Article

The Journal of Applied Behavioral Science
2019, Vol. 55(1) 50–72

DOI: 10.1177/0021886318794606
journals.sagepub.com/home/jab
SAGE

Towards Team-Centered Informatics: Accelerating Innovation in Multidisciplinary Scientific Teams Through Visual Analytics

Suresh K. Bhavnani[1], Shyam Visweswaran[2], Rohit Divekar[3], and Allan R. Brasier[4]

Abstract

A critical goal of multidisciplinary scientific teams is to integrate knowledge from diverse disciplines for the purpose of developing novel insights and innovations. For example, multidisciplinary translational teams (MTTs) which typically include physicians, biologists, statisticians, and informaticians, aim to integrate biological and clinical knowledge leading to innovations for improving health outcomes. However, such teams face numerous barriers in integrating multidisciplinary knowledge, which is further exacerbated by the explosion of molecular and clinical data generated from millions of patients. Here, we explore the use of a visual analytical representation to help MTTs integrate molecular and clinical data with the goal of accelerating translational insights. The results suggest that the visual analytical representation functioned as a "computational evolving boundary object" which (a) evolved through several emergent states that progressively helped integrate diverse disciplinary knowledge, (b) enabled team members to play primary and supportive roles in evolving the data representation resulting in a more egalitarian team structure, and (c) enabled the team to arrive at novel translational insights leading to domain and methodology publications. However, the interventions also revealed limitations in the approach motivating new visual analytical approaches. These results suggest

[1]The University of Texas Medical Branch, Galveston, TX, USA
[2]University of Pittsburgh, Pittsburgh, PA, USA
[3]Mayo Clinic, Rochester, MN, USA
[4]University of Wisconsin, Madison, WI, USA

Corresponding Author:
Suresh K. Bhavnani, The University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555, USA.
Email: skbhavnani@gmail.com

(a) implications for theory related to modeling computational evolving boundary objects (CEBOs) as an instance of team-centered informatics, and (b) implications for practice related to the design and use of interactive features that enable teams to fluidly evolve CEBOs through emergent states, with the goal of deriving novel insights from large multiomics datasets.

Keywords
visual analytics, computational evolving boundary objects, biomedical insights, multidisciplinary scientific teams, team-centered informatics

Introduction

Numerous organizations worldwide are investing huge resources in assembling multidisciplinary scientific teams to address complex real-world problems (Klein, 2010; Trochim, Marcus, Masse, Moser, & Weld, 2008). This realization is motivated by several studies which show that multidisciplinary scientific teams have many advantages compared to researchers working alone including increased research productivity (Hall et al., 2012), faster and broader dissemination of results across disciplinary boundaries (Stipelman et al., 2014), and generation of highly significant outcomes and practical applications (Hage & Meeus, 2009).

However, scientific teams increasingly face significant barriers in generating innovative solutions to complex problems. For example, multidisciplinary translational teams (MTTs), a specific form of multidisciplinary scientific teams that typically include physicians, biologists, statisticians, and informaticians, face the challenge of integrating their biological and clinical knowledge with the goal of translating the results into innovations to improve health outcomes (Ameredes et al., 2015; Wooten et al., 2015). Furthermore, the challenge of integrating information from diverse disciplines is exacerbated by the explosion of multiomics data including genomic, proteomic, metabolomic, phenomic, and social data collected from millions of patients (Bietz et al., 2016).

Fortunately, a growing understanding of the social and psychological antecedents of effective teams (Falk-Krzesinski et al., 2010; Stokols, Hall, Taylor, & Moser, 2008), combined with powerful computational visual analytical methods (Thomas & Cook, 2005) designed for helping researchers integrate diverse types of knowledge, provide new opportunities for empowering multidisciplinary scientific teams. Here, we explore the role that visual analytics can play in enabling MTTs to integrate diverse information from large biomedical data sets, with the goal of generating novel insights and innovations for improving health outcomes.

We begin by discussing the barriers faced by multidisciplinary scientific teams in integrating knowledge from multiple disciplines, and how the "integrative capacity" framework provides a starting point for exploring potential interventions. Next, we discuss the cognitive motivations underlying the field of visual analytics, and its role in helping humans integrate complex information. We then present a case study where we designed and used a visual analytical method as an intervention in an MTT to help

integrate molecular and clinical information. We conclude by discussing the strengths and limitations of the intervention, and the theoretical and practical implications of the results for helping scientific teams integrate diverse disciplinary knowledge with the goal of generating insights and innovative solutions.

Background

Barriers to Integrating Diverse Knowledge in Multidisciplinary Science Teams

Multidisciplinary science teams consist of specialists from different disciplines working toward a common goal and exist in virtually every sphere of science (Falk-Krzesinski et al., 2010). In contrast to interdisciplinary and transdisciplinary science teams whose goal is to create knowledge that extends beyond the scope of any one discipline, multidisciplinary science teams have the goal of integrating knowledge from diverse sources to solve challenging scientific problems (Klein, 2010). For example, MTTs, a specific form of multidisciplinary scientific team typically consist of physicians, biologists, statisticians, and informaticians who attempt to integrate knowledge from each of their disciplines with the goal of translating that knowledge into innovations to improve health outcomes (Ameredes et al., 2015; Wooten et al., 2015).

Unfortunately, while multidisciplinary scientific teams are critical for addressing complex real-world problems, they encounter numerous barriers leading to missed opportunities for generating novel insights. These barriers include (a) narrow social identification which makes scientists strongly identify with values, methods, and norms of their own discipline, resulting in evaluating members within their own discipline more favorably compared to those outside (Hewstone, Rubin, & Willis, 2002); (b) disagreement on team goals emanating from strong discipline identification such as basic scientists favoring theoretical advances compared to applied scientists favoring practical applications (Bunderson & Sutcliffe, 2002); (c) differing conceptualization of a problem leading to little overlap in methods of analysis and interpretation (Lélé & Norgaard, 2005); (d) too wide a breadth of knowledge across the team members with an inability to leverage it to solve a common problem (Dougherty, 1992; Liyanage & Barnard, 2003); (e) physical and temporal differences in workspaces resulting in poor coordination of tasks best achieved in collocated settings (Olson & Olson, 2000); (f) too dense or too sparse social and informational networks resulting in either being closed off to new ideas, or not having enough alignment with ideas in a discipline (Perry-Smith, 2006); and (g) little or too much familiarity between team members resulting in either reduced cohesion and trust, or formation of closed cliques of friends (Stokols, Harvey, Gress, Fuqua, & Phillips, 2005) that miss opportunities for cross-fertilization of ideas from those outside their clique.

While many of these barriers have been analyzed separately with specific interventions to address them, the concept of "integrative capacity" has been proposed as a conceptual framework to unify the barriers and respective interventions (Salazar, Lant,

Fiore, & Salas, 2012). This framework proposes that specific interventions can be progressively introduced in a team to improve integrative capacity resulting in emergent states such as shared team identity and trust. This coevolution of knowledge and integration has the potential of overcoming integration barriers and precipitating cutting-edge insights and innovations.

The integrative capacity framework further suggests specific interventions in the form of propositions to address the barriers. For example, team leaders could help diverse team members to identify a shared goal, develop a common problem conceptualization, and minimize power and status differences. Furthermore, leaders could generate "boundary objects" defined as "an object which lives in multiple social worlds and which has different identities in each," and which "act as anchors or bridges" for the team members (Star & Griesemer, 1989, pp. 409, 414). Such objects could facilitate a collective understanding of diverse disciplinary knowledge. For example, similar to an organizational chart which shows how each employee fits into an overall organization, a boundary object in a scientific team is designed to enable team members integrate their disciplinary knowledge into the shared team knowledge. Finally, team members who belong to more than one disciplinary or professional group could help bridge social networks within subgroups in the team (Gray, 2008), in addition to developing conflict and affect management skills to foster greater trust and collaboration (Klein, Knight, Ziegert, Lim, & Saltz, 2011). The authors emphasize that improving integrative capacity does not depend on any individual team member, but rather on the interaction of the social, psychological, and cognitive systems across all members in the team. While each of these propositions are based on empirical studies in the literature, the authors encourage further empirical investigations of the interrelationships between the proposed antecedents of improving integrative capacity, and their effect on knowledge integration in teams (Salazar et al., 2012).

The Role of Visual Analytics in Enabling Knowledge Integration and Translational Insights

While the integrative capacity framework provides broad guidelines for interventions to help improve integration in multidisciplinary teams, MTTs have the specific problem of integrating large and diverse data sets consisting of molecular and clinical information (Wooten et al., 2015). One approach that has shown promise in helping humans integrate such large and complex data sets are computational methods from the rapidly maturing field of visual analytics. Visual analytics is defined as the science of analytical reasoning, facilitated by interactive visual interfaces (Thomas & Cook, 2005). The primary goal of visual analytics is to augment cognitive reasoning by translating symbolic data (e.g., numbers in a spreadsheet) into visualizations (e.g., a scatter plot), which can be manipulated through interaction (e.g., highlight only some data points in the scatter plot).

Visualizations are powerful because they leverage the massively parallel architecture of the human visual system consisting of the eye and the visual cortex of the brain (Card, Mackinlay, & Shneiderman, 1999). This parallel cognitive architecture enables

the rapid comprehension of multiple graphical relationships simultaneously, which often leads to insights about relationships in complex data such as similarities, trends, and anomalies (Thomas & Cook, 2005). For example, the detection of an outlier in a scatter plot is fast because the graphical relationships between the outlier and the rest of the points can be processed in parallel by the visual cortex. Such parallel processing is independent of the number of nonoutlying points and therefore scales up well to large amounts of data. In contrast, finding an outlier in a spreadsheet of numbers involves numerical comparisons to identify the outlier, which is dependent on the much slower symbolic processing areas of the human brain. Such symbolic processing is serial in nature, and therefore highly dependent on the number of data points, which when large can quickly overwhelm an analyst. Data visualizations therefore help to shift processing from the slower symbolic processing areas of the human brain, to the faster graphical parallel processing of the visual cortex enabling comprehension of large and complex data sets such as those currently available for complex diseases such as asthma and Alzheimer's disease (Card et al., 1999). Furthermore, visual analytic representations such as a bar chart shift information from an internal to an external representation reducing working memory loads and therefore enable the analyst to use the freed up resources for deriving critical insights from the data (Zhang & Norman, 1994).

While static visualizations of data can be powerful if they are aligned with tasks, data, and mental representations, they are often not sufficient for comprehending complex data. This is because data analysis typically requires many different tasks performed on the same data such as discovery, inspection, confirmation, and explanation (Bhavnani, Bellala, Victor, Bassler, & Visweswaran, 2012), each requiring different views of the data. Furthermore, when analysis is done in teams consisting of different disciplines, each member often requires a different representation of the same data. For example, in an MTT, a molecular biologist might be interested in which cytokines are coexpressed across patients, whereas a clinician might be interested in the clinical characteristics of patients with similar cytokine profiles, and later how they integrate with the molecular information. To address these changes in task and mental representation, visualizations require interactivity or the ability to transform parts or the entire visual representation. Interactive changes enable a visual analytical representation to be iteratively changed based on the associations discovered during the analytical process. This iterative process, with a focus on facilitating reasoning and making sense of complex information individually and in groups (Thomas & Cook, 2005), distinguishes visual analytics from information visualization which has typically focused on visually presenting the results of analysis. These features make visual analytical representations especially useful for tasks such as inferring biological pathways from molecular and clinical information in MTTs.

Visual Analytics As a Boundary Object to Enable Knowledge Integration and Novel Insights

While the field of visual analytics has generated numerous representations such as interactive bar charts and interactive maps (Heer, Bostock, & Ogievetsky, 2010),

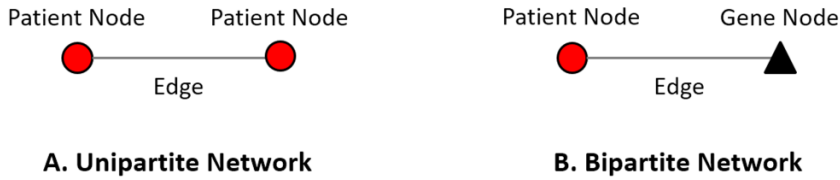


Figure 1. A unipartite network (A) has nodes of the same type connected in pairs by an edge, whereas a bipartite network (B) has nodes of two types connected in pairs by an edge.

one of the most advanced forms of visual analytical representations are networks (Newman, 2010). Networks are powerful because they not only enable an interactive visualization of complex associations, but because they are based on a graph representation, also enable quantitative analysis and validation of the patterns that are identified in the visualization. This visual and quantitative coupling therefore enables cognitive comprehension of patterns in addition to their statistical significance, both of which are critical in science.

A network (also called a graph; Newman, 2010) consists of a set of nodes, connected in pairs by edges; nodes represent one or more types of entities (e.g., patients or genes). Edges between nodes represent a specific relationship between the entities (e.g., a patient has a mutation on a specific gene). Figure 1A shows a unipartite network where nodes are of the same type (commonly used to analyze co-occurrence of genes across patients, or to analyze protein–protein interaction networks; Jensen et al., 2009). In contrast, Figure 1B shows a bipartite network where nodes are of two types, and edges exist only between different types of nodes such as between patient nodes (circles) and gene nodes (triangles), with the edges representing a relationship between them such as a patient has a mutation.

As shown in Figure 2A, the bipartite visual analytical method takes as input any data set consisting of patients and their characteristics (e.g., mutated genes), and automatically outputs a quantitative and visual description of patient subgroups (Bhavnani et al., 2015). The quantitative output provides the number, size, and statistical significance of patient subgroups and their most highly co-occurring characteristics. The visual output displays the quantitative information of the patient subgroups through a network diagram as shown in the figure.

A key advantage of a bipartite network visualization is that besides showing the number and size of the disease subgroups, it also reveals the relationships within and across disease subgroups. For example, the network visualization in Figure 2A reveals that patients in the left subgroup have a more uniform profile compared to patients in the right subgroup. Furthermore, three patients in the right subgroup share a characteristic that occurs most frequently in the left subgroup (shown by the darker edges between the subgroups), whereas none of the patients in the left subgroup share a characteristic frequently occurring in the right subgroup. Such relationships could enable clinicians and basic scientists to infer, for example, that the disease-causing mechanisms and resulting interventions in the right subgroup involve complex interactions, which could overlap

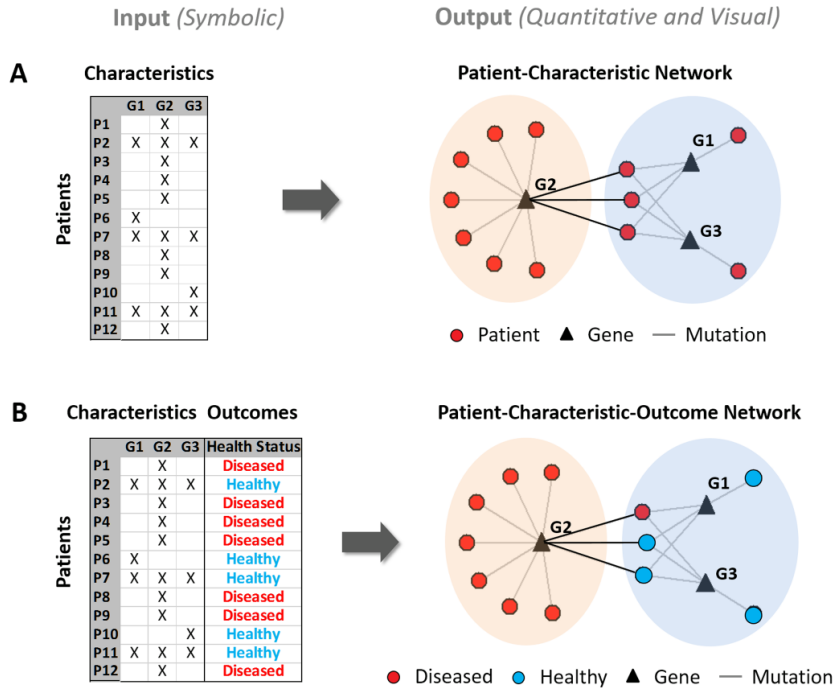


Figure 2. Examples of how a data set of patients and characteristics (e.g., genes) can be modeled as a bipartite network (A), in addition to integrating clinical information (e.g., outcomes such as diseased and healthy) through node color (B) in the same representation.

with the left subgroup. Insights such as these are difficult to derive from many computational approaches such as classification (Hastie, Tibshirani, & Friedman, 2001; Kehl & Ulm, 2006; Lipkovich, Dmitrienko, Denne, & Enas, 2011) and unipartite clustering (Abu-Jamous, Fa, & Nandi, 2015; Fitzpatrick et al., 2011) which are not designed to reveal relationships within and across patient subgroups simultaneously.

When the data contain clinical outcomes, that information can be overlaid onto the network using node color. For example, as shown in Figure 2B, the patient nodes can be colored red and blue to represent diseased and healthy, respectively. The resulting patient-characteristic-outcome network enables an integration of molecular and clinical outcomes in a unified visual-quantitative representation. This unified representation, besides externalizing complex associations that reduce working memory loads, enables each discipline to comprehend how their disciplinary knowledge relates to the whole, and therefore increases the chances of an integrated insight about the mechanisms underlying the disease. For example, physicians and biologists in an MTT could together infer that the gene mutations in the left cluster in Figure 2B increase risk of the disease because they are associated with mostly the diseased, whereas the gene mutations in the right cluster are protective as they are involved with mostly the healthy. This integration of biological and clinical knowledge could

lead to inferences related to the underlying risky and protective biological mechanisms in the disease being analyzed. While the examples show binary mutation associations and disease outcomes, bipartite networks can also model continuous associations using edge thickness, in addition to continuous disease outcomes using variations in the color of nodes.

Because the patient-characteristic-outcome network represents molecular, clinical, and statistical information, it spans the disciplinary worlds of physicians, biologists, and analysts while integrating the information in a unified representation, and therefore meets the definition of a boundary object (Star & Griesemer, 1989). Given its digital nature, we hypothesized that the patient-characteristic-outcome network would function as a *computational* boundary object to enable MTTs gain an integrated understanding of molecular and clinical information, leading to translational insights into the mechanisms of the disease and potential treatments. Furthermore, we wished to investigate how such a representation would affect specific team processes elucidated by the integrated capacity framework.

While several studies have analyzed computational boundary objects, they have primarily focused on the macro impact such boundary objects have had on an organization. For example, the introduction of anonymous blogs into an organization empowered employees to make candid recommendations on how to improve operations (Daniel, Hartnett, & Meadows, 2017). Because many of these recommendations from employees were implemented by the management, the evolving content of the blog revealed a shift in the organizational boundaries related to decision making. Similarly, the introduction of three-dimensional (3D) modeling software into design and construction organizations enabled drawings of the different trades (structure, plumbing, electricity) to be automatically generated from the 3D model (Taylor, 2007). However, the shift to a single computational model of the building from which other drawings were derived, required the designers to provide more accurate design details earlier in the design process. This shift contrasted to the approximate drawings they had previously produced using 2D software, because the engineers would redo the drawings at a later stage creating more accurate construction drawings. The introduction of the 3D system therefore expanded the boundary of team members who were responsible for providing accurate details of the building design. In contrast to such studies which have focused on the macro impact of computational boundary objects on organizational boundaries and relations, the current study focuses on the micro-phenomena related to the use of a computational boundary object in a multidisciplinary science team enabling a more detailed analysis of its use and impact on team dynamics.

Method

Our goal in the current study was to introduce the visual analytical data modeling method shown in Figure 2 as an intervention in MTTs to analyze its effects on generating novel translational insights.

Team Selection

We selected MTTs based on three criteria: (a) had possession of a data set containing molecular and clinical or demographical data of patients, (b) had the goal of identifying the biological mechanisms in the disease based on molecular and clinical information, and (c) had previous analysis and/or publications of the data set so we could determine whether the insights made during the analysis were novel and primarily based on the visual analytical intervention.

Data Collection

Building on the theoretical constructs from the integrative capacity framework (Salazar et al., 2012), and the iterative process inherent in visual analytics (Thomas & Cook, 2005), we introduced the visual analytical representation in each team with a focus on analyzing (a) its role as a boundary object that spanned disciplines and (b) the associated effects on team dynamics and outcomes. We recorded the members of the team and their primary discipline, prior publications on the data set selected for analysis, changes that were made to the data representation, who made the changes, and their motivations. We also recorded biomedical or methodological insights, in addition to publications, and grants at each step of the analysis.

Analysis

The above data were coded under the categories of data representation, team dynamics, and outcomes:

1. *Data Representation*. We focused on three concepts related to changes in the data representation: (a) *Trigger for Change* in the representation defined as the reason why the representation was changed (e.g., clusters being difficult to comprehend). (b) *Method of Change* in the representation that was defined as any computational, statistical, or manual change to the data representation resulting in a new state of the representation. (c) *Resultant State* of the representation defined as any visual or quantitative transformation of the data (e.g., normalization of the data, mapping of data into a network, and calculating the significance of the clustering). Minor modifications (e.g., change of font or zooming) which had no impact on the overall comprehension of the data were not considered as a changed representational state.
2. *Team Dynamics*. We focused on analyzing team dynamics at the granularity of changes in data representation. We therefore coded whether a team member had a *primary* contribution in changing the representation by suggesting or leading the change in representation, or a *supportive* contribution by observing, agreeing, or not agreeing for changing the representation. As the roles in each team were well-defined (e.g., the visual analyst had little knowledge of the domain, and conversely the physician had little knowledge of visual

analytics), the roles shifted based on the expertise needed in each step of the analysis, resulting in an unambiguous coding. However, analysis of such team dynamics with members that have overlapping expertise would require more fine-grained analysis with interrater consensus for the coding.

3. *Outcomes*. The outcomes were coded in the following categories: (a) *Biomedical Insights* defined as a cause–effect relationship derived from observed relationships in the data (Davidson & Sternberg, 1995), which were not previously published by the team members; (b) *Methodological Insights* related to the methodology; and (c) *Publications or Grants* which were based on, or included the biomedical and methodological insights made through the intervention.

Results

We present a detailed description of the effects of the visual analytical intervention in one MTT focused on severe asthma, and then summarize our results from our ongoing analysis of three other MTTs.

Team Structure and Goals

We selected an MTT focused on analyzing phenotypes of severe asthma that met our three inclusion criteria: possessing a data set containing molecular and clinical or demographical data of patients, having the goal for understanding the biological mechanisms in the disease, and having completed previous analysis and/or publications of the data set using conventional statistical methods to enable us to determine if the insights were novel. This team consisted of an asthma physician who led the MTT, and included a biologist specializing in proteomics, a statistician, a visual analyst, and a computer programmer.

Data and Prior Publication

The data consisted of 108 asthma patients, and 18 cytokines (protein molecules involved in intercellular communication) known to be implicated in asthma. Additionally, the data included clinical variables consisting of six lung function variables, and status on asthma severity (severe vs. not severe). The data had been analyzed and published (Brasier et al., 2008) using conventional statistical methods such as *t* tests, and regression. The published approach first classified patients based on severe and nonsevere patients, and then identified subphenotypes of patients using proteomics profiles. The cytokine-based subclassification was therefore dependent on the a priori classification of severe and not severe. However, it is well known that this classification does not predict the outcome of asthma (Bousquet et al., 2010). The team was therefore motivated to reanalyze the data using the visual analytical approach without the severity status, to identify heterogeneity in asthma and respective molecular pathways, with the goal of designing targeted treatments.

Visual Analytical Results

Figure 3 shows the seven main stages through which the data representation evolved. Each change in the data representation is associated with the primary or supportive role played by team members in changing the data representation, in addition to the outcomes related to insights, publications, and grants. Inspection of the seven stages revealed that they fell into the following four broad areas related to data preparation, visual representation, quantitative verification, and inference:

1. Data preparation
    a. The analysis began with the raw data in a spreadsheet (not shown in Figure 3) where the rows represented patients, the columns represented cytokines, and the cells contained raw cytokine expression values. As algorithms that use Euclidean space to display data require the values to have a uniform quantitative and interpretive range, the raw cytokine values were first converted into normalized values using the minimum–maximum (min-max) range-normalization method. This method transforms the differing ranges of data in each variable be from 0 to 1 by making the minimum value = 0, the maximum value = 1, and mapping of the rest of the values proportionally. The visual analyst played the primary role in making this decision based on prior experience in conducting network analysis of similar data, and was supported by a programmer who implemented the data transformation. The resulting data representation was a spreadsheet containing range-normalized cytokine expression values.
2. Visual representation and exploration
    b. The normalized symbolic data were then visually represented as a bipartite network using Pajek (a well-known network visualization and analysis application; Nooy, Mrvar, & Batagelj, 2011). For this transformation, the analyst needed to determine which *entities* in the data should represent the nodes in the network, and which *relationships* between the entities in the data should represent the edges in the network. After conferring with the physician and biologist, the visual analyst represented the nodes as either patients or cytokines, and edges to represent the range-normalized cytokine expression values. Furthermore, to capture the degree of cytokine expression, the edge thicknesses were made proportional to the normalized cytokine expression values. Therefore, thick edges represented higher cytokine expression values compared to thin edges. Finally, the size of the node was made proportional to the total expression value of the connecting edges. Therefore, large patient nodes have overall higher aggregate normalized cytokine expression values compared with smaller patient nodes. The resultant data representation was a bipartite network as described above, where the network nodes were layout out around a circle.
    c. As the circular layout of the network did not make salient the similarities and differences among nodes, the visual analyst applied a standard *Kamada–Kawai* layout algorithm (Kamada & Kawai, 1989) to the network. This

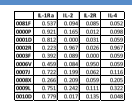
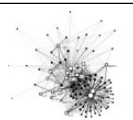
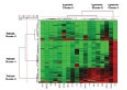
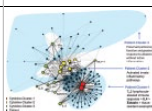
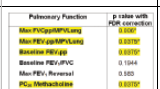
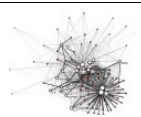
		Data Representation			Team Dynamics		Outcomes		
		Trigger for Change	Method of Change	Resultant State	Primary Role	Supportive Role	Biomedical Insights	Methodological Insights	Publications & Grants
Data Preparation	A	Bipartite networks need a uniform quantitative and interpretive range for all values to enable comparisons and clustering	Used min-max range normalization method of raw cytokine values		Visual Analyst	Programmer			
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 represenation. Node size represented sum of cytokine expression, and edge thickness represented degree of cytokine expression		Visual Analyst	Physican, Biologist		Circular layout of nodes does not make salient how nodes are similar or different	
	C	Lay out the network to reveal similarity and differences among nodes	Used force-directed layout algoirthm to lay out nodes in a 2D plane		Visual Analyst, Statistican	Physican, Biologist	Existence of patient subgroups based on cytokine profiles	Statistican 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 encirle 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 signifcant clincal variables across clusters		Visual Analyst, Statistician,	Programmer			
Inference of Patterns	G	Comprehend how statistically significant clincial 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 pubication, and 2 peer-reviewed methodological publications; included in CTSA proposal which was funded

Figure 3. Evolution of the data representation through seven stages with associated effects on team dynamics and outcomes.

resulted in nodes with a similar pattern of connecting edge weights to be pulled together, and those with different patterns to be pushed apart. Studying the network layout, the physician and biologist noted the existence of distinct clusters of patients and cytokines, in addition to intercluster relationships such as which patient clusters are most closely related with which cytokine clusters. This complex but understandable clustering generated high interest in the team. However, the statistician, seeing the output of a bipartite network analysis for the first time, questioned the normalization method used, which was different from the standard normalization approaches he had used. This led both to do additional reading and experiments, which resulted in an understanding of the distinction between a regular normalization method, compared to the range normalization method being used. Both agreed that the min–max range normalization method was valid, but could be sensitive to new patients being added to the current data set, potentially changing the maximum and minimum values of any variable. As there were no plans to add additional patients to the data set, the team decided to move ahead with the min–max range normalization method.

3. Quantitative verification
   d. While the layout algorithm revealed the overall topology of the data suggesting strong clustering, it was not designed to identify the cluster boundaries. The visual analyst therefore used quantitative clustering methods to identify the boundaries of the clusters and to test the significance of the clustering. This was done using hierarchical clustering, visualized through a heatmap and dendrogram. The results identified three patient clusters and three cytokine clusters. However, the programmer who had used clustering before, questioned the linkage function selected for the hierarchical clustering, but agreed that as there were no guidelines to pick one over the other, and therefore they should proceed with the analysis. The physician and biologist who had conducted such clustering before supported the continuation of the analysis. The visual analyst also used random permutations of the network data to determine if the clustering could have occurred by chance, which was implemented by the programmer.
   e. Given that the network and heatmap representations were different views of the same data and difficult to cognitively integrate, the visual analyst superimposed the cluster boundaries over the network to produce an integrated visualization of patient and cytokine clusters. This was achieved by coloring the nodes such that nodes in each cluster shared the same color. However, this approach resulted in too many differently colored nodes, which did not improve comprehension. The analyst therefore decided to use two different methods for denoting cluster boundaries. Using Pajek, he first used the same color for all cytokine nodes within a cluster. He then copied the network into PowerPoint and drew shapes to denote the boundaries of the patient clusters. This representation, which combined the visual and

quantitative results was presented to the team for interpretation. The biologist and physician agreed that based on their prior knowledge, the clustering of the cytokines made biological sense, but they were unable to infer the molecular pathway with only that information. The physician therefore requested additional analysis to determine how the patient clusters differed in their clinical lung function variables.

f. The visual analyst and the statistician together selected *Kruskal–Wallis* (a nonparametric statistical test to address the skewed distribution of the data) for identifying which of the six patient lung function variables were significantly different across the three patient clusters. Three of the variables were significantly different after correcting for multiple testing. The results were displayed in a table containing all six variables, with the three that were statistically significant highlighted in yellow.

4. Inference of biological mechanisms and asthma phenotypes

g. To associate the significant variables with the patient clusters, the analyst decided to color the patient nodes based on the significant variables. However, as the patient nodes could have only one color at a time, the visual analyst generated three additional networks, one for each of the three significant variables. These networks were shown using PowerPoint animation to enable an understanding of how the three patient clusters were associated with the clinical variables, and their relationship to cytokine clusters. Using the integrated representation of molecular and clinical information, the biologist and clinician began the process of interpreting the network. They first referred to the Ingenuity Pathway Analysis database to verify their inferences, checked prior publications on similar pathways, and then arrived at a consensus. Figure 4 shows an enlarged image of the network to show the pathway that was identified for each patient cluster.

Given the importance of the novel insight, we present as data a verbatim description of the pathway inferences and implications for precision medicine written by the physician and biologist from our publication:

> The frequent co-occurrence of Eotaxin and IL-4 (Cytokine-Cluster-1) is congruent with a known sequence of molecular changes in asthma patients who often have a T-helper-2 (Th2) lymphocyte-skewed immune response. This response results in the secretion of IL-4, which in turn triggers Eotaxin production by bronchial epithelial cells (Fujisawa et al., 2000). The resulting downstream actions include the activation and recruitment of tissue-resident eosinophils, a hallmark of early stage asthma. The presence of Eotaxin and IL-4 in lung fluids therefore appears to indicate key sub-stages of a complex molecular pathway in asthma, which explains their high co-occurrence in the network.
>
> To comprehend the biological significance of cytokines in Cytokine-Cluster-2 (IL-5, IFN-γ, MIP1a, MIG, IL-17, and MIP-1β), they were entered into the Ingenuity Pathway Analysis (IPA) application. The results suggest that the frequent co-occurrence of these cytokines is regulated by the innate inflammatory nuclear factor-κB pathway (NF-κB).

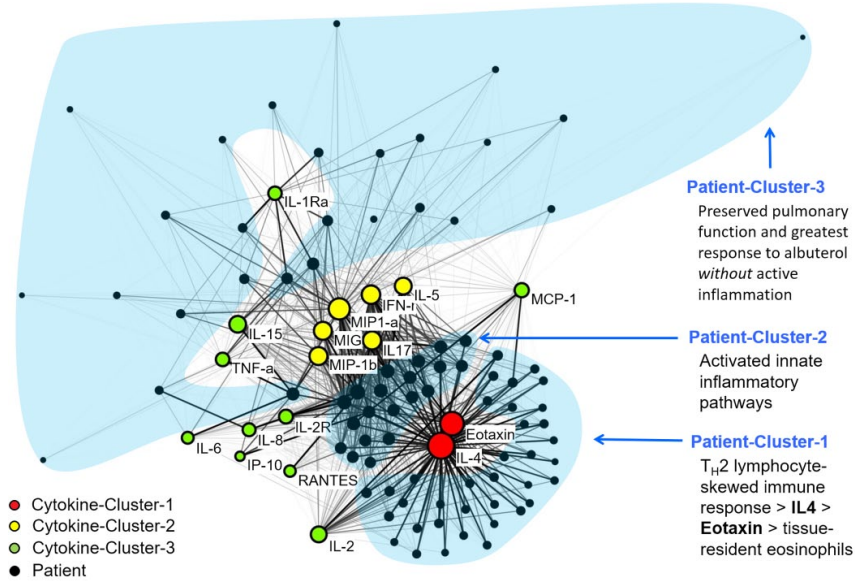


Figure 4. The visual analytical representation showing the interrelationship between three patient clusters with three cytokine clusters, and the resulting inferences of pathways for each patient cluster based on an integrated understanding of molecular and clinical information.

NF-κB is a potent pro-inflammatory transcription factor that activates expression of cytokine networks. In addition, persistent NF-κB activation has been linked to uncontrolled/acute exacerbations of asthma (Gagliardo et al., 2003). The frequent co-occurrence of this set of cytokines therefore implies the presence of a distinctly different pro-inflammatory state, when compared to the IL-4–Eotaxin process discussed above.

The above cytokine clusters combined with the pulmonary functions of the patients, provide a biological explanation for the patient clusters. The strong relationship of Patient-Cluster-1 to Cytokine-Cluster-1 suggests that patients in this cluster have disease driven primarily by Th2 inflammation. In contrast, Patient- Cluster-2 has a strong relationship to both Cytokine-Clusters-1 and -2. This result implies that patients in Patient-Cluster-2 have a component of activated innate inflammatory pathways. Additional evidence for this inference of state-based clusters is evidenced by differences in pulmonary function across the clusters discussed earlier. Patient-Cluster-3 which has the lowest cytokine values for both of the above cytokine clusters also has the largest number of significant differences in obstructive airway disease parameters in pulmonary function testing, and lowest airway reactivity response to methacholine compared to Patient-Clusters-1 and -2. This result implies that Patient-Cluster-3 represents a subgroup of asthmatics with preserved pulmonary function and greatest response to albuterol without active inflammation.

Informed by these underlying molecular processes, the analysis of patients and cytokines therefore implies a state-based classification of asthma patients. The results also provide evidence for the growing consensus (Bousquet et al., 2010) that asthma is a

> dynamic disease where the same patient could enter different asthmatic states based on environmental and/or other triggers. The biologist and physician concluded that future studies that include such information could lead to a better understanding of the relationship between triggers and resulting asthmatic states, which could translate into more effective personalized treatment and prevention approaches for each patient (Bhavnani et al., 2011, p. S29).

The above integration of clinical and molecular information, along with the inferences resulted in two domain publications, and two peer-reviewed methodological publications, including an informatics distinguished paper award for defining a new molecular-based taxonomy of asthma. The analysis was included as part of the Clinical and Translational Science Awards (CTSA) Program proposal to the National Institutes of Health (NIH), which was subsequently funded.

Discussion

The micro-level analysis of how visual analytics was used as a computational boundary object in an MTT provided a deeper understanding of evolving representations of the data, shifts in primary and supportive roles of the team members, and novel translational outcomes:

Evolving Representation

The results suggest that the process of integrating molecular and clinical information in a visual representation enabled innovative translational insights. However, the data representation and its content evolved through several intermediate states using *representation shifts* (table, network, heatmap with dendrogram), *representation transformations* (circular network, force-directed network, layered network with clusters, layered network with significant clinical variables), and *content modifications* (raw table values, normalized table values, clusters, significance) driven by feedback from different team members, and leading to emergent representations and insights. Therefore, similar to results reported by other studies on computational boundary objects discussed earlier (Daniel et al., 2017; Taylor, 2007), the boundary object was not a static object, but rather one that evolved having its own emergent states, in addition to having emergent effects on team dynamics. However, in contrast to previous studies where the boundary object was reported as changing mainly in *content* (e.g., text changes in a blog, graphical additions to a 3D model), here we observe that the boundary object in addition to changing in content, also changed in *representation* by shifting and transforming. This distinction between the content and representation of a boundary object is critical as it provides the necessary specificity for describing a class of boundary objects which can be referred to as computational evolving boundary objects (CEBOs), defined as digital boundary objects that evolve in representation and content in response to the needs of integrating disciplinary knowledge in multidisciplinary teams. Future designs of CEBOs could include interactive features such as filtering, layering, and intersecting different types of information in multiple

representations and views of the data enabling team members to make fluid and rapid modifications to data representations, resulting in new insights for improving health outcomes. Such designs will be critical for supporting discoveries in the massive amounts of multiomics data including genomic, proteomic, metabolomics, and phenomic from millions of patients becoming available to scientists in the not-too-distant future. Therefore, while the evolution of data representations and their content can be expected when conducting any computational analysis, analyzing that evolution through the lens of a boundary object enabled the definition and design of CEBOs from a sociomaterial perspective, with potential impact on the ability of a wide range of teams to integrate diverse disciplinary knowledge.

Shifting Primary and Supportive Roles

The results also showed that there was a shift in the primary and supportive roles that the team members played at different stages of the analysis. MTTs in medical schools tend to have an implied hierarchy with physicians having the highest status, and analysts playing an important but supportive role. This hierarchy is exemplified by the mainly middle-author status that analysts tend to have in such team publications. However, the use of the visual analytical representation for enabling novel insights required involvement of the different team members with shifting primary and supportive roles. While the initial steps of data preparation and visual representation were led by the visual analyst and the statistician, they required important input from the clinician and biologist for designing the visual representation. However, this primary role shifted as the data representation became more complex and required the domain experts to drive the subsequent changes to the data representation and resulting interpretations. In the last stage, the integration of the cluster boundaries determined by the cytokine profiles, together with the clinical variables required most of the team working together resulting in the inference of the biological pathways. The evolving data representations and their content, based on input from different team members, resulted in a more egalitarian team structure, with different members playing primary and supportive roles at different stages. While a shift in the boundaries of roles and relations has been reported in several organizational studies (Barrett, Oborn, Orlikowski, & Yates, 2012; Kellogg, Orlikowski, & Yates, 2006), here we observe that the primary and supportive roles changed throughout the project, revealing a fluid shift of boundaries in response to and affecting the evolution of the data representations with no single member dominating the overall contributions. Furthermore, as discussed below, the authorship and type of publications also reflected this shift in team boundaries.

Novel Translational Insights

The results show that the visual analytics functioned as a boundary object enabling the team to integrate knowledge from visual analytics, statistics, molecular biology, and the clinical treatment of asthma, resulting in a novel translational insight of disease

mechanisms in each patient cluster. These insights were new and not previously published in the team's prior publications. The novel results were published in two domain and two methodological publications (where the visual analyst was first author), suggesting that the intervention was not only instrumental in precipitating novel insights but also affected the team dynamics resulting in a more egalitarian team structure. As the identification of patient subgroups and their respective biological pathways was an innovation that transcended any single discipline, it suggests evidence of "collective intelligence" (Woolley, Chabris, Pentland, Hashmi, & Malone, 2010), an emergent characteristic of highly successful teams.

Generalization of Results in Other MTTs

Based on the lessons learned from observing the effects of applying the visual analytical intervention in the Phenotypes of Severe Asthma MTT, we have since used it to analyze data sets produced by other MTTs that met the inclusion criteria discussed earlier. These include the following three MTTs and their data sets which have been analyzed: (a) Prevention of preterm births using data from 22 preterm delivery mothers and 28 controls, in addition to their profile on 10 methylation sites; (b) reducing hospital readmission using data of 17,402 readmitted patients with chronic obstructive pulmonary disease with matched controls from the 2013 Medicare database; and (c) 1,411 Alzheimer's cases and controls and their profile on 88,709 single-nucleotide polymorphisms. Each of the above three teams was led by a physician or biologist, and included a statistician and visual analyst.

Our preliminary analysis of the first two MTTs has revealed that the overall outcomes are similar to what we observed in the phenotypes of severe asthma MTT. In both teams, the representation evolved to integrate clinical or molecular associations with outcomes, leading to an insight that the domain experts agreed were worthy of publication. The results were published with a focus on domain or methodological contributions with the visual analyst either being the sole corresponding author, or being a joint corresponding author. However, our approach failed to find any patterns in the data set provided by the Alzheimer's MTT. The network showed nodes appearing to be randomly organized in a topology colloquially referred to as the "network hairball." This limitation of the visual analytical approach when analyzing very high dimensional data due to the large number of variables, was subsequently addressed through the design of new algorithms that find and display patterns in large and dense networks, such as what are increasingly becoming common in MTTs analyzing large data sets (Bhavnani et al., 2017). The interventions in real-world settings with real biomedical data therefore helped reveal the strengths and limitations of our visual analytical representation.

Conclusions

Throughout history, humans have assembled into teams to address problems that are difficult for individuals to solve alone. Such teams have ranged from villages of people collectively growing crops, to armies of soldiers coordinating defenses against

invaders. However, a key difference between such teams, and multidisciplinary scientific teams, is the high degree of specialization that scientific team members tend to have. While such specialization is necessary for solving complex real-world problems, scientific teams face considerable social, cultural, and cognitive barriers reducing a team's ability for knowledge integration and innovation. Here, we explored the role that visual analytics can play as a computational boundary object to help integrate knowledge across diverse disciplines in an MTT, with the goal of enabling novel insights and innovations. The results of testing such an approach led to implications for theory and practice.

Implications for Theory

The results suggest that a boundary object was not just a static object, but rather one that evolved through several emergent states progressively integrating different types of disciplinary knowledge with shifts and transformations in representation, in addition to changes in content, and ultimately enabling the team to have translational insights. Furthermore, the process of evolving the boundary object appeared to make the teams more egalitarian as it enabled each member to play primary and supportive roles in motivating a change in the representation thereby building a common ground for communication critical for generating insights. These results provide support and deepen our understanding of boundary objects and their role as an intervention for increasing the integrative capacity of scientific teams. Furthermore, the results also suggest that the bipartite network's ability to model patients, characteristics, and clinical outcomes in a unified externalized representation, played an important role in enabling the team to progressively integrate biological, clinical, and statistical knowledge. However, our analysis of this intervention in three other MTTs helped reveal a critical limitation in the representation when analyzing high-dimensional data, which in turn led to new methodological innovations.

Abstracting results from the current and prior studies, future research should explore a model for CEBOs which includes (a) *states* of the boundary object defined by changes in representation and content; (b) *triggers* for evolving the boundary object driven by domain and methodological insights from the team members; (c) *methods* for changing the state of the boundary object; (d) *roles* played by the team members in triggering, implementing, and interpreting each new state of the boundary object; and (e) *outcomes* at the level of each change of state, and at the level of the entire project which could include insights, innovations, publications, trust, and collective intelligence. Such a model should accelerate the design and analysis of CEBOs representing an instance of a broad class of "Team-Centered Informatics" approaches that enable multidisciplinary scientific teams get an integrated understanding of human biology and disease from large and diverse datasets, with the goal of improving health. Besides being useful within teams, Team-Centered Informatics solutions could also be useful to integrate knowledge across teams and organizations, and across other disciplines beyond biomedicine.

Implications for Practice

As discussed above, the results suggest that the visual analytical approach was effective in generating novel translational insights in three of the four MTTs in which we tested the intervention. However, because the tools for conducting network visualization and analysis are currently designed for advanced users, the approach required a team member who had computational expertise in making the modifications needed by the team. Future research should explore how a small set of application features such as layering, filtering, and intersecting multiomics variables could enable members of MTTs without extensive computational expertise to make rapid changes to CEBOs based on the needs of their team. Furthermore, as multidisciplinarity becomes mainstream in science and medicine, students and analysts should be taught how to use CEBOs not just as static representations, but as continuously evolving in content and representation helping to integrate diverse information with the goal of enabling translational teams make novel insights.

Such theoretical and practical advances, especially in the age of exploding and diverse data in virtually every field, could in the future enable multidisciplinary scientific teams to overcome critical team-centric barriers, with the goal of more effectively addressing the increasingly complex real-world problems that humans face.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was conducted with the support of the Institute for Translational Sciences at the University of Texas Medical Branch, supported in part by a Clinical and Translational Science Award (UL1 TR001439) from the National Center for Advancing Translational Sciences, National Institutes of Health, by a Patient-Centered Outcomes Research Institute contract (ME-1511-33194), and by a National Library of Medicine of the National Institutes of Health award (R01 LM012095). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Patient-Centered Outcomes Research Institute.

References

Abu-Jamous, B., Fa, R., & Nandi, A. K. (2015). *Integrative cluster analysis in bioinformatics*. Chichester, West Sussex, England: John Wiley.

Ameredes, B. T., Hellmich, M. R., Cestone, C. M., Wooten, K. C., Ottenbacher, K. J., Chonmaitree, T., . . . Brasier, A. R. (2015). The multidisciplinary translational team (MTT) model for training and development of translational research investigators. *Clinical and Translational Science*, *8*, 533-541. doi:10.1111/cts.12281

Barrett, M., Oborn, E., Orlikowski, W. J., & Yates, J. (2012). Reconfiguring boundary relations: Robotic innovations in pharmacy work. *Organization Science*, *23*, 1448-1466.

Bhavnani, S. K., Bellala, G., Victor, S., Bassler, K. E., & Visweswaran, S. (2012). The role of complementary bipartite visual analytical representations in the analysis of SNPs: A

case study in ancestral informative markers. *Journal of the American Medical Informatics Association*, *19*, e5-e12. doi:10.1136/amiajnl-2011-000745

Bhavnani, S. K., Chen, T., Ayyaswamy, A., Visweswaran, S., Bellala, G., Rohit, D., & Kevin, E. B. (2017). Enabling comprehension of patient subgroups and characteristics in large bipartite networks: Implications for precision medicine. *AMIA Joint Summits on Translational Science proceedings*, *2017*, 21-29.

Bhavnani, S. K., Dang, B., Bellala, G., Divekar, R., Visweswaran, S., Brasier, A., & Kurosky, A. (2015). Unlocking proteomic heterogeneity in complex diseases through visual analytics. *Proteomics*, *15*, 1405-1418. doi:10.1002/pmic.201400451

Bhavnani, S. K., Victor, S., Calhoun, W. J., Busse, W. W., Bleecker, E., Castro, M., . . . Brasier, A. R. (2011). How cytokines co-occur across asthma patients: from bipartite network analysis to a molecular-based classification. *Journal of Biomedical Informatics*, *44* (Suppl. 1), S24-S30. doi:10.1016/j.jbi.2011.09.006

Bietz, M. J., Bloss, C. S., Calvert, S., Godino, J. G., Gregory, J., Claffey, M. P., . . . Patrick, K. (2016). Opportunities and challenges in the use of personal health data for health research. *Journal of the American Medical Informatics Association*, *23*, e42-e48. doi:10.1093/jamia/ocv118

Bousquet, J., Mantzouranis, E., Cruz, A. A., Ait-Khaled, N., Baena-Cagnani, C. E., Bleecker, E. R., . . . Zuberbier, T. (2010). Uniform definition of asthma severity, control, and exacerbations: document presented for the World Health Organization Consultation on Severe Asthma. *Journal of Allergy and Clinical Immunology*, *126*, 926-938. doi:10.1016/j.jaci.2010.07.019

Brasier, A. R., Victor, S., Boetticher, G., Ju, H., Lee, C., Bleecker, E. R., . . . Calhoun, W. J. (2008). Molecular phenotyping of severe asthma using pattern recognition of bronchoalveolar lavage-derived cytokines. *Journal of Allergy and Clinical Immunology*, *121*, 30-37. e36. doi:10.1016/j.jaci.2007.10.015

Bunderson, J. S., & Sutcliffe, K. M. (2002). Comparing alternative conceptualizations of functional diversity in management teams: Process and performance effects. *Academy of Management Journal*, *45*, 875-893. doi:10.2307/3069319

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. San Francisco, CA: Morgan Kaufmann.

Daniel, E.M., Hartnett, E.J., Meadows, M. (2017). Don't throw rocks from the side-lines: A sociomaterial exploration of organizational blogs as boundary objects. *Information Technology & People*, *30*, 542-561. doi:10.1108/ITP-02-2015-0036

Davidson, J.E., & Sternberg, R.J. (1996). *The nature of insight*. Cambridge: MIT Press.

Dougherty, D. (1992). Interpretive barriers to successful product innovation in large firms. *Organization Science*, *3*, 179-202. doi:10.1287/orsc.3.2.179

Falk-Krzesinski, H. J., Börner, K., Contractor, N., Fiore, S. M., Hall, K. L., Keyton, J., . . . Uzzi, B. (2010). Advancing the science of team science. *Clinical and Translational Science*, *3*, 263-266. doi:10.1111/j.1752-8062.2010.00223.x

Fitzpatrick, A. M., Teague, W. G., Meyers, D. A., Peters, S. P., Li, X., Li, H., . . . Moore, W. C. (2011). Heterogeneity of severe asthma in childhood: Confirmation by cluster analysis of children in the National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. *Journal of Allergy and Clinical Immunology*, *127*, 382-389. doi:10.1016/j.jaci.2010.11.015

Fujisawa, T., Kato, Y., Atsuta, J., Terada, A., Iguchi, K., Kamiya, H., . . . Hirai, K. (2000). Chemokine production by the BEAS-2B human bronchial epithelial cells: differential regulation of eotaxin, IL-8, and RANTES by TH2- and TH1-derived cytokines. *Journal of Allergy and Clinical Immunology*, *105*(1 Pt. 1), 126-133.

Gagliardo, R., Chanez, P., Mathieu, M., Bruno, A., Costanzo, G., Gougat, C., . . . Vignola, A. M. (2003). Persistent activation of nuclear factor-κB signaling pathway in severe uncontrolled asthma. *American Journal of Respiratory and Critical Care Medicine*, *168*, 1190-1198. doi:10.1164/rccm.200205-479OC

Gray, B. (2008). Enhancing transdisciplinary research through collaborative leadership. *American Journal of Preventive Medicine*, *35*, S124-S132. doi:10.1016/j.amepre.2008.03.037

Hage, J., & Meeus, M. (2009). *Innovation, science and institutional change. A handbook of research*. Oxford, England: Oxford University Press.

Hall, K. L., Stokols, D., Stipelman, B. A., Vogel, A. L., Feng, A., Masimore, B., . . . Berrigan, D. (2012). Assessing the value of team science: A study comparing center- and investigator-initiated grants. *American Journal of Preventive Medicine*, *42*, 157-163. doi:10.1016/j.amepre.2011.10.011

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. New York, NY: Springer.

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM*, *53*(6), 59-67. doi:10.1145/1743546.1743567

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*, 575-604. doi:10.1146/annurev.psych.53.100901.135109

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., . . . von Mering, C. (2009). STRING 8: A global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research*, *37*, D412-D416. doi:10.1093/nar/gkn760

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, *31*, 7-15.

Kehl, V., & Ulm, K. (2006). Responder identification in clinical trials with censored data. *Computational Statistics & Data Analysis*, *50*, 1338-1355. doi:10.1016/j.csda.2004.11.015

Kellogg, K. C., Orlikowski, W. J., & Yates, J. (2006). Life in the trading zone: Structuring coordination across boundaries in postbureaucratic organizations. *Organization Science*, *17*, 22-44. doi:10.1287/orsc.1050.0157

Klein, J. T. (2010). A taxonomy of interdisciplinarity. In J. T. Klein & C. Mitcham (Eds.), *The Oxford handbook of interdisciplinarity* (pp. 15-30). Oxford, England: Oxford University Press.

Klein, K. J., Knight, A. P., Ziegert, J. C., Lim, B. C., & Saltz, J. L. (2011). When team members' values differ: The moderating role of team leadership. *Organizational Behavior and Human Decision Processes*, *114*, 25- 36. doi:10.1016/j.obhdp.2010.08.004

Lélé, S., & Norgaard, R. B. (2005). Practicing interdisciplinarity. *BioScience*, *55*, 967-975. doi:10.1641/0006- 3568(2005)055[0967:PI]2.0.CO;2

Lipkovich, I., Dmitrienko, A., Denne, J., & Enas, G. (2011). Subgroup identification based on differential effect search: A recursive partitioning method for establishing response to treatment in patient subpopulations. *Statistics in Medicine*, *30*, 2601-2621. doi:10.1002/sim.4289

Liyanage, S., & Barnard, R. (2003). Valuing of firms' prior knowledge: A measure of knowledge distance. *Knowledge and Process Management*, *10*, 85-98. doi:10.1002/kpm.171

Newman, M. E. J. (2010). *Networks: An introduction*. Oxford, England: Oxford University Press.

Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory social network analysis with Pajek* (2nd ed.). Cambridge, England: Cambridge University Press.

Olson, G. M., & Olson, J. S. (2000). Distance matters. *Human–Computer Interaction*, *15*, 139-178. doi:10.1207/s15327051hci1523_4

Perry-Smith, J. E. (2006). Social yet creative: The role of social relationships in facilitating individual creativity. *Academy of Management Journal*, *49*, 85-101. doi:10.5465/amj.2006.20785503

Salazar, M. R., Lant, T. K., Fiore, S. M., & Salas, E. (2012). Facilitating innovation in diverse science teams through integrative capacity. *Small Group Research*, *43*, 527-558.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, "translations" and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, *19*, 387-420. doi:10.1177/030631289019003001

Stipelman, B., Hall, K., Zoss, A., Okamoto, J., Stokols, D., & Börner, K. (2014). *Mapping the impact of transdisciplinary research: A visual comparison of investigator-initiated and team-based tobacco use research publications*. Retrieved from http://cns.iu.edu/docs/publications/2014-stipelman-mappingimpact.pdf

Stokols, D., Hall, K. L., Taylor, B. K., & Moser, R. P. (2008). The science of team science: Overview of the field and introduction to the supplement. *American Journal of Preventive Medicine*, *35*, S77-S89. doi:10.1016/j.amepre.2008.05.002

Stokols, D., Harvey, R., Gress, J., Fuqua, J., & Phillips, K. (2005). In vivo studies of transdisciplinary scientific collaboration: Lessons learned and implications for active living research. *American Journal of Preventive Medicine*, *28*, 202-213. doi:10.1016/j.amepre.2004.10.016

Taylor, J. E. (2007). *How are technological changes in boundary objects enacted in design and construction networks?* Paper presented at the Sloan Industry Studies, Cambridge, MA.

Thomas, J. J., & Cook, K. A. (Eds.). (2005). *Illuminating the path: The R&D agenda for visual analytics*: Piscataway, NJ: IEEE Press.

Trochim, W. M., Marcus, S. E., Masse, L. C., Moser, R. P., & Weld, P. C. (2008). The evaluation of large research initiatives: A participatory integrative mixed-methods approach. *American Journal of Evaluation*, *29*, 8-28.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, *330*, 686-688. doi:10.1126/science.1193147

Wooten, K. C., Calhoun, W. J., Bhavnani, S., Rose, R. M., Ameredes, B., & Brasier, A. R. (2015). Evolution of multidisciplinary translational teams (MTTs): Insights for accelerating translational innovations. *Clinical and Translational Science*, *8*, 542-552. doi:10.1111/cts.12266

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science*, *18*, 87-122.