms there is a need to include how the patient clusters differ in clinical variables	Too many colored nodes can be confusing so select shapes instead of node color to denote patient clusters																																																								
	F Identify the clinical variables that are statistically significant across the patient clusters	Used Kruskal-Wallis with FDR correction to identify statistically significant clinical variables across clusters	<table border="1"> <thead> <tr> <th>Pulmonary Function</th> <th>p value with FDR correction</th> </tr> </thead> <tbody> <tr> <td>Max FEV<sub>0pp</sub>/MPV/Lung</td> <td>0.008</td> </tr> <tr> <td>Max FEV<sub>0pp</sub>/MPV/Lung</td> <td>0.0375*</td> </tr> <tr> <td>Baseline FEV<sub>0p</sub></td> <td>0.0375*</td> </tr> <tr> <td>Baseline FEV<sub>0</sub>/FVC</td> <td>0.1944</td> </tr> <tr> <td>Max FEV<sub>0</sub> Reversal</td> <td>0.583</td> </tr> <tr> <td>PC<sub>50</sub> Methacholine</td> <td>0.0375*</td> </tr> </tbody> </table>	Pulmonary Function	p value with FDR correction	Max FEV <sub>0pp</sub> /MPV/Lung	0.008	Max FEV <sub>0pp</sub> /MPV/Lung	0.0375*	Baseline FEV <sub>0p</sub>	0.0375*	Baseline FEV <sub>0</sub> /FVC	0.1944	Max FEV <sub>0</sub> Reversal	0.583	PC <sub>50</sub> Methacholine	0.0375*	Visual Analyst, Statistician,	Programmer																																												
Pulmonary Function	p value with FDR correction																																																														
Max FEV <sub>0pp</sub> /MPV/Lung	0.008																																																														
Max FEV <sub>0pp</sub> /MPV/Lung	0.0375*																																																														
Baseline FEV <sub>0p</sub>	0.0375*																																																														
Baseline FEV <sub>0</sub> /FVC	0.1944																																																														
Max FEV <sub>0</sub> Reversal	0.583																																																														
PC <sub>50</sub> Methacholine	0.0375*																																																														
Inference of Patterns	G Comprehend how statistically significant clinical variables relate to clusters	Used Pajek application to generate a series of networks each with a different statistically significant clinical variable used to color the patient nodes; used IPA to identify known pathways associated with the cytokine clusters		Biologist, Physician, Visual Analyst	Statistician, Programmer	Identification of three different biological pathways corresponding to each patient subgroup; proposed targeted treatments for each patient subgroup		2 peer-reviewed domain publication, and 2 peer-reviewed methodological publications; included in CTSA proposal which was funded																																																							

# Results: Stage-1

## 1. Data Preparation (data cleanup and normalization)

Raw Data



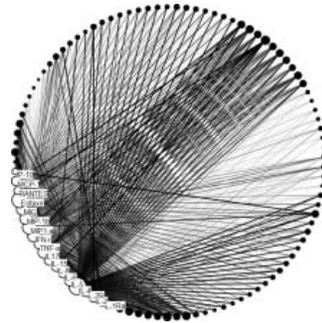
	IL-1Ra	IL-2	IL-2R	IL-4
0081F	0.537	0.094	0.085	0.052
0000P	0.921	0.165	0.012	0.098
0001D	0.812	0.000	0.031	0.059
0002R	0.223	0.967	0.026	0.967
0003F	0.392	0.089	0.000	0.059
0006V	0.459	0.084	0.950	0.059
0007J	0.722	0.199	0.062	0.116
0008X	0.266	0.209	0.059	0.205
0009L	0.751	0.242	0.111	0.322
0010D	0.779	0.017	0.135	0.048

Primary Role: Informatician  
Supporting Role: Programmer

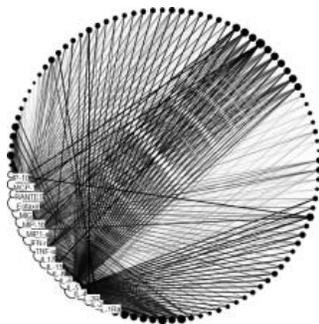
# Results: Stage-2

## 2. Visual Representation and Discovery

	IL-1Ra	IL-2	IL-2R	IL-4
0081F	0.537	0.094	0.085	0.052
0000P	0.921	0.165	0.012	0.098
0001D	0.812	0.000	0.031	0.059
0002R	0.223	0.967	0.026	0.967
0003F	0.392	0.089	0.000	0.059
0006V	0.459	0.084	0.950	0.059
0007J	0.722	0.199	0.062	0.116
0008X	0.266	0.209	0.059	0.205
0009L	0.751	0.242	0.111	0.322
0010D	0.779	0.017	0.135	0.048



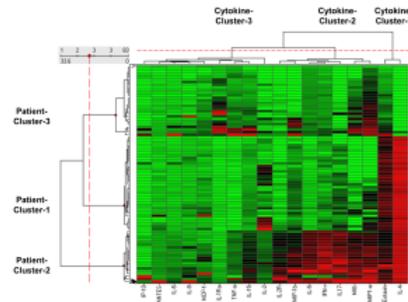
- **Primary Role:** Informatician
- **Supporting Role:** Physician, Biologist
- **Insight:** Circular Layout not Useful



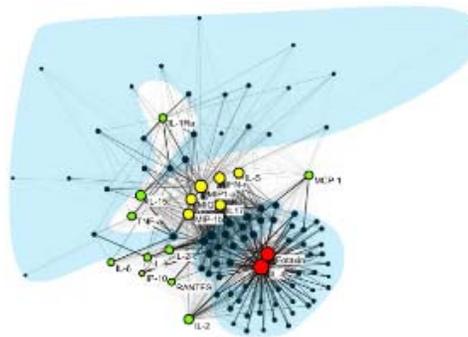
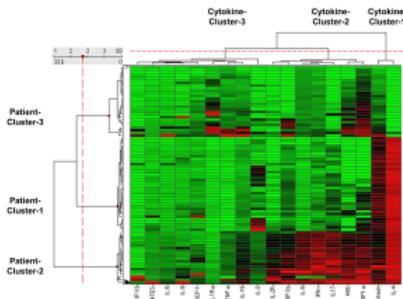
- **Primary Role:** Informatician, statistician (disagrees with normalization method, reaches consensus that it is equivalent)
- **Supporting Role:** Physician, Biologist
- **Insight:** Existence of Patient Subgroups warrants quantitative verification

# Results: Stage-3

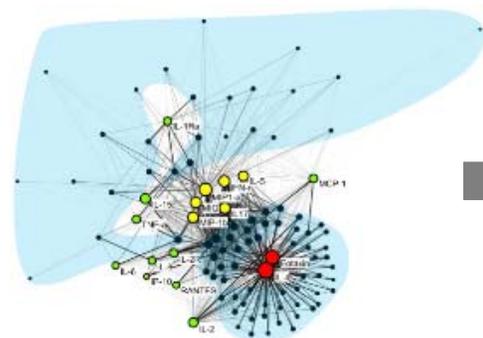
## 2. Quantitative Verification



- **Primary Role:** Informatician, programmer
- **Supporting Role:** Physician, Biologist
- **Insight:** Programmer questions cluster linkage function, but agrees there are no objective guidelines for selection. Heatmap not as useful as network for visualizing clusters



- Primary Role:** Informatician, Physician, Biologist
- Supporting Role:** Statistician
- Insight:** Cluster boundaries need node color and shapes to be comprehensible; Clinical variables of each patient subgroup



Pulmonary Function	p value with FDR correction
Max FVC <sub>pp</sub> /MPV <sub>Lung</sub>	0.006*
Max FEV <sub>1,pp</sub> /MPV <sub>Lung</sub>	0.0375*
Baseline FEV <sub>1,pp</sub>	0.0375*
Baseline FEV <sub>1</sub> /FVC	0.1944
Max FEV <sub>1</sub> Reversal	0.583
PC <sub>20</sub> Methacholine	0.0375*

- Primary Role:** Informatician, Statistician
- Supporting Role:** Programmer
- Insight:** Clinical variables need to be layered on network

# Results: Stage-4

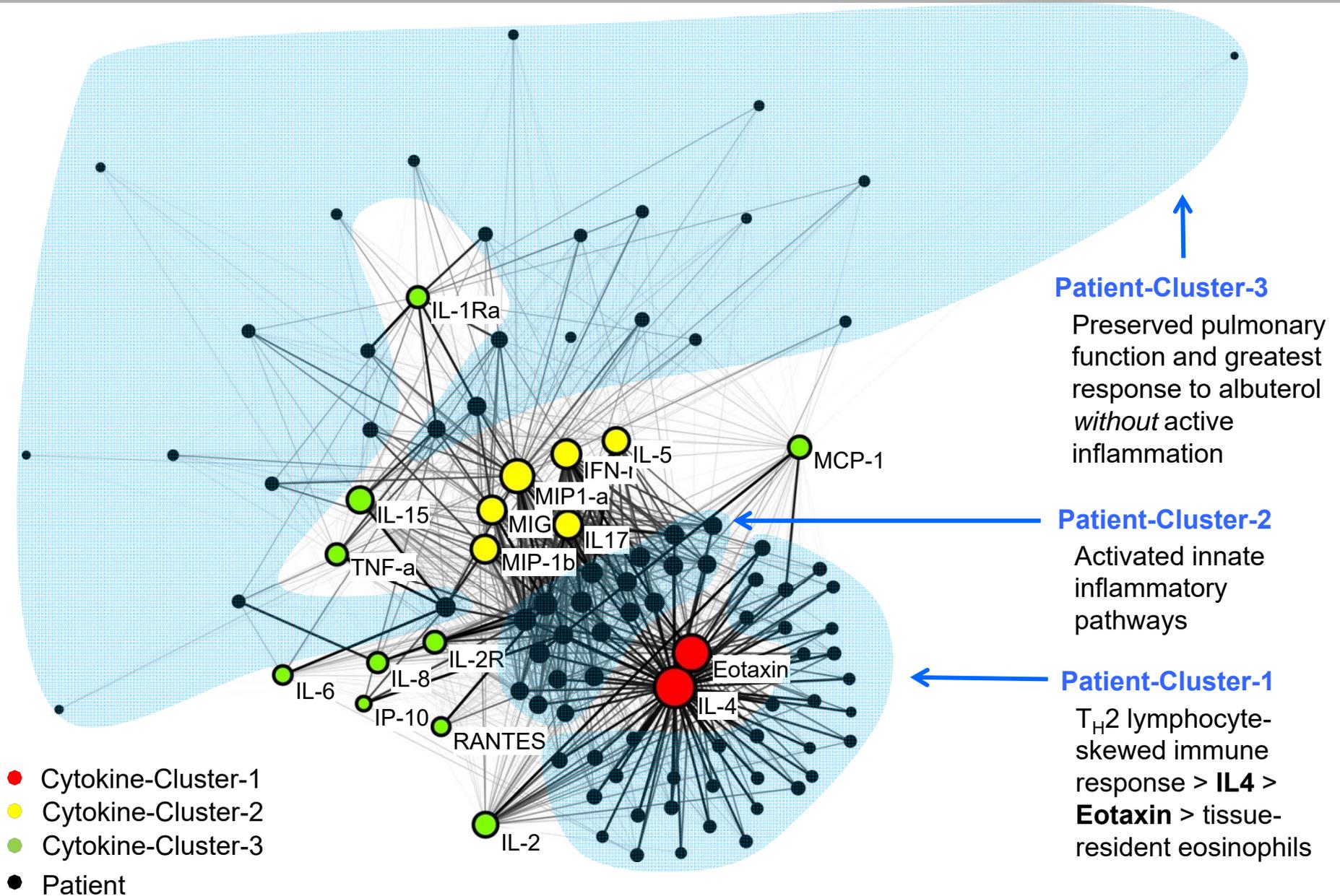
## 4. Inference of Patterns

Pulmonary Function	p value with FDR correction
Max FVC <sub>pp</sub> /MPV <sub>Lung</sub>	0.006*
Max FEV <sub>1pp</sub> /MPV <sub>Lung</sub>	0.0375*
Baseline FEV <sub>1pp</sub>	0.0375*
Baseline FEV <sub>1</sub> /FVC	0.1944
Max FEV <sub>1</sub> Reversal	0.583
PC <sub>20</sub> Methacholine	0.0375*



- **Primary Role:** Biologist, Physician, Informatician
- **Supporting Role:** Statistician, Programmer
- **Insight:** Molecular Pathways for each patient subgroup

# Inference of Biological Mechanisms

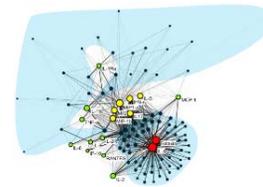
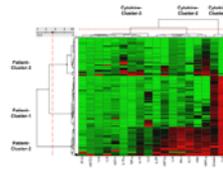


Two informatics publications and two domain publications

# Summary of Results

## 1. Evolving Data Representations

	H_1Ba	H_2	H_2R	H_4
0081F	0.537	0.094	0.085	0.052
0000P	0.921	0.165	0.012	0.098
0001D	0.812	0.000	0.031	0.059
0002R	0.223	0.987	0.026	0.967
0001B	0.392	0.089	0.000	0.099
0006V	0.459	0.084	0.950	0.059
0007J	0.722	0.199	0.062	0.116
0008X	0.266	0.209	0.059	0.205
0009L	0.751	0.242	0.111	0.322
0010D	0.779	0.017	0.135	0.048



Pulmonary Function	p value with FDR correction
Max FVCpp/MPVLung	0.006*
Max FEV,pp/MPVLung	0.0375*
Baseline FEV,pp	0.0375*
Baseline FEV/FVC	0.1944
Max FEV, Reversal	0.583
PC <sub>20</sub> Methacholine	0.0375*



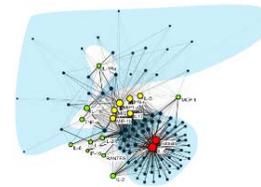
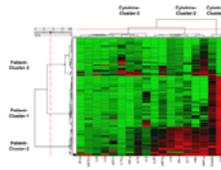
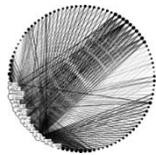
- Implications

- Visual analytical methodology produced data representations which was a boundary object enabling multidisciplinary insights
- Boundary object evolved based on multidisciplinary engagement and insights
- Data representations can be conceptualized as a *Computationally Evolving Boundary Object* (CEBO)

# Summary of Results

## 2. Shifting Primary and Supporting Roles

	H_1Ba	H_2	H_2H	H_4
0081f	0.537	0.094	0.085	0.052
0000P	0.921	0.165	0.012	0.098
0001D	0.812	0.000	0.031	0.059
0002R	0.223	0.987	0.026	0.967
0003B	0.392	0.089	0.000	0.099
0006V	0.459	0.084	0.950	0.059
0007J	0.722	0.199	0.062	0.116
0008X	0.266	0.209	0.059	0.205
0009L	0.751	0.242	0.111	0.322
0010D	0.779	0.017	0.135	0.048



Pulmonary Function	p value with FDR correction
Max FVCpp/MPVLung	0.006*
Max FEV <sub>pp</sub> /MPVLung	0.0375*
Baseline FEV <sub>pp</sub>	0.0375*
Baseline FEV/FVC	0.1944
Max FEV, Reversal	0.583
PC <sub>20</sub> Methacholine	0.0375*

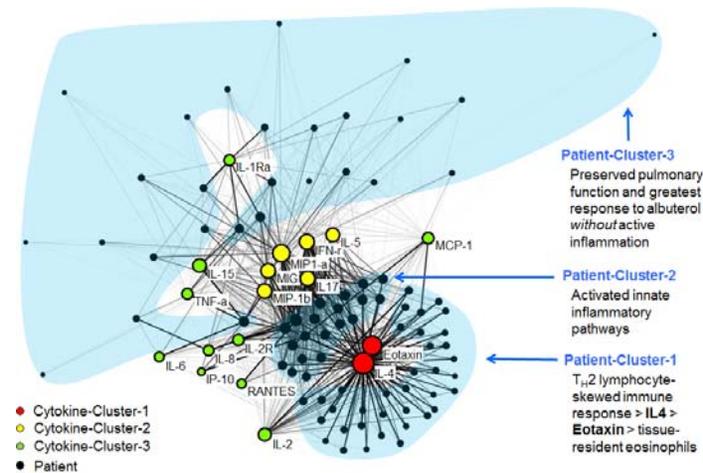


- Implication

- Visual Analytical methodology appeared to enable team members to play primary and supporting roles at different analytical stages, resulting in an egalitarian team dynamic

# Summary of Results

## 3. Novel Translational Insight



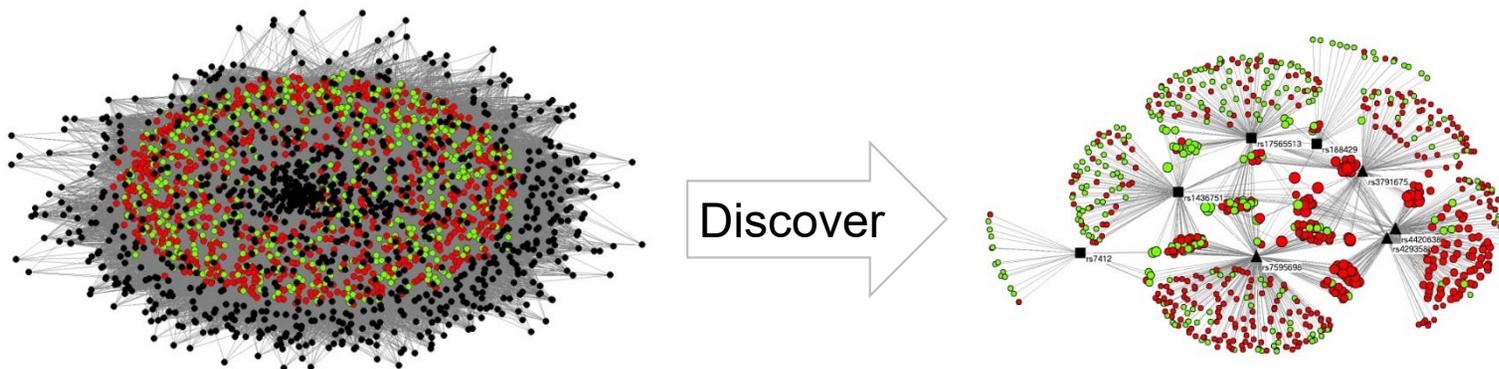
Two informatics publications and two domain publications; inclusion in CTSA proposal that was funded

- **Implication**

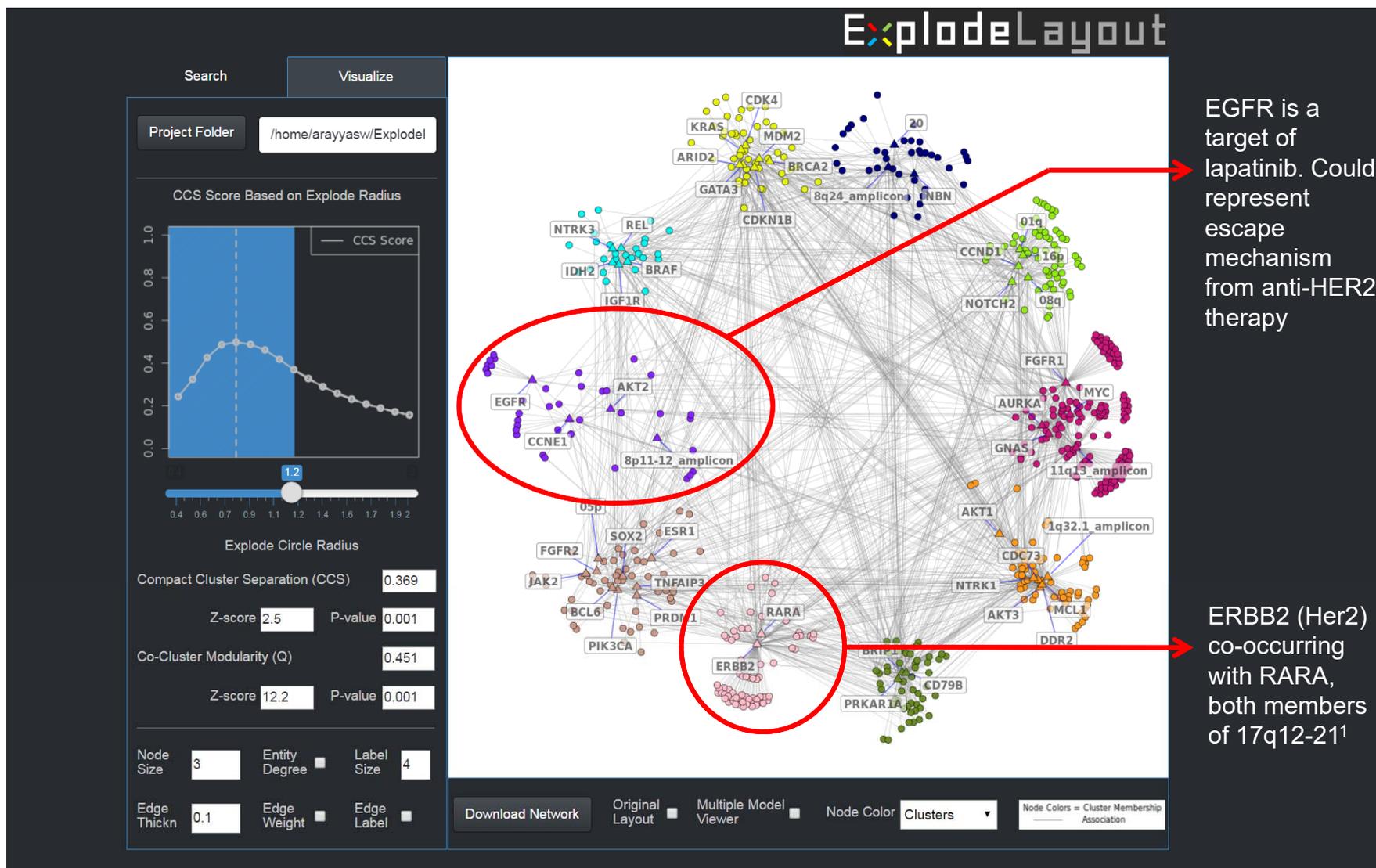
- Visual Analytical methodology appeared to transform the typical medical team structure, into an egalitarian team structure
- Egalitarian team structure enabled parity in the domain and analytical contributions

# Conclusions and Future Research

- **Motivation**
  - Multi-omics data require multidisciplinary scientific teams to enable translational discoveries
  - However such teams face significant hurdles to integrate diverse knowledge
- **Role of Visual Analytics**
  - Bipartite networks can model multiple datatypes
  - Visual analytical representations can play the role of a **computationally evolving boundary object** (CEBO)
  - Outcome included novel translational insights published in informatics and domain journals
- **Future Research**
  - CEBOs are just one instantiation of “Team-Centered Informatics” which requires further investigation and evaluation
  - Large datasets require more advanced algorithms and tools to enable success as computationally evolving boundary objects



# Breast Cancer Patients and Amplification Events



Results Published: Madhavan, S., Ratwani, R., Warner, J., Bhavnani, S.K. Making Precision Oncology Data More Usable for Research and Care. *Proceedings of AMIA Summit on Translational Bioinformatics* (2017).

# Use of ExplodeLayout in a Cancer Team

- **Cancer Team**
  - Oncologist
  - Biologist
  - Statistician
  - Biomedical Informatician
- **Data**
  - Goal: Analyze heterogeneity in breast cancer patient specimens based on SCNAs
  - Extracted 894 metastatic breast cancer specimens
  - Analyzed 50 most-frequent SCNAs after aggregation
- **Analytical Method**
  - Represented the data as a bipartite network
  - Identified clusters of specimens and SCNAs using co-cluster modularity
  - Visualized the network using ExplodeLayout
  - Interpreted the breast cancer heterogeneity suggested by the clusters