How Can An Interactive Visual Analytics Tool Help Biomedical Scientists Investigate Genotype-Phenotype Relationships?

By

Erik Jon Segerdell

A Thesis

Presented to the Department of Medical Informatics and Clinical Epidemiology
and the Oregon Health & Science University
School of Medicine
in partial fulfillment of
the requirements for the degree of

Master of Science

November 2015

Department of Medical Informatics and Clinical Epidemiology
School of Medicine
Oregon Health & Science University

Certificate of Approval

This is to certify that the Master's thesis of
Erik Segerdell
has been approved

Beth Wilmot, Ph.D., Mentor

Melissa Haendel, Ph.D.

Harry Hochheiser, Ph.D.

Michelle Hribar, Ph.D.

Contents

Acknowledgments **vi**

Abstract **vii**

1 Introduction **1**

1.1 The animal model data landscape 1

1.2 Biomedical data visualization 2

1.2.1 Graph-structured data 2

1.2.2 The state of current tools 4

1.3 The study 6

1.3.1 Purpose and philosophy 6

1.3.2 Discount usability and user-centered design 7

2 Methods **10**

2.1 Overview 10

2.2 Design and evaluation strategy 11

2.2.1 Iterative design 11

2.2.2 Users and user sessions 12

2.3 Mockups, prototype, and user tasks 14

2.3.1 First iteration 14

2.3.2 Second iteration 17

2.4 Biological use cases 21

2.5 Data sources 22

2.6 Software 23
2.7 Evaluation methods 23

3 Results **26**
3.1 Overview 26
3.2 The mockups 27
3.2.1 Scenario development and interface features 27
3.2.2 User feedback and requests 28
3.3 The functional prototype 31
3.3.1 Feature selection 31
3.3.2 Usability evaluation 32

4 Discussion **35**
4.1 User-centered discount methods in practice 35
4.2 Future requirements 37
4.3 Future considerations 39
4.4 The takeaways 40

5 Conclusion **42**

Bibliography **43**

Figures **53**

Tables **80**

List of Figures

1 A mouse database gene page. 54

2 A graph of some human and model organism data. 55

3 A graph highlighting gaps in knowledge. 56

4 Overview of the study. 57

5 Task A mockup: graph visualization and contextual details panel. . . 58

6 Task A mockup: “Information” tab. 58

7 Task A mockup: “Relationship” tab. 59

8 Task A mockup: “Evidence” tab. 59

9 Task A mockup: slider for filtering related nodes. 60

10 Task A mockup: filter applied to related nodes. 60

11 Task B mockup: gene family. 61

12 Task B mockup: gene and related nodes, including metanode. 61

13 Task B mockup: node selection. 62

14 Task B mockup: clearing of unselected nodes. 62

15 Task B mockups: node expansion, filtered by data type. 63

16 Task B mockup: addition of nodes related to a metanode. 63

17 Task B mockup: path between selected nodes. 64

18 Task C mockups: gene, variants, and disease. 64

19 Task C mockup: exploration of related animal model data. 64

20 Task C mockup: addition of related phenotype. 65

21 Task C mockup: further exploration to develop a biological picture. . 66

22 Functional prototype: help dialog window. 67

23 Functional prototype: legend. 67

24 Functional prototype: initial display. 68

25 Functional prototype: node details. 69

26 Functional prototype: right-click popup on a node. 70

27 Functional prototype: node expansion and edge details. 71

28 Functional prototype: addition of individual nodes. 72

29 Functional prototype: multiple paths in the graph. 73

30 System Usability Scale. 74

31 Sketches of search scenarios. 75

32 Sketches of alternative graph-based views. 76

33 SUS scoring. 77

34 D3 cluster layout. 78

35 D3 zoomable partition layout. 79

List of Tables

1 Iterative interface design and usability evaluation plan. 81

2 Questions for the users. 83

3 Data and relationship types supported by the functional prototype. . 84

4 User feedback on the mockups. 85

5 Feature requests emanating from discussion of the mockups. 86

6 Candidate user interface features for graph visualization. 87

7 User feedback on the functional prototype. 88

8 Feature requests emanating from prototype evaluation. 89

9 System Usability Scale scores. 90

Acknowledgments

I would like to express my sincere gratitude to my thesis mentor Dr. Beth Wilmot at Oregon Health & Science University for her guidance and enthusiastic support of my Master's project. I sincerely thank the rest of my Thesis Advisory Committee: Dr. Melissa Haendel, Dr. Harry Hochheiser, and Dr. Michelle Hribar for their encouragement, hard questions, and expert suggestions.

Thank you to Dr. Armand Bankhead, Dr. Shannon McWeeney, Dr. Jayashree Kalpathy-Cramer; and to colleagues at the Department of Medical Informatics and Clinical Epidemiology, Knight Cancer Institute, and Xenbase for their support as I juggled work and school the past four years. My thanks also goes to Lara Fournier, Fred Loney, and the OHSU Ontology Development Group for their constructive comments on my public proposal and thesis presentations.

Finally, I am grateful to my six colleagues who found time in their busy schedules to participate in this study as testers and whose knowledge and creativity were crucial to its success.

Abstract

Genotypes and phenotypes are fundamental to many areas of the life sciences. Biomedical and animal model databases contain a wealth of information about genetics, genomics, embryonic development, phenotypes, and diseases, presenting biomedical researchers the opportunity to uncover human disease mechanisms. Currently, however, investigators lack suitable informatics tools to effectively meet the challenge of traversing these rich datasets, synthesizing the biology described therein, and identifying gaps in our knowledge that need further research. In this study, I address the question of how a visual analytics tool can help scientists investigate these important biological relationships. I developed a prototype of a tool that enables users to visually explore connections between human and animal model data rendered as a network graph. Two iterations of a prototype were produced, including mockups and a functional software-based application. A small group of users was recruited to provide feedback, identify needs, and evaluate the prototype's usability. With different user interface design and evaluation approaches from which to choose and a brief time frame in which to conduct the study, I employed a user-centered design philosophy, qualitative "discount" evaluation methods, and software libraries well-

suited to making a visually appealing and highly responsive interface. The study generated abundant high-quality user feedback, a generally usable early software prototype, and a comprehensive set of recommendations for subsequent tool development – confirming the effectiveness of these choices.

1. Introduction

1.1 The animal model data landscape

Genotypes and phenotypes are fundamental to many areas of the life sciences. Seeking biological insights into the relationship and distinction between organisms' observable traits of above the molecular level – *phenotypes* – and the genetic makeup that they inherited from their parents – *genotypes* – has long been a key goal of research [1, 2]. Today, a vast and growing wealth of information about genetics, genomics, phenotypes, embryonic development, and environmental interactions populates biomedical and model organism databases. These data present biomedical researchers the opportunity to uncover human disease mechanisms and drive discovery of new therapies, provided they are able to make sense of massive quantities of information. Currently, investigators lack suitable bioinformatics tools to effectively meet this challenge.

Animal model systems are the backbone of our understanding of biological processes and the testing of gene-based disease hypotheses [3]. Model organism databases such as those for the laboratory mouse, zebrafish, and the fly *Drosophila* contain an enormous amount of genetic, genomic, and phenotypic data [4, 5]. Their web inter-

faces tend to be primarily text-based, which allows biologists to inspect and evaluate data in exquisite detail but may require them to click through many tabs or pages to synthesize information and discern complex relationships. An example from the mouse informatics database is shown in Figure 1.

The interconnectivity of different types of biological data can be quite complex and daunting from the standpoint of creating a resource for querying and exploring information in a scientifically productive and efficient way. To visualize this, Figure 2 integrates, in the form of a *graph*, some human and mouse data about a disease and its related genotypes and phenotypes. Graphs are mathematical models of network structures, a collection of *nodes* and *edges*, or *relationships*, that represent how entities – represented as nodes – relate to each other and the world [6]. Remember, the information presented in the figure amounts only to partial information about one disease, and it is merely a sliver of all the data that reside in model organism databases.

1.2 Biomedical data visualization

1.2.1 Graph-structured data

In *data visualization*, information is pictorially displayed for the purposes of sense-making, analysis, and communication [7]. As stated by Few (2014), "Important stories live in our data and data visualization is a powerful means to discover and understand these stories, and then to present them to others." [7].

Figure 2 hints that graph structures might be one promising motif for visualizations that involve many types of information [8]. How can they help us make sense of biomedical data? Figure 3 illustrates a problem into which my study attempts to gain some insight. Imagine that you have a potential *path* (a sequence of two or more nodes that are connected by edges) from a phenotype of interest to a disease of interest that is *incomplete* – and it is a path that would be complete if there were a known biological relationship between two particular human genes. It may be the case that the two genes are known to be in the same genetic or protein interaction pathway or that the mouse orthologs of the human genes exhibit similar mutant phenotypes or gene expression patterns. If scientists have the ability to picture the data in this way and within these contexts, this can potentially lead them to formulate the hypothesis that some critical functional relationship between the two human genes exists. In other words, this may lead them to start synthesizing and explaining the biology and developing a scientific narrative. This sort of functionality may be especially relevant to the study of rare and undiagnosed diseases, for which available data may be limited to phenotype descriptions and various candidate genomic variations [9].

This class of diseases is a major focus of the Monarch Initiative, a collaboration of groups at Oregon Health & Science University, the University of Pittsburgh, and several other institutions, which plans to deploy a web application featuring semantic and statistical model based tools for users to navigate, manipulate, and interpret model organism research data [10]. Their visual analytics environment will enable exploration of biomedical data that is organized in a graph database. HitWalker provides a more clinically oriented local example of a network visualization initia-

tive. It integrates genomic information with complementary functional analyses in order to prioritize sequence variants in cancer, and match specific therapies with patients [11]. Its visualization component is based on protein-protein networks overlaid with relevant metadata. Outside of OHSU, one can find graph-based tools like the gene function explorer GeneMANIA, which has a layout superficially similar to the prototype I present in this report [12]. They often use *force directed* layouts in which nodes are spaced about a central node with all of the edges being of equal length. As becomes evident in this study, the choice of layout is important because of its impact on the graph's interpretability.

1.2.2 The state of current tools

Current bioinformatics tools provide good support for querying and classifying relationships, but scientists might benefit greatly from future tools that facilitate more robust biological model-building and hypotheses generation – i.e., scientific "storytelling" [13]. In a study that concentrated on protein-protein interaction network tools, Mirel (2009) found that while tools did a decent job of supporting tasks related to finding relationships of interest (i.e., *exploratory* analysis), they were less adept at supporting tasks that let researchers gain "causal insights in systems biology analyses" (*explanatory* analysis) [13]. She identified potential features in bioinformatics visualizations, in terms of content, interactivity, and workspaces, that might enable investigators to do so [13]. They include several that might be found in graph-like interfaces and that informed some of the work in this study (viz. Section 2.3). Examples

include:

- Updated interactions across overlaid networks.
- Color-coding.
- Zooming, filtering, and node expansion or contraction.
- Inferences of significant sets of relationships based on other interactions.
- Weighted edges in a graph.
- Overlays of different interactions and associations.
- Indications of functional distinctions between types of entities.
- Nodes or edges that are statistically encoded.

It is also poorly understood how scientists use such tools to engage in "sense making," the iterative development of mental models based on the information that they have extracted [14, 15]. While exploring gene expression data from high throughput experiments for the purpose of uncovering functional relationships relevant to disease mechanisms, scientists were found by Mirel and Görg (2014) to:

> use interactive data visualizations and read deeply in the research literature. Little is known, however, about the actual flow of reasoning and behaviors (sense making) that scientists enact in this analysis, end-to-end. Understanding this flow is important because if bioinformatics tools are to be truly useful they must support it [15].

These shortcomings provide some fertile ground for exploring how well certain user interface elements in a visual analytics tool can facilitate explanatory analysis, gaining insight into users' mental processes as they use it, and – a vital outcome of this study, as discussed in Section 4.2 – locking down the interface requirements that a visual analytics tool must meet to be useful.

1.3 The study

1.3.1 Purpose and philosophy

The immediate purpose of this study is to assess how a visual analytics tool – i.e., an interactive visual interface that facilitates analytical reasoning [16] – can help biomedical scientists investigate the web of data and relationships centered on genes, genotypes, phenotypes, and diseases. Its broader purpose is to inform the bioinformatics community on productive processes and useful resources for making better tools in general. What philosophy should be pursued for the interface design process? Which software tools and templates are best suited to designing and implementing these tools? What are the evaluation methods that work, especially in situations where designers have limited time and resources to bring to bear?

This study led me to emphasize the *user* and elicit rich feedback from the kinds of scientists and students to whom the application will be geared. The study follows users' experiences and perceptions of two iterations of a prototype application. Manually constructed sketches and mockups were designed to indicate what the sys-

tem can do, while freely encouraging discussion and fresh ideas about how it can best facilitate scientists' needs. A functional software-based prototype served several purposes: 1) enabling a thorough usability evaluation of an early prototype based on time-tested methods; 2) investigating and adapting software libraries that would make the prototype visually appealing and responsive; and 3) providing a dynamic environment in which users can learn what they can or will be able to accomplish with the system, while encouraging further ideas and refining requirements for the eventual implementation of a tool beyond the conclusion of the study.

1.3.2 Discount usability and user-centered design

Discount usability engineering offered a promising approach for getting this study off the ground because of its "cheap, fast, and early focus on usability" [17, 18]. It consists of flexible, agile, easy-to-learn methods for assessing interface designs – centering on only a handful of test users who can be recruited in a short time – such as [17, 19, 20, 21, 22]:

- Storyboarding and sketches of use scenarios (viz. Sections 2.2.1 and 3.2.1).
- Narrowed-down prototypes containing a subset of features (viz. Section 2.3).
- Think-aloud usability evaluations (viz. Sections 2.7 and 3.3.2).
- "Quick and dirty" usability surveys (viz. Sections 2.7 and 3.3.2).
- Heuristic evaluation (viz. Section 2.7).

Application prototypes can vary in sophistication, from *low-fidelity* mockups produced by hand on paper or using drawing software (viz. Section 2.3.1) to a *high-fidelity* one consisting of a software application that is functional to some degree (viz. Section 2.3.2) [23]. Discount usability methods typically involve paper prototypes, but because the emphasis of this study is on *interactive* tools, it was desirable to create a functional prototype at some point and leverage these methods to evaluate it.

Ideas in contextual and scenario-based design give the proposed development and evaluation approaches a sturdy, *user-centered* framework. In the bioinformatics world, the Zebrafish Information Network – which today is one of many sources of animal model data in the Monarch system – was a pioneer of user-centered design and has shown that this approach can result in an accessible tool for widely-distributed research scientists [24].

In scenario-based design,

> Narrative descriptions of envisioned usage episodes are ... employed in a variety of ways to guide the development of the system that will enable these use experiences. Like other user-centered approaches, scenario-based design changes the focus of design work from defining system operations (i.e., functional specification) to describing how people will use a system to accomplish work tasks and other activities. However, unlike approaches that consider human behavior and experience through formal analysis and modeling of well-specified tasks, scenario-based design is a relatively lightweight method for envisioning future use possibilities [25].

This approach benefits the design process by emphasizing people and their experiences, directing "attention to the use-appropriateness of design ideas" and discouraging system designers from making premature commitments to early design ideas [25]. The relatively crude mockups developed for the first part of this study (viz. Section 3.2) were designed to avoid that trap.

In contextual design, user-centered design is composed of three primary layers, which dovetail nicely with storyboards, rough mockups, and iterative prototype development [22]. In Section 2.2.1 of the Methods, I specifically identify which parts of the present study correspond to these layers:

- *Practice design* – How does the user move through the system in the course of their activity? Use cases and use scenarios are developed to explore tasks and activities. Storyboards lend coherence to the user's experience.

- *Interaction design* – What is the layout of the screen and users' basic interaction with it? What is the design of the content being presented? Users can make sense of a structure that is clean and consistent.

- *Visual design* – What are the graphical elements, colors, animations, and the details of interaction? Good aesthetic design contributes to the user's experience. It should reinforce the purpose and structure of the system.

In the course of this report, I describe how the various discount engineering methods were applied to a genotype-phenotype visual analytics tool and assess the extent to which user-centered approaches helpfully informed development of the prototypes.

2. Methods

2.1 Overview

To examine how an interactive visual analytics tool can help biomedical scientists investigate genotype-phenotype relationships, I designed a qualitative usability study consisting of user-centered design and evaluation of two iterations of a prototype application. Figure 4 shows the flow of the work in the study. I first developed biological use cases that would provide inspiration for a small dataset involving typical biological connections between genotypes and phenotypes. I created a series of tasks for users to simulate using low-fidelity mockups of a visualization tool and then drew by hand a simple interface that illustrates possible outcomes of the tasks. Six test users were recruited for the study. After obtaining their feedback on the mockups, I developed a second-iteration, high-fidelity prototype application incorporating a set of interactive features supported by several open-source software libraries. A new set of user tasks was created for a usability evaluation. I met with the same six users to obtain their assessment of the prototype and elucidate future needs and requirements for a fully-developed application.

In short, these methods incorporate the user-centered, discount engineering approaches that I discussed in the Introduction. In this section, I lay out in detail the design and evaluation strategy involving the low- and high-fidelity iterations of the visual analytics tool (the "mockups" and "prototype," respectively). I then specify the users' characteristics and their tasks, present illustrations of the mockups and prototype, describe briefly the biological background for the use cases, describe data sources and software, and finally delineate the usability evaluation methods that were applied to the second-iteration prototype.

2.2 Design and evaluation strategy

2.2.1 Iterative design

I developed two iterations of a visual analytics tool for investigating and exploring genetic, phenotypic, and disease relationships in data provided by the Monarch Initiative. The user interface's visualization element consists of nodes and edges representing a biomedical *knowledge graph*. In order to finish the prototypes within the study's proposed three-month time frame, I employed a small subset of data, which is based on two general use cases, and several – but not all – of the relationship types supported by Monarch.

This iterative interface design and evaluation process is spelled out in detail in Table 1. For the first iteration, I created a preliminary set of manually constructed, low-fidelity mockups focusing on a few application features and met with users to

discuss, sketch, and refine use scenarios. These designs include simple searches for genes and diseases, a graph layout containing biological entities, and tabbed panels containing textual details about the entities. The mockups and tasks were designed to encourage discussion and do not focus on the functional specifics of how users perform actions in the interface. They essentially encompass the *practice* and *interaction* components of the contextual design process outlined in the Introduction, with the goal of creating a cohesive interface structure and user experience. For the second iteration, I developed a functional, interactive, web application – a high-fidelity prototype – based on the input acquired from the first round of user sessions and on recommendations from Monarch personnel, created concrete use cases for usability evaluation, and met again with users to assess the interface. The functional prototype was designed with a formal usability evaluation in mind, to assess the system's ease of use and its intuitiveness, and to further drive the formation of ideas and needs for future iterations and full implementation of the system. It touches heavily on the *visual* component of the contextual design process and it opened an avenue to try out software packages that might give the application visual appeal and high responsiveness to user actions.

2.2.2 Users and user sessions

Six users were recruited for the study:

- Two medical clinicians.
- Two biomedical students.

- Two biomedical researchers.

The users were volunteers who comprised a *convenience sample*; i.e., they were convenient to identify and recruit given limited available time and resources and they were chosen based on subjective criteria [26]. This quantity of users is appropriate given that the study focuses on an early stage of development, is qualitative in nature, and would not have benefited from a large number of users identifying the same problems over and over [23]. Although no firm consensus exists of the minimum number of users needed for an evaluation such as this, I initially aimed for a minimum of five and exceeded that by one [23].

I selected and invited these particular users because they constitute what I believed to be a typical cross-section of *primary users*, meaning users who will use this application directly (in contrast to *secondary users* or stakeholders, such as principal investigators, who might benefit from the system but not interact directly with it) [23]. I sought individuals whose profiles include at least an introductory college-level biology and genetics background, ample experience and competence in computer use, and the motivation to try out any available resources for exploring biomedical information of interest.

Visits with users typically lasted 50-60 minutes and consisted of oral discussion and a short written questionnaire. Discussions were recorded via note-taking and were not audio- or video-recorded. I met with each user individually. The two rounds of user sessions, corresponding to the two iterations of interface design and prototyping, occurred approximately one month apart.

I began my first visit with each user by giving a brief overview of the Monarch system and the data that it contains, explaining that it integrates information about genes, genotypes, animal models, phenotypes, and diseases. I asked users how they might look for information in such a system, without suggesting any specifics about visual presentation, sketching on paper some possibilities when necessary to drive the conversation.

Throughout the sessions, I sought to obtain from users their responses to several questions, via discussion before, during, and after their hands-on use of the interface and via a written questionnaire. They are centered on user requirements; user tasks and work; and simple usability rules, or heuristics (Table 2).

2.3 Mockups, prototype, and user tasks

2.3.1 First iteration

For the low-fidelity mockups, three user tasks (Tasks A-C, below) served as the basis of discussion and brainstorming. Users were instructed to interpret how they would carry out the activities outlined in Table 1, part I.

The Monarch system supports a multitude of relationship (edge) types. To reduce complexity and save time, I defined a limited set that would appear in the mockups:

- Gene-variant.
- Variant-genotype.
- Genotype-phenotype.

- Gene-disease.
- Phenotype-disease.
- Genetic interactions.
- Gene homology.

Task A The initial task is a simple one that begins with a zebrafish phenotype of interest and requires the user to interpret information about related zebrafish genes in a graph layout (Figures 5 to 10). The user can see information about the strength and type of evidence linking the phenotype to several genes. She can also apply a filter that displays only that gene node which has the strongest association with the phenotype, which in this case is based on how many publications report the link. The user is also encouraged to think of other ways in which she might wish to filter the nodes (for example, by gene family or by choosing more specific phenotypes). This series of mockups also introduces an interface element that would prove to be an important component of the functional prototype later: a contextual details panel occupying the right part of the display. Clicking on a node causes the panel to dynamically display information about related nodes (Figure 5) and a description (Figure 6). Clicking on an edge results in the panel showing the components of the relationship (Figure 7) and evidence (Figure 8). This panel provides ample space to display information about a graph element that does not legibly fit in the graph component itself.

Task B The next task involves making sense of different relationship types as users explore mouse Plexin family genes and try to discover links to cancer in humans (Figures 11 to 17). The concept of a *metanode* – a single node representing more than one node of the same type – comes into play as an attempt to prevent the graph display from becoming too cluttered (see Figure 12). The contextual details panel now contains controls for revealing related nodes of certain types in the graph display, including the option to place them in view as a single metanode (Figure 14). Figures 15 and 16 illustrate how this may be helpful. The PLXND1 gene has five genetic interactants, four of which are related to different diseases. Rather than painstakingly revealing each PLXND1 interactant one-by-one and then revealing its related disease, the user could take two quick actions to accomplish the same thing, first by placing a metanode representing the five interactants on the display and then commanding the interface to show all diseases related to any of those interactants. Once she identifies a disease of interest, she can then unveil paths back to the PLXND1 gene that might contain more entities of interest, such as a responsible genetic variant (Figure 17). In the course of this task, the user is encouraged to identify other kinds of information that would be helpful for deciding where to explore next in the graph (for example, a visual indication of how many nodes must be traversed in a potential path before reaching some data type of interest, such as a human disease). Furthermore, the tasks up to this point suggest that data will be visualized as a graph. The user is asked to consider alternative approaches to presenting the data.

Task C The final task involves making sense of different relationship types as users explore the human NF2 (a.k.a. MERLIN) gene and related variants and diseases – with the goal of identifying cancers related to these genes and forming ideas about underlying gene functions. The task begins in the same vein as the exploration of genetic interactants and diseases in the previous task, but this time resulting in the user uncovering a genetic variant related to a disease called neurofibromatosis, type 2 (Figure 18). Figures 19 to 21 illustrate subsequent exploration of related animal model phenotypes, a second gene (PAK1), and two types of cancer. The user is encouraged to infer information from the picture that has finally emerged. She is told that neurofibromatosis manifests itself as a number of tumor types. Based on that information as well as her own domain knowledge and the names of the diseases and phenotypes that she sees in the display, is it possible to discern some sort of biological connection between NF2 and PAK1? Is it helpful to explore animal genotypes and phenotypes like this, or does she prefer to focus on human genes and variants? Is there anything else that she wants to explore?

2.3.2 Second iteration

The high-fidelity prototype is a web application, viewable in a typical browser, that lets users explore data in a force-directed graph and view contextual information shown in a panel on the right side of the page. It thus mimics the general layout of the original mockups. The x and y coordinates of the nodes in the graph have no significance and can change from session to session. (For clarity's sake, the *graph* I

refer to in the prototype description means the visible collection of nodes and edges in the display, not the underlying knowledge graph or database.) The prototype may be freely downloaded from GitHub [27].

The prototype supports the data and relationship types summarized in Table 3.

At the top of the display, users can click icons to open a help dialog window (Figure 22), a legend for the graph visualization (Figure 23), and an application configuration box (not shown). The prototype does not include a fully featured search function, instead providing two canned queries that users can access from a drop-down menu. Figure 24 shows the prototype after a search for Plexin genes has been performed and Figure 25 shows contextual details for a gene which the user has clicked on and pinned to the display. In Figure 26, a right-click popup presents options for expanding the graph. Figure 27 shows contextual details for an edge. In Figures 28 and 29, the user has added individual nodes by clicking on them in the hierarchical view in the details panel. Figure 29 gives examples of every data type represented in the small dataset that fuels the prototype. Near the top of the interface users can find a "History" menu and a "Show only pinned" button, which enables the user to clear from the display any unpinned nodes.

Some features are partially implemented. A history of visited nodes, or bread-crumbs, is available, but no user actions can yet be performed on them. Users can drag and reposition nodes on the display, but only one at a time. The aforementioned hierarchical view is currently a stub; in a fully developed application, users would be able to drill down the hierarchy using the + icons next to each item. Edge strength is based only on number of curated publications, as it is in the original mockups; in

the Monarch system, edge type and weight determinations may be considerably more complex.

All told, the prototype supports concrete use cases that provided the basis for three user tasks (Tasks A-C, below) performed during the think-aloud usability evaluation (Section 2.7, below). All six users – the same individuals from the earlier sessions – were shown the exact same prototype; no modifications or improvements were made to the prototype or the planned tasks between one user session and the next, regardless of any feedback that was received. Users were directed to carry out the activities outlined in Table 1, part II, in the order presented below.

Preliminary task The user finds the help button and opens it. The system displays a help modal window.

Task A

1. The user finds the search menu and selects the first item, "plxnd1." The system initializes a force-directed graph consisting of two gene nodes and one edge.

2. The user pins the mouse Plxnd1 gene on the graphical display. The system applies a red outline to the node and displays gene information in a selection details panel.

3. The user opens a popup menu on the Plxnd1 node and chooses to display related genetic variants. The system adds all related genetic variant nodes and edges to the graph.

4. The user selects one of the edges (lines linking the gene and variants). The system displays relationship information and sources in the selection details panel.

5. The user opens one of the publications that are provided as sources. The system launches an outside webpage containing the publication.

6. The user elects to remove all of the variants from the display. The system removes all of the variant nodes from the graph.

7. The user elects to add to the display only the variant that has the most additional items related to it in the database. The system adds the variant to the graph.

Task B

1. The user reloads the display and repeats Task A steps 1 and 2, but this time pins the human PLXND1 gene on the display.

2. The user repeats Task A step 7, but this time sequentially adds the following to the display: a) a disease related to PLXND1, b) a phenotype related to the disease, c) a gene related to the disease, d) a mouse phenotype that is the equivalent of the human one, e) a genotype related to the mouse phenotype, and f) a variant related to the genotype. At the end of this user action sequence, the system displays a cyclic graph consisting of several nodes and edges.

3. The user continues to explore the data in undirected fashion, time permitting.

Task C

1. The user finds the configuration button and opens it. The system displays a configuration modal window.

2. The user elects to have descriptions of selected nodes appear in a contextual popup instead of the selection details panel. He or she closes the configuration window and opens a popup menu on any node. The system displays the description in the popup.

3. The user reloads the display, opens the configuration window, and elects to apply a color scale to edges in the graphical display. He or she closes the configuration window and displays variants related to the mouse Plxnd1 gene. The system varies the colors of the edges from black to red according to the number of sources supporting the relationship (i.e., the edge widths).

2.4 Biological use cases

The use cases explore two general classes of phenotypes that are associated with the hallmarks of cancer: induction of angiogenesis and proliferative signaling [28]. Semaphorins are axon guidance molecules that interact with Plexins and Neuropilins; they have been found to have a crucial role in the regulation of cancer progression and tumor angiogenesis [29]. Semaphorins are also involved in blood vessel defects such as truncus arteriosus [30]. Mutations in the NF2 gene (also known as MERLIN) have been implicated in thyroid cancer, mesothelioma, and melanoma, suggesting

that the normal gene functions as a tumor suppressor in many cell types; NF2 is a linker between transmembrane proteins and the actin cytoskeleton [31].

2.5 Data sources

The prototypes display subsets of the biomedical data found in the Monarch system. To create the knowledge graph, data were obtained with the Monarch Initiative's Dipper package, a data ingestion pipeline which generates Resource Description Framework (RDF) triples from public scientific resources, including [32]:

- NCBI Gene (gene descriptions) [33].
- Online Mendelian Inheritance in Man (human genes, phenotypes, and diseases) [34].
- ClinVar (human genomic variants) [35].
- BioGrid (genetic interactions) [36].
- PANTHER (gene orthology) [37].
- Human Phenotype Ontology [38].
- Mouse Genome Informatics (laboratory mouse genotypes and phenotypes) [4].
- Zebrafish Information Network (zebrafish genotypes and phenotypes) [5].

For the functional prototype, mouse gene descriptions were queried and manually curated from the Mouse Genome Informatics website [39]; human phenotype descriptions were curated from the Human Phenotype Ontology website [40].

2.6 Software

Storyboards and mockups were crafted in Balsamiq, a rapid wireframing software tool that mimics sketching on a whiteboard [41].

The prototype Genotype-Phenotype Explorer web application was built with several open-source JavaScript libraries: AngularJS, which provides the prototype with a model-view-controller architecture [42]; Data-Driven Documents (D3), for data visualization [43]; jQuery, for client-side scripting [44]; and Bootstrap, which provides templates for interface panels and buttons [45]. Data for the knowledge graph were obtained August 15-17, 2015, and were extracted from downloaded files via Sparql (SPARQL Protocol and RDF Query Language) queries and manual curation. Node colors in the visualization component were chosen from a color-blind palette [46]. The prototype has been tested and is functional in Chrome, Firefox, and Safari browsers running on Mac OS X 10.8 and later, and in Firefox on Microsoft Windows 7. To properly view it in Chrome, it must be served from a local web host. All application code and data are publicly available on GitHub [27].

2.7 Evaluation methods

Think-aloud usability study Each user navigated the functional prototype in a web browser while continuously "thinking out loud" – that is, verbalizing his or her thoughts during the process of carrying out certain tasks in the user interface. This protocol was designed to encourage users to describe the actions that they are taking

and to explain or interpret their actions.

Semi-structured interview This method comprised a short written survey and an oral debriefing interview with each user. General measures of the functional prototype's usability were taken with a standard usability questionnaire, the System Usability Scale (SUS), a ten-item Likert scale that provides a "global view of subjective assessments of usability" (Figure 30) [21]. The questionnaire was issued only after users had an opportunity to use the prototypes in the course of the think-aloud session, and users were prompted to record their immediate responses rather than taking a prolonged time to answer the questions. They were able to write additional comments on the form if they so wished.

Heuristic evaluation This method consists of broad rules of thumb that informed the questions in Table 2. It involves examining the interface and judging whether it is in compliance with recognized principles, and spotting usability problems that can be addressed during the iterative design process [20]. Nielsen (1995) identified 10 such heuristics [47]:

1. Visibility of system status.
2. Match between system and the real world.
3. User control and freedom.
4. Consistency and standards.
5. Error prevention.

6. Recognition rather than recall.

7. Flexibility and efficiency of use.

8. Aesthetic and minimalist design.

9. Help users recognize, diagnose, and recover from errors.

10. Help and documentation.

3. Results

3.1 Overview

In this section, I present the feedback and evaluation results from the two rounds of user sessions focusing on two iterations of a genotype-phenotype exploration tool. I break down the positive and negative critiques of the first-iteration mockups and list the interface features that users expressed their desire to have in a functional application. I then provide an overview of candidate features for the interactive prototype – which was coded after the first round of user sessions – based both on the users' feedback and on Monarch project recommendations. Finally, I once again break down the critiques and recommendations emanating from users' evaluation of the second-iteration prototype and present usability scores for the user interface. The results coming out of this study should inform subsequent development of a visual analytics tool, as I explain in the Discussion section later. Detailed results are shown in Tables 4 to 9.

3.2 The mockups

3.2.1 Scenario development and interface features

Before embarking on this study, I discussed with Monarch project personnel (M. Haendel, H. Hochheiser, N. Washington, pers. comm., April 23, 2015) interesting potential features that could be well integrated in some prototypes. Altogether, it was desirable to achieve a visualization environment consisting of nodes and edges that users would be able to explore. At a high level, interactive features should enable users to start at a node of interest, reveal related nodes as they explore the graph, and filter nodes based on data or relationship types. Users should be able to anchor or "pin" nodes to the display and have access to a history of their actions.

Another significant feature to consider was the use of edges' visual properties to indicate the strength of association between nodes. Users should be able to filter related nodes accordingly in the course of exploration. The main types of edges in the Monarch system are directly asserted edges based on curated data and inferred edges based on similarity algorithms.

Before meeting with users, I created a series of mockups (Figures 5 to 21), covering three tasks, that would enable a preliminary evaluation of these features and that would encourage general discussion and brainstorming about the central question of this study: how an interactive visual tool can help biomedical scientists investigate and analyze the aforementioned relationships. The mockups rest on the simplest sorts of search scenarios which I anticipated users would want to carry out: identifying a

phenotype or gene of interest and placing that piece of information in the visualization environment. The use cases in general involve starting with a phenotype or gene and then exploring connections to human sequence variants and cancer.

Before reviewing the mockups, I gave each user a brief description of the Monarch system and sought some insight into how they would begin using such a system if they had come across it for the first time. Users thought that the following would be desirable ways to start finding information:

- Starting with one piece of information of interest and inspecting related pieces information.
- Starting with two pieces of information of interest and exploring all of the possible links between the two.

These simple scenarios are sketched in Figure 31. The mockups and prototype generally enable users to perform the first of these two scenarios.

3.2.2 User feedback and requests

Upon introduction to the mockups, users attempted to do the three tasks which were described in the Methods section (Table 1, "Iteration I"). User feedback is summarized in Table 4; the various user comments and responses are grouped thematically and roughly in the order in which they arose in the discussion, and each is color-coded as either a favorable or unfavorable critique. The table indicates which individual users – the two clinicians, two students, and two researchers – provided each piece of feedback; several critiques were repeated by multiple users.

Overall, there is little apparent difference in responses between the clinician, student, and researcher groups, although the sample size of each is admittedly tiny.

A few notable observations:

Users immediately grasped what the nodes and edges mean and found such an interface to be intuitive. They grasped the relationship between the graph display and the accompanying contextual detail panel. They also expressed concerns that the display would quickly become overwhelming and unreadable as new nodes are added. Most had some trouble interpreting text labels in the mockups that may have used verbiage that was too jargonish or technical (e.g., human and mouse gene abbreviations).

When shown the simple metric for strength of evidence based on the number of curated publications, users made sense of it and liked it. There was, however, less interest in – and less understanding of – being able to differentiate between different kinds of evidence (e.g., asserted vs. curated, or electronic vs. curated) and having a way to favor one over the others in the course of exploring data.

Several users offered quite detailed ideas on alternate ways to display entities and relationships. These include:

- Different shapes for different species.
- Different colors for different species (shapes might be difficult to distinguish).
- Different colors for different phenotypes.
- Different edge styles (dotted, dashed, etc.) for different species.

- A color scale indicating "hotness" that is applied to edges to indicate strength of association (varying widths alone might be difficult to distinguish).

- Variations in node textures and borders to convey information.

The last of the three tasks (Figures 18 to 21) attempted to prompt users to make biological inferences concerning gaps in the data. Only one user correctly interpreted that the human NF2 and PAK1 genes interact; the other users did not perceive the interface as being helpful in this regard.

In addition to their critiques of the interface, users chimed in often about features that they would like to see in a functional application. These are listed in Table 5 and include several interactive features as well as configurability options. The student group in particular favored robust configurability of how nodes and edges are displayed, perhaps even including alternatives to the force-directed graph layout itself.

One student sketched alternatives layouts which he felt would be desirable avenues for inspecting relationships. They are reproduced in Figure 32:

- A rigid tree view of nodes and edges that can be expanded left-to-right or top-to-bottom from a single node.

- A treemap view, in which squares representing entities are nested in one another and can be clicked to zoom in and out of the data.

- A hierarchical view, in which an element in a breadcrumb trail can be inspected by drilling down into its child or related entities.

Three users said that a text-based hierarchical or ontological tree should appear side-by-side with the graph-based layout, perhaps in the right-side panel, giving users two complementary and connected ways of viewing. Users could then drill down the hierarchy or ontology and then jump from a node within it to a graph view of the node. This would be especially useful in cases where a node is linked to a great many nodes of some type, such as gene interactants or sequence variants. Perhaps users could see a single metanode representing a batch of interactants or variants in the graph display, while an accompanying panel would list the interactants or variants individually and provide immediate access to useful data. This piece of feedback would evolve later into a vital component of the functional prototype.

3.3 The functional prototype

3.3.1 Feature selection

After the first round of user sessions was complete, I considered the feedback and decided which interactive and visual features could be built into a functional web-based prototype in the one month period before the second round of evaluations was scheduled to begin. I also pared down slightly the set of data and relationship types that the prototype would support (Table 3). All candidate interface features are shown in Table 6 along with an indication of whether they ultimately were included in the prototype (Figures 24 to 29). It is important to point out that these decisions should not be taken as endorsements or the lack thereof of a particular feature; several

crucial and enormously desirable application elements were left out or exist only as stubs in the prototype because of their complexity and the unfeasibility of fully implementing them in a brief time frame.

3.3.2 Usability evaluation

Think-aloud study Upon opening the prototype application in a web browser, users attempted to perform the three tasks which were described in the Methods section (Table 1, "Iteration II"). Users were prompted to talk aloud as they navigated the interface and carried out tasks. Their feedback provides insight into the application's usability in its current state as well as their needs and desires for future iterations and, eventually, a full-fledged bioinformatics tool. Feedback is summarized in Table 7; it is organized the same way as in the earlier user feedback table.

Some general key observations:

- Users were eager to try out the application's features without much prompting.
- They tended to remember how to do tasks based on what they had done earlier.
- The integration between the graph visualization and the details panel was obvious to them.
- Users liked the use and choice of colors to represent different relationship types.
- They quickly grasped how to add nodes to the display, using either the graph's interactive elements or the details panel.

- Information about relationships (edges) was not as intuitive as contextual information concerning nodes.
- Users had little trouble finding links to outside resources (i.e., the Monarch site and PubMed).
- They were unenthusiastic about the configuration options (viz. Task C).
- Although users skimmed the help dialog at the start, they tended not to go back to it if they got stuck.

On a higher level, users liked the clear biological connections that they saw and enjoyed being able to trace a path. While some users envisioned this application primarily as a solid learning and reference tool, others believed that having such clear biological connections and the ability to visually trace a path through them would help researchers talk out scenarios and solve problems. They provided a significant set of ideas and requirements, listed in Table 8, for subsequent iterations of the application. Some of these items reiterate feature requests that came out of the earlier mockup evaluations, emphasizing their importance for future inclusion.

System Usability Scale Each user filled out a System Usability Scale (SUS) questionnaire following the think-aloud evaluation. Individual responses to all questions are shown in Table 9. Following are the average qualitative responses to the ten questions on a scale of "strongly disagree" to "strongly agree"; for all but the final question, the mean response was slightly or generally favorable.

1. I think that I would like to use this system frequently: **slightly agree.**

2. I found the system unnecessarily complex: **disagree.**

3. I thought the system was easy to use: **agree.**

4. I think that I would need the support of a technical person to be able to use this system: **disagree.**

5. I found the various functions in this system were well integrated: **agree.**

6. I thought there was too much inconsistency in this system: **slightly disagree.**

7. I would imagine that most people would learn to use this system very quickly: **agree.**

8. I found the system very cumbersome to use: **slightly disagree.**

9. I felt very confident using the system: **agree.**

10. I needed to learn a lot of things before I could get going with this system: **slightly agree.**

Overall scores for the ten questions were converted to a scale of 0 to 100. The average score for the six respondents was 72, ranging from 60 to 82.5. A score above 68 is considered an above average usability score (Figure 33) [48].

4. Discussion

4.1 User-centered discount methods in practice

User-centered interface design and evaluation is an elaborate, multistep process involving identification of stakeholders and users, requirements analysis, conceptual design, the choice of software components, integration of those components, and usability evaluation – and many of these steps, especially integration, evaluation, and the refinement of requirements, can churn through many iterations before achieving a satisfactory result [23]. Furthermore, these steps can be undertaken with a variety of methods from which to choose. This study's central *challenge* was to compress many of these development phases into a time frame that realistically allowed only two iterations of a prototype and brief evaluation sessions that could fit into biomedical students' and researchers' busy schedules. Nevertheless, this study's central *outcome* was achieved: it provides insight into how an interactive visual analytics tool can help scientists investigate interconnected biomedical data and it should meaningfully inform future iterations of the application and its eventual integration into the Monarch system.

The strategy of developing two prototype iterations – one low-fidelity and one high-fidelity – and the application of discount evaluation techniques all proved quite effective. All six users were comfortable offering critiques and recommendations and they provided, for the most part, a remarkable amount of feedback on the system. Low-fidelity mockups seemed to enhance their comfort. Users could see that manually drawn layouts did not require countless hours of coding work to prepare, which might have made them reticent to criticize something – a pitfall that lightweight scenario-based design is indeed designed to avoid [25].

The second prototype offered enough interactivity and visual appeal to immediately engage users. System Usability Scale assessments affirmed that the system, despite being just the first software-based version of the application, meets an acceptable usability standard. Most users agreed that at the least, the application has promise as a biomedical learning and reference resource. Two of them went a step further and concluded that a fully developed tool would indeed help researchers talk out biological scenarios, solve problems, and identify gaps in knowledge needing further investigation.

One stubborn shortcoming of the study is that it generated a rather perfunctory and vague sense of how users would query information in the system when first encountering it and not knowing much about it. Some brief sketches and storyboarding helped (viz. Figure 31) but did not generate a lot of concrete ideas. With meeting times short, I moved quickly to the pre-prepared mockups, which essentially amounted to advanced storyboards, and these rapidly spurred a great deal of discussion. In the course of the tool's future development, it will be necessary to think of ways to im-

mediately engage new cohorts of users and evaluators and rapidly convey the benefits of the system.

4.2 Future requirements

This study produced a rich set of requirements for future iterations and eventual implementation of a genotype-phenotype exploration tool. The early mockups generated a lot of ideas, some of which were tested in the functional prototype and some not. A few key feature requests were made repeatedly by users throughout the study but were too complex to fully or even partially implement. In my view, however, they are essential:

Search The functional prototype lacks a proper search interface. Users should be able to seed the visualization with one or more nodes – found via a typical search box and auto-suggest function – and any links between them. This should be coordinated with user-defined filters that can restrict the display to certain data and relationship types and specific edge strength thresholds.

Hierarchy and path navigation An expandable, text-based hierarchical or ontological display of Monarch entities should complement and be tightly knit with the interface's visualization component. This need was completely unanticipated before the first round of user sessions. Even though it was implemented only as a stub in the functional prototype, users were immediately drawn to it and several used it as the primary means for adding nodes to the graph. The hierarchy and graph components

should both contain powerful and consistent visual indications of the richness of data that a user would find when exploring in a given direction.

Configurability The Monarch knowledge base contains more data types and information about far more species than the functional prototype supports (five and two, respectively), which puts limits on adapting the prototype's relatively simple node shapes and color palette to a full-scale tool. Rather than arbitrarily assigning a plethora of shapes and colors to data elements, the tool may better serve users by applying a few basic default shapes and colors to graphical elements and allowing users to tailor these to their liking. For example, all nodes could appear as circles by default, while users can access a configuration panel to render human entities as triangles and mouse entities as squares, or alternatively to distinguish particular data types by shape. Colors could be assigned to particular combinations of species and data types using color-palette widgets. Applying different textures may be another option. Edges in the graph may be configured to appear with various line styles (dotted, dashed, etc.) and in various colors, as well.

Alternate visualizations Based on my observations of users and my own experience using the prototype, the force-directed graph layout of the visualization component has clear benefits: it is aesthetically pleasing and extremely versatile in allowing users to drag, pin, and organize nodes however they like in the display. This is crucial for users who, as envisioned, will be using the system to formulate literal pictures of biological connections and reveal gaps in knowledge. One drawback is that node

text labels quickly tend to overlap other visual elements, a particular problem with biomedical data because of the extremely long names of some items (typical is the mouse genotype officially called "Plxnd1$^{tm1.1Tmj}$/Plxnd1$^{tm1.1Tmj}$; Tg(Tek-cre)1Ywa/0 [involves: 129S/SvEv * 129S4/SvJaeSor * C57BL/6 * SJL]"). Therefore, the tool could offer configurable alternatives to the force-directed graph layout, as sketched in Figure 32, in which labels might render more legibly. D3 templates are available for promising options like cluster and zoomable partition layouts (Figures 34 and 35) [49, 50].

4.3 Future considerations

In addition to the essential requirements, a few aspects of the tool merit special consideration before marching onward with future iterations:

Edge strength The prototypes use a very simple metric to determine "strength of association" between nodes, and most users were comfortable with it. One user, interestingly, interpreted edge width between a gene and variants as the biological *impact* of the mutation, i.e., the relative severity of the resulting phenotype. There is a need to probe deeper into how edge widths can meaningfully inform and guide users in their exploration.

Phenotype and disease representation A few users recognized the unique challenge in traversing the abundance of concepts in phenotype and disease ontologies and their super- and subclasses while avoiding a spaghetti-like tangle of terms. For this

task, a future prototype might explore the benefits of treemaps (the middle "nested rectangle" sketch in Figure 32), using colors or size to indicate the richness of data associated with a concept [51]. Whatever the approach, great care must be taken not to overwhelm users with phenotype information.

Speed and performance Application performance was not evaluated in the functional prototype. Because of time constraints, no attempt was made to undergird the application with a proper graph data store and query language such as Neo4j and Cypher [52, 53]. Upon loading into a browser, the prototype's complete dataset, comprising only 67 entities and 92 relationships, is read into the application from curated JSON (JavaScript Object Notation) files. The small amount of data is no hindrance to performance and the system responded to users' input instantaneously. Scaling the application to the Monarch system's vast data store will require robust graph query performance that allows the application to maximally benefit from the responsiveness, efficiency, and low latency of D3 visualization [54].

4.4 The takeaways

What important lessons for the bioinformatics community have come from this study? First, it validated several user interface features (viz. Section 1.2.2) which have been proposed to enhance explanatory analysis of highly interconnected biomedical data [13]. Some of the features listed in the Mirel (2009) article evolved into essential core elements of my functional, interactive prototype: color coding of nodes, filtering

and node expansion, and weighted edges. User feedback alluded to others in the list at least tangentially: overlaying networks or other biological information, inferring or displaying alternate pathways and associations, and statistically encoding the visual properties of nodes and edges. Bringing to life certain coordinated features that scientists have long desired in bioinformatics tools has produced encouraging results.

Crucially, this study also demonstrated that user-centered design and evaluation is a productive approach – and, I believe, still an under-appreciated and underused one – and for this tool it was a better alternative to creating a tool that forced users to adapt to it rather than the other way around. Second, I did not fully appreciate at the outset the sheer number of software libraries and templates that are freely available and that allow one to quickly piece together interactive tools that are visually appealing and instantaneously responsive to user actions. That latter point is a key to making software that is enjoyable to use and that users will come back to use again. The choices I made for constructing the prototype – D3 and AngularJS in particular – are proven libraries that should allow the application to be scaled up to a full-blown informatics tool. Finally, the discount usability methods were extremely effective in evaluating the prototype and eliciting needs and priorities for subsequent development. I recommend them for those of us who lack the time and resources to carry out exhaustive, detailed usability studies or who find such a task rather intimidating. With these in our arsenal, we can realize the potential of our tools to benefit scientists.

5. Conclusion

Bioinformatics visualization tools abound. Developing one that can immediately engage users and succeed as a bona fide research tool rather than merely being another face in the crowd is a stiff challenge. This study has rested on the belief that a profound emphasis on user-centered design and evaluation is a fruitful way to meet it. Early prototypes and user scenarios were built with flexibility and freedom in mind, and they succeeded overall in spurring users to brainstorm ideas in great quantity and with considerable quality, articulate their needs as researchers and clinicians, and suggest sometimes unorthodox approaches to improving the application. There is little doubt that this tool for investigating genotypes, phenotypes, and diseases can be a very good reference and learning tool, but subsequent work is needed to scale up the prototype to a fully functional tool with robust user configurability and a variety of ways to visualize biological relationships. Its future as a problem-solving visual analytics tool depends on the ongoing commitment to embracing as many users with as diverse a biomedical background as possible at the heart of the creative process.

Bibliography

[1] Frederick B. Churchill. William Johannsen and the genotype concept. *Journal of the History of Biology*, 7(1):5–30, 1974.

[2] Andrew R. Deans, Suzanna E. Lewis, Eva Huala, Salvatore S. Anzaldo, Michael Ashburner, James P. Balhoff, David C. Blackburn, Judith A. Blake, J. Gordon Burleigh, Bruno Chanet, Laurel D. Cooper, Mélanie Courtot, Sándor Csösz, Hong Cui, Wasila Dahdul, Sandip Das, T. Alexander Dececchi, Agnes Dettai, Rui Diogo, Robert E. Druzinsky, Michel Dumontier, Nico M. Franz, Frank Friedrich, George V. Gkoutos, Melissa Haendel, Luke J. Harmon, Terry F. Hayamizu, Yongqun He, Heather M. Hines, Nizar Ibrahim, Laura M. Jackson, Pankaj Jaiswal, Christina James-Zorn, Sebastian Köhler, Guillaume Lecointre, Hilmar Lapp, Carolyn J. Lawrence, Nicolas Le Novère, John G. Lundberg, James Macklin, Austin R. Mast, Peter E. Midford, István Mikó, Christopher J. Mungall, Anika Oellrich, David Osumi-Sutherland, Helen Parkinson, Martín J. Ramírez, Stefan Richter, Peter N. Robinson, Alan Ruttenberg, Katja S. Schulz, Erik Segerdell, Katja C. Seltmann, Michael J. Sharkey, Aaron D. Smith, Barry Smith, Chelsea D. Specht, R. Burke Squires, Robert W. Thacker, Anne

Thessen, Jose Fernandez-Triana, Mauno Vihinen, Peter D. Vize, Lars Vogt, Christine E. Wall, Ramona L. Walls, Monte Westerfeld, Robert A. Wharton, Christian S. Wirkner, James B. Woolley, Matthew J. Yoder, Aaron M. Zorn, and Paula Mabee. Finding our way through phenotypes. *PLoS Biology*, 13(1):e1002033, 2015.

[3] Danielle Simmons. The use of animal models in studying genetic disease: transgenesis and induced mutation. *Nature Education*, 1(1):70, 2008.

[4] Judith A. Blake, Carol J. Bult, James A. Kadin, Joel E. Richardson, and Janan T. Eppig. The mouse genome database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Research*, 39(Suppl. 1):842–848, 2011.

[5] Judy Sprague, Leyla Bayraktaroglu, Yvonne Bradford, Tom Conlin, Nathan Dunn, David Fashena, Ken Frazer, Melissa Haendel, Douglas G. Howe, Jonathan Knight, Prita Mani, Sierra A. T. Moxon, Christian Pich, Sridhar Ramachandran, Kevin Schaper, Erik Segerdell, Xiang Shao, Amy Singer, Peiran Song, Brock Sprunger, Ceri E. Van Slyke, and Monte Westerfield. The Zebrafish Information Network: the zebrafish model organism database provides expanded support for genotypes and phenotypes. *Nucleic Acids Research*, 36:D768–D772, 2008.

[6] David Easley and Jon Kleinberg. Graphs. In *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010.

[7] Stephen Few. Data visualization for human perception. In Mads Soegaard and Rikke Friis Dam, editors, *The Encyclopedia of Human-Computer Interaction*. Interaction Design Foundation, Aarhus, Denmark, 2014.

[8] Ian Robinson, Jim Webber, and Emil Eifrem. *Graph Databases*. O'Reilly, Sebastopol, Cal., 2013.

[9] The Undiagnosed Diseases Program. `http://www.genome.gov/27544402`. Cited November 20, 2015.

[10] The Monarch Initiative. `http://monarchinitiative.org`. Cited November 20, 2015.

[11] Daniel Bottomly, Beth Wilmot, Jeffrey W. Tyner, Christopher A. Eide, Marc M. Loriaux, Brian J. Druker, and Shannon K. Mcweeney. HitWalker: variant prioritization for personalized functional cancer genomics. *Bioinformatics*, 29(4):509–510, 2013.

[12] David Warde-Farley, Sylva L. Donaldson, Ovi Comes, Khalid Zuberi, Rashad Badrawi, Pauline Chao, Max Franz, Chris Grouios, Farzana Kazi, Christian Tannus Lopes, Anson Maitland, Sara Mostafavi, Jason Montojo, Quentin Shao, George Wright, Gary D. Bader, and Quaid Morris. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research*, 38(Suppl. 2):214–220, 2010.

[13] Barbara Mirel. Supporting cognition in systems biology analysis: findings on

users' processes and design implications. *Journal of Biomedical Discovery and Collaboration*, 4:2, 2009.

[14] Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*, 2005.

[15] Barbara Mirel and Carsten Görg. Scientists' sense making when hypothesizing about disease mechanisms from expression data and their needs for visualization support. *BMC Bioinformatics*, 15:117, 2014.

[16] Pak Chung Wong and Jim Thomas. Visual analytics. *Computer Graphics and Applications, IEEE Transactions on*, 24(5):20–21, 2004.

[17] Jakob Nielsen. Usability engineering at a discount. In *Proceedings of the Third International Conference on Human-Computer Interaction: Designing and Using Human-Computer Interfaces and Knowledge Based Systems*, 1989.

[18] Jakob Nielsen. Guerilla HCI: using discount usability engineering to penetrate the intimidation barrier. Nielsen Norman Group, 1994.

[19] Jakob Nielsen. Thinking aloud: the #1 usability tool. Nielsen Norman Group, 2012.

[20] Jakob Nielsen. Heuristic evaluation. In Jakob Nielsen and Robert L. Mack, editors, *Usability Inspection Methods*. John Wiley & Sons, New York, 1994.

[21] John Brooke. SUS: A 'quick and dirty' usability scale. In Patrick W. Jordan,

Bruce Thomas, Bernard A. Weerdmeester, and Ian L. McClelland, editors, *Usability Evaluation in Industry*. Taylor & Francis, London, 1996.

[22] Karen Holtzblatt and Hugh Beyer. Detailed design and validation. In *Contextual Design: Evolved*, Synthesis Lectures on Human-Centered Informatics No. 24. Morgan & Claypool, San Rafael, Cal., 2015.

[23] Debbie Stone, Caroline Jarrett, Mark Woodroffe, and Shailey Minocha. *User Interface Design and Evaluation*. Morgan Kaufmann, San Francisco, 2005.

[24] Eckehard Doerry, Sarah Douglas, Arthur E. Kirkpatrick, and Monte Westerfield. User-centered design for widely-distributed scientific communities. Technical report, University of Oregon, Eugene, Ore., 1997.

[25] Mary Beth Rosson and John M. Carroll. Scenario-based design. In Julie A. Jacko and Andrew Sears, editors, *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. Lawrence Erlbaum Associates, New York, 2nd edition, 2002.

[26] Mike Battaglia. Convenience sampling. In Paul J. Lavrakas, editor, *Encyclopedia of Survey Research Methods*. SAGE Publications, Thousand Oaks, Cal., 2008.

[27] Genotype-Phenotype Explorer. `https://github.com/seger/gpex/releases/tag/v0.1`. Cited November 20, 2015.

[28] Douglas Hanahan and Robert A. Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674, 2011.

[29] Chenghua Gu and Enrico Giraudo. The role of semaphorins and their receptors in vascular development and cancer. *Experimental Cell Research*, 319(9):1306–1316, 2013.

[30] Leonard Feiner, Andrea L. Webber, Christopher B. Brown, Min Min Lu, Li Jia, Paul Feinstein, Peter Mombaerts, Jonathan A. Epstein, and Jonathan A. Raper. Targeted disruption of semaphorin 3C leads to persistent truncus arteriosus and aortic arch interruption. *Development*, 128(16):3061–3070, 2001.

[31] Ivan Stamenkovic and Qin Yu. Merlin, a 'magic' linker between extracellular cues and intracellular signaling pathways that regulate cell motility, proliferation, and survival. *Current Protein & Peptide Science*, 11(6):471–484, 2010.

[32] Data Ingestion Pipeline for SciGraph. `https://github.com/monarch-initiative/dipper`. Cited November 20, 2015.

[33] Gene – NCBI. `http://www.ncbi.nlm.nih.gov/gene`. Cited November 20, 2015.

[34] Ada Hamosh, Alan F. Scott, Joanna S. Amberger, Carol A. Bocchini, and Victor A. McKusick. OMIM.org: Online Mendelian Inheritance in Man

(OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research*, 33(Database Iss.):1–10, 2005.

[35] Melissa J. Landrum, Jennifer M. Lee, George R. Riley, Wonhee Jang, Wendy S. Rubinstein, Deanna M. Church, and Donna R. Maglott. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, 42(D1):980–985, 2014.

[36] Chris Stark, Bobby-Joe Breitkreutz, Teresa Reguly, Lorrie Boucher, Ashton Breitkreutz, and Mike Tyers. BioGRID: a general repository for interaction datasets. *Nucleic Acids Research*, 34:D535–D539, 2006.

[37] Huaiyu Mi, Anushya Muruganujan, and Paul D. Thomas. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, 41(D1):377–386, 2013.

[38] Sebastian Köhler, Sandra C. Doelken, Christopher J. Mungall, Sebastian Bauer, Helen V. Firth, Isabelle Bailleul-Forestier, Graeme C. M. Black, Danielle L. Brown, Michael Brudno, Jennifer Campbell, David R. Fitzpatrick, Janan T. Eppig, Andrew P. Jackson, Kathleen Freson, Marta Girdea, Ingo Helbig, Jane A. Hurst, Johanna Jähn, Laird G. Jackson, Anne M. Kelly, David H. Ledbetter, Sahar Mansour, Christa L. Martin, Celia Moss, Andrew Mumford, Willem H. Ouwehand, Soo Mi Park, Erin Rooney Riggs, Richard H. Scott, Sanjay Sisodiya, Steven Van Vooren, Ronald J. Wapner, Andrew O. M. Wilkie, Caroline F. Wright, Anneke T. Vulto-Van Silfhout, Nicole de Leeuw,

Bert B. A. de Vries, Nicole L. Washingthon, Cynthia L. Smith, Monte Westerfield, Paul Schofield, Barbara J. Ruef, Georgios V. Gkoutos, Melissa Haendel, Damian Smedley, Suzanna E. Lewis, and Peter N. Robinson. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Research*, 42(D1):966–974, 2014.

[39] MGI – Mouse Genome Informatics. `http://www.informatics.jax.org`. Cited November 20, 2015.

[40] The Human Phenotype Ontology. `http://human-phenotype-ontology.github.io`. Cited November 20, 2015.

[41] Balsamiq. `https://balsamiq.com`. Cited November 20, 2015.

[42] AngularJS. `https://angularjs.org`. Cited November 20, 2015.

[43] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3: Data-Driven Documents. *Visualization and Computer Graphics, IEEE Transactions on*, 2011.

[44] jQuery. `https://jquery.com`. Cited November 20, 2015.

[45] Bootstrap. `http://getbootstrap.com`. Cited November 20, 2015.

[46] Martin Krzywinski, Inanc Birol, Steven Jones, and Marco Marra. Getting into visualization of large biological data sets: 20 imperatives of information design, 2012. Poster presented at the 2nd IEEE Symposium on Biological Data Visualization (BioVis 2012), Seattle, Wash.

[47] Jakob Nielsen. 10 usability heuristics for user interface design. Nielsen Norman Group, 1995.

[48] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual SUS scores mean: adding an adjective rating scale. *Journal of Usability Studies*, 4(3):114–123, 2009.

[49] Cluster Layout. `https://github.com/mbostock/d3/wiki/Cluster-Layout`. Cited November 20, 2015.

[50] Partition Layout. `http://mbostock.github.io/d3/talk/20111018/partition.html`. Cited November 20, 2015.

[51] Gene Ontology with Treemap. `http://www.cs.umd.edu/hcil/bioinfovis/treemap.shtml`. Cited November 20, 2015.

[52] Neo4j. `http://neo4j.com`. Cited November 20, 2015.

[53] Cypher Introduction. `http://neo4j.com/docs/stable/cypher-introduction.html`. Cited November 20, 2015.

[54] Zhicheng Liu and Jeffrey Heer. The effects of interactive latency on exploratory visual analysis. *Visualization and Computer Graphics, IEEE Transactions on*, 2014.

[55] MGI – Mouse Genome Informatics – Sod1 Gene Detail, The Jackson Laboratory, Bar Harbor, Me. http://www.informatics.jax.org. Cited November 20, 2015.

[56] OMIM – Online Mendelian Inheritance in Man. http://omim.org. Cited November 20, 2015.

Figures

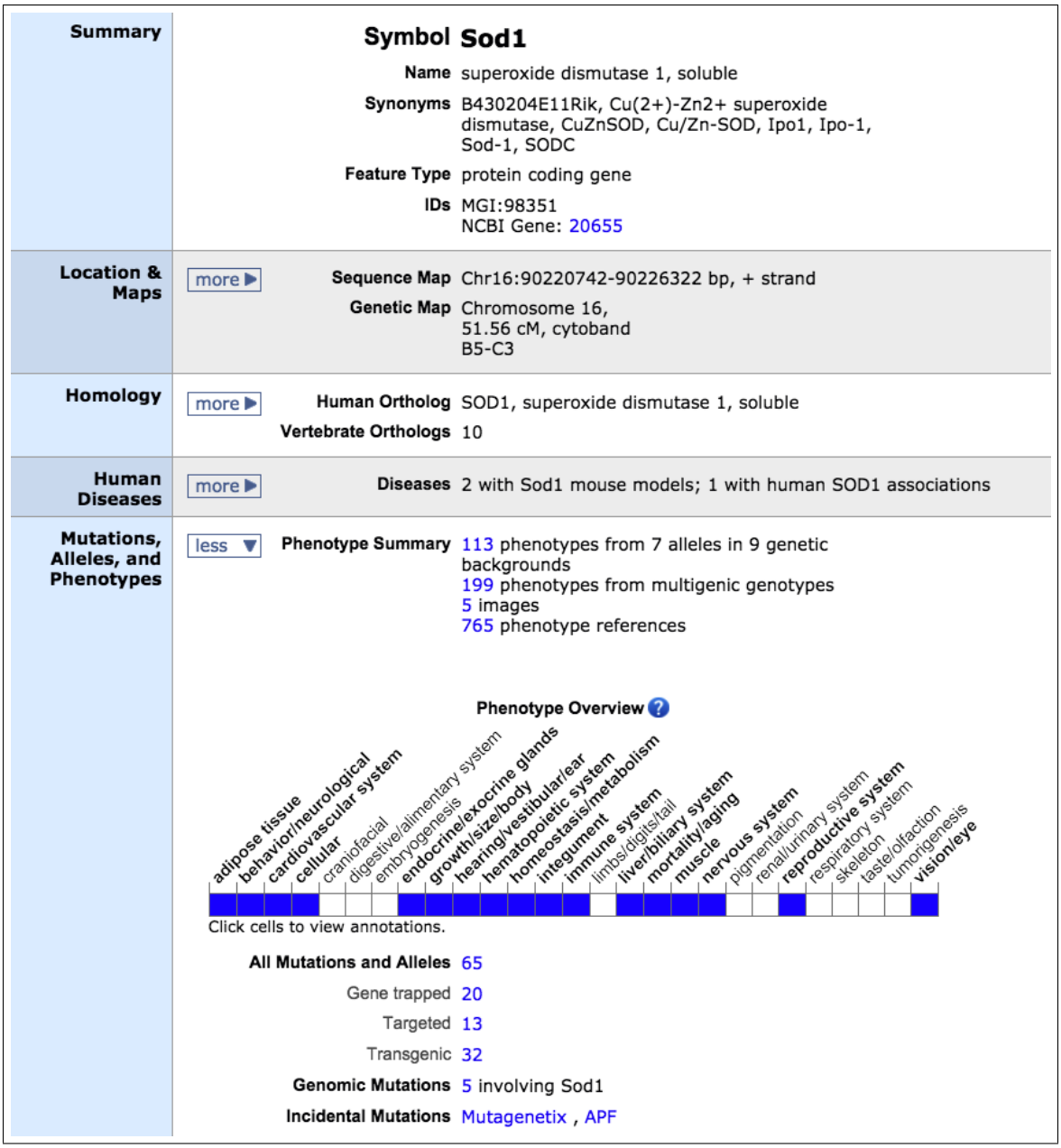

Figure 1: Some sequence and phenotype data on a typical Mouse Genome Informatics gene page [55].

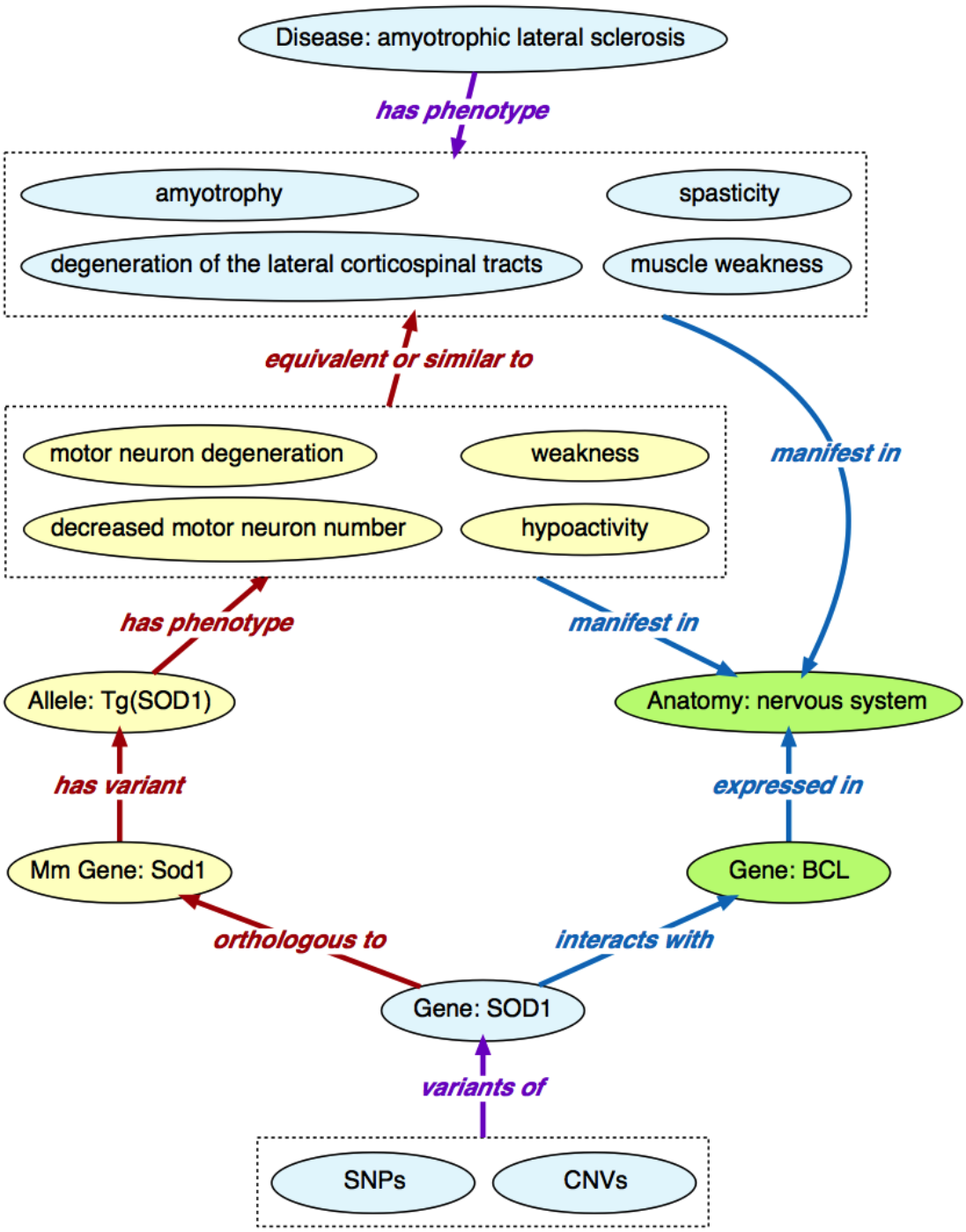

Figure 2: A graph that integrates partial human and animal model data about a disease (ALS) and related genotypes and phenotypes [39, 56]. Human data are indicated by light blue nodes, mouse data by yellow nodes, and mammalian (human or mouse) data by green nodes. Multiple paths are possible between the human and mouse phenotypes (near top) and the human genotype (bottom). Boxes bound by dotted lines contain groups of similar nodes, or *metanodes*. Arrows indicate how each relationship should be read from one node to the other (e.g., "The BCL gene is expressed in the nervous system").

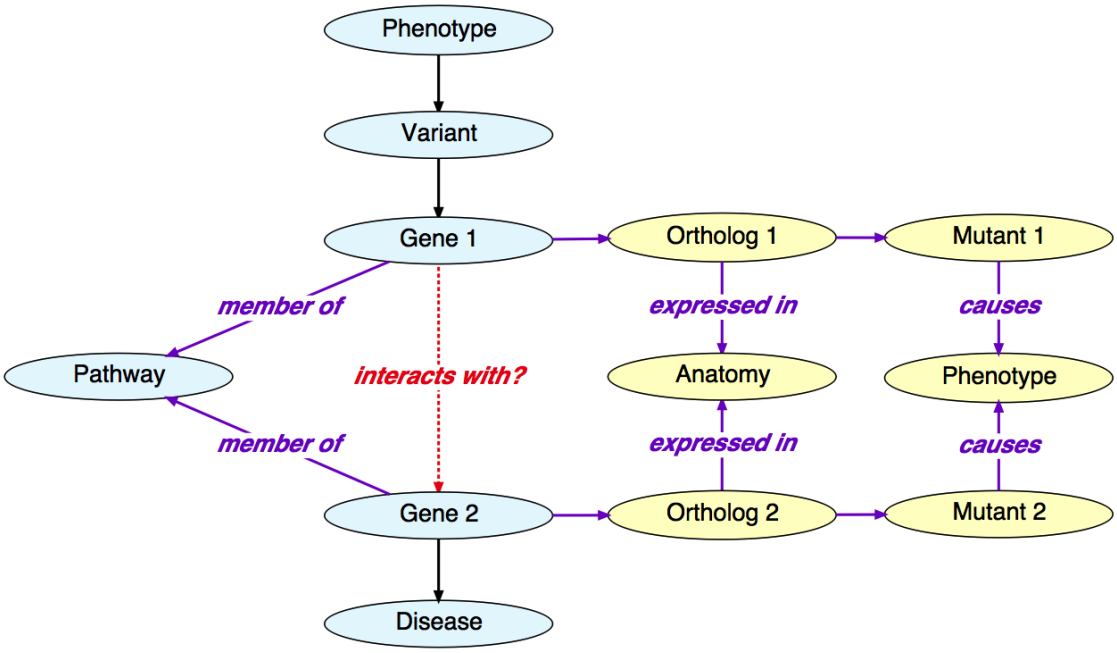

Figure 3: A graph that highlights gaps in knowledge. Given the membership of two human genes (light blue nodes) in the same genetic or protein interaction pathway, as well as mutant phenotype and gene expression data about mouse orthologs (yellow nodes), visualization strategies could guide users to formulate the hypothesis that some critical interaction between the human genes exists.

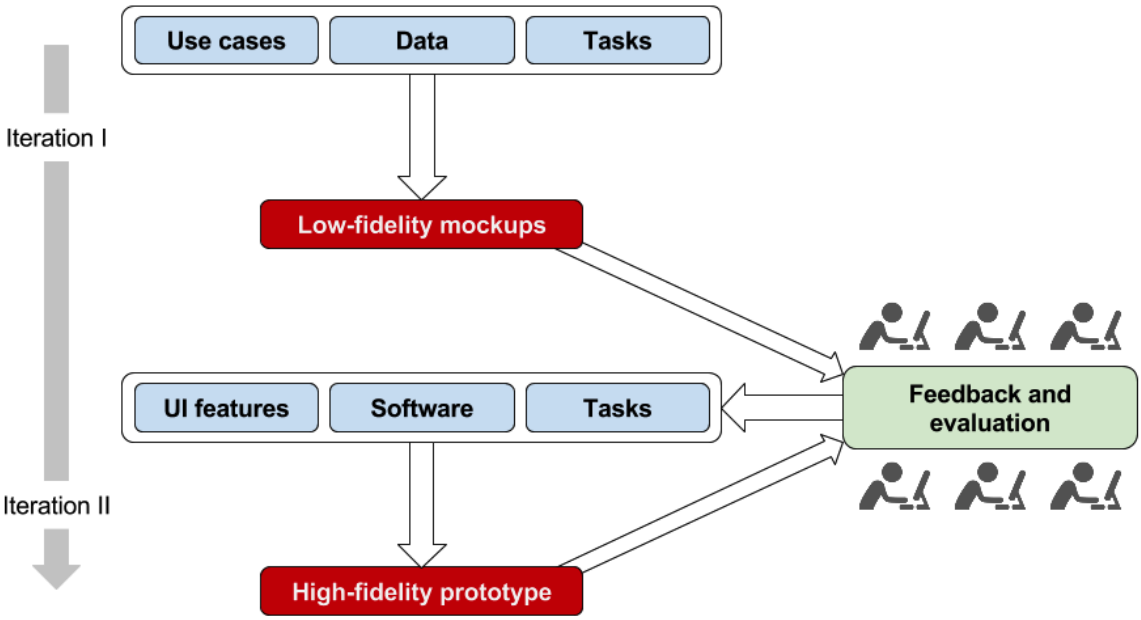

Figure 4: An overview of the study. Use cases and user tasks were developed for first-iteration low-fidelity mockups of a biomedical visual analytics tool. A biological dataset relevant to the use cases was obtained. Six test users were recruited for the study. After obtaining user feedback on the mockups, a second-iteration high-fidelity prototype was developed with a set of interactive features supported by several open-source software libraries. A new set of user tasks was created for a qualitative usability evaluation. The same six users participated in the evaluation.

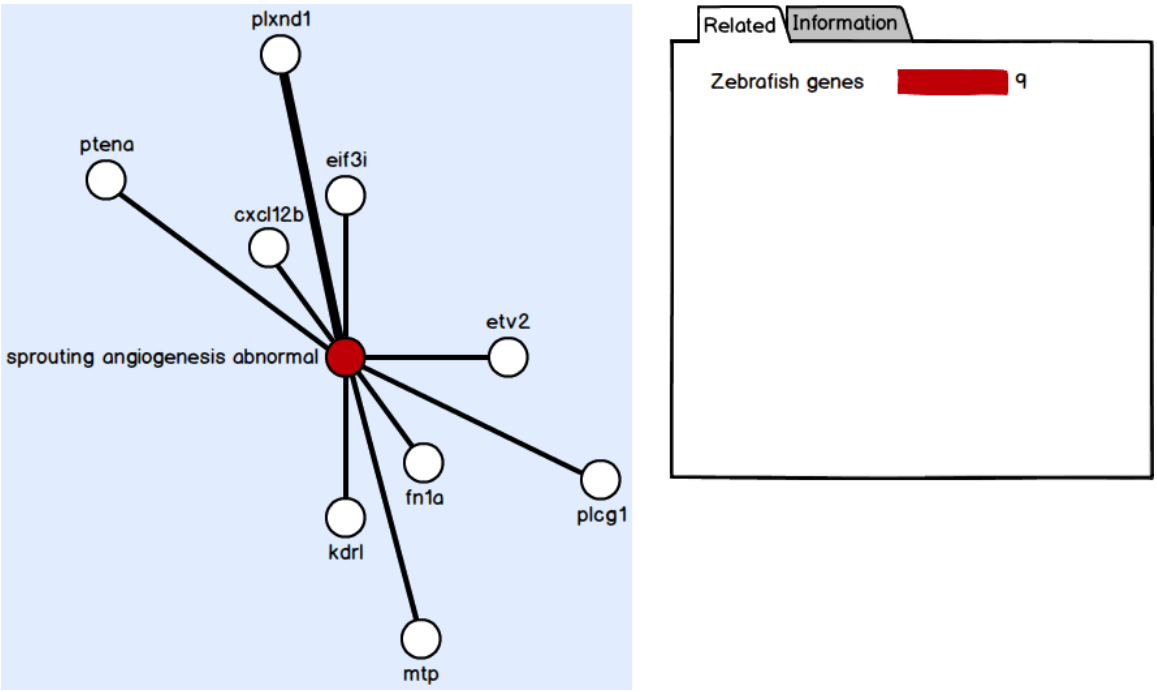

Figure 5: The first in a series of mockups of an interactive genotype-phenotype exploration interface. A graph comprising nodes and edges is displayed at left and a contextual details panel at right. A zebrafish angiogenesis phenotype (red fill) is linked to nine genes.

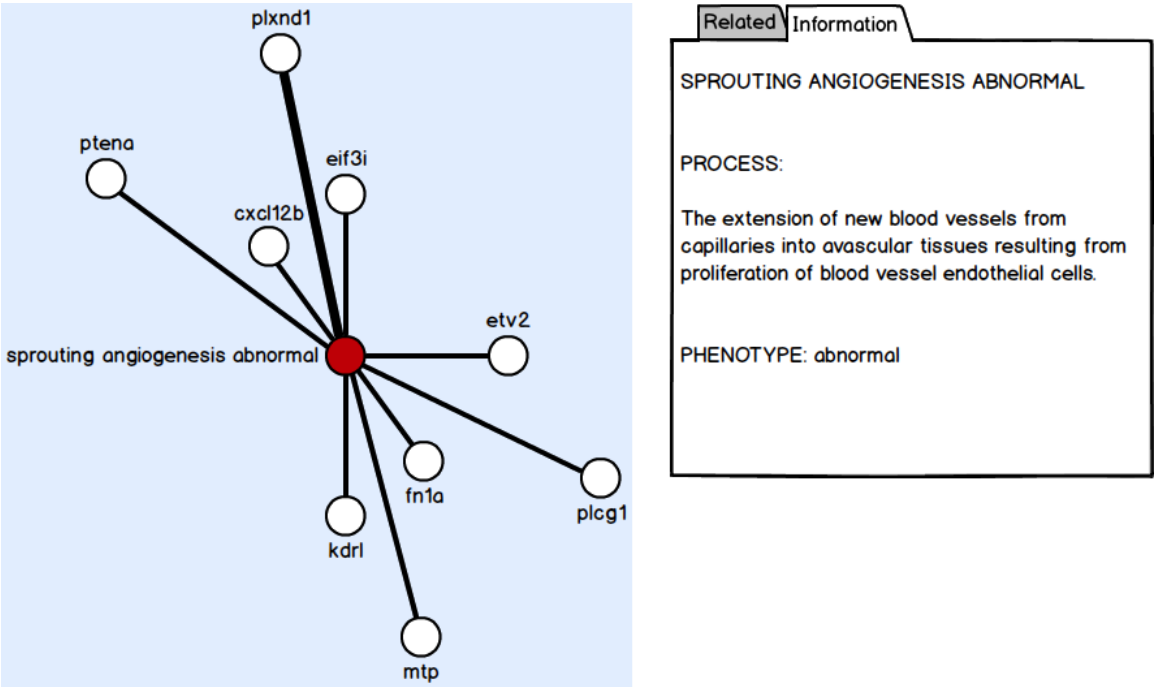

Figure 6: The "Information" tab contains a description of the node selected by the user (red fill).

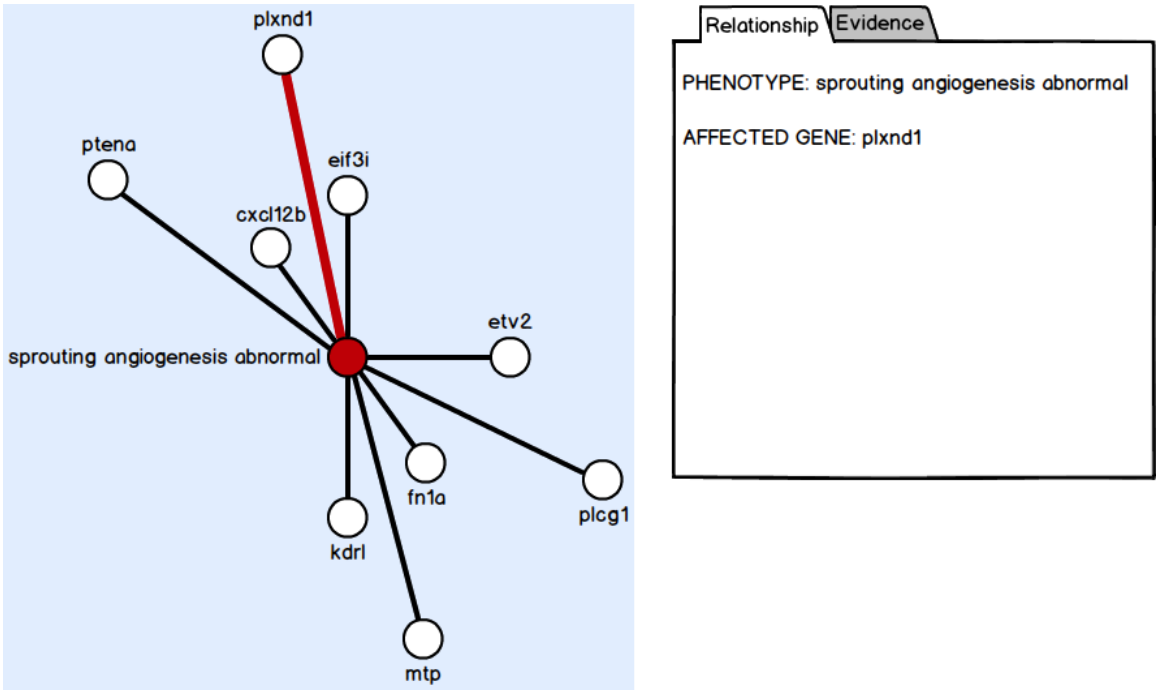

Figure 7: The "Relationship" tab contains a description of the selected edge (red stroke).

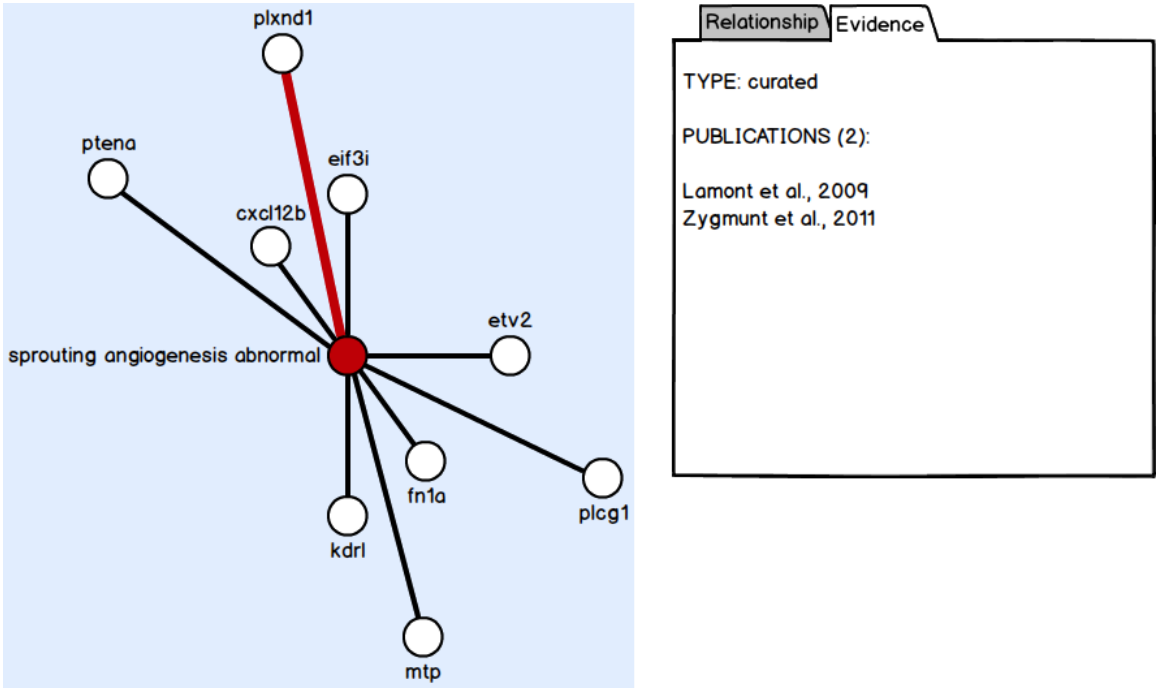

Figure 8: The "Evidence" tab contains a list of sources supporting the relationship between the linked nodes.

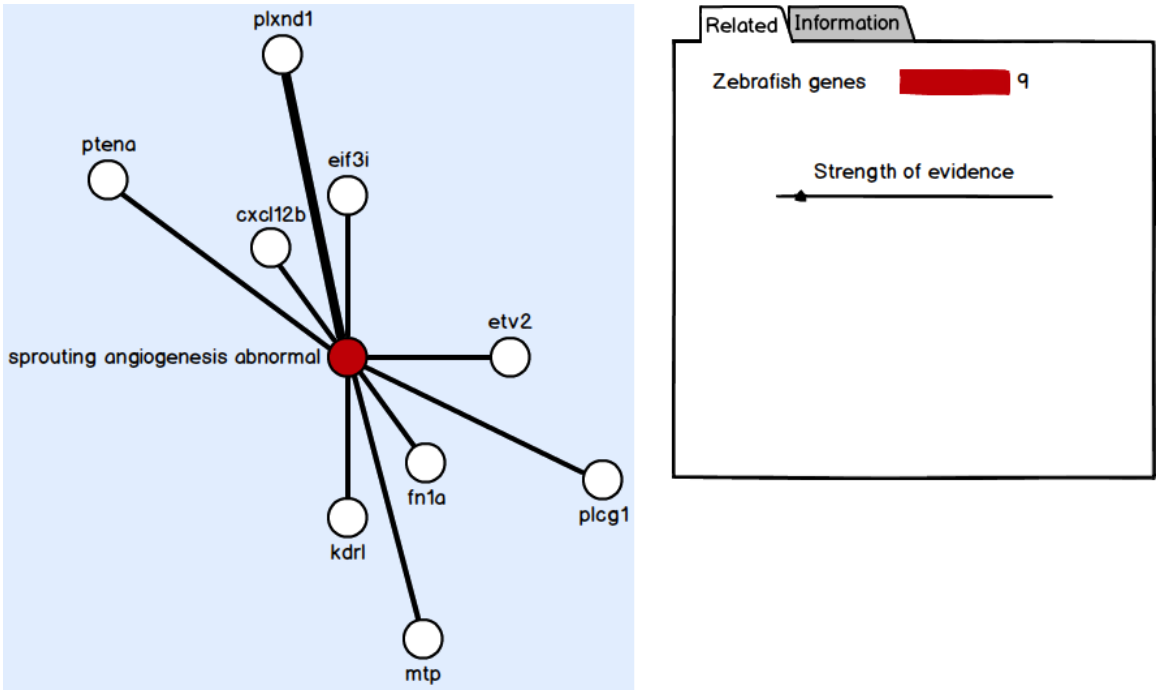

Figure 9: A slider in the panel enables the user to apply a filter on related nodes based on the strength of evidence for their associations, which is indicated by differences in edge width.

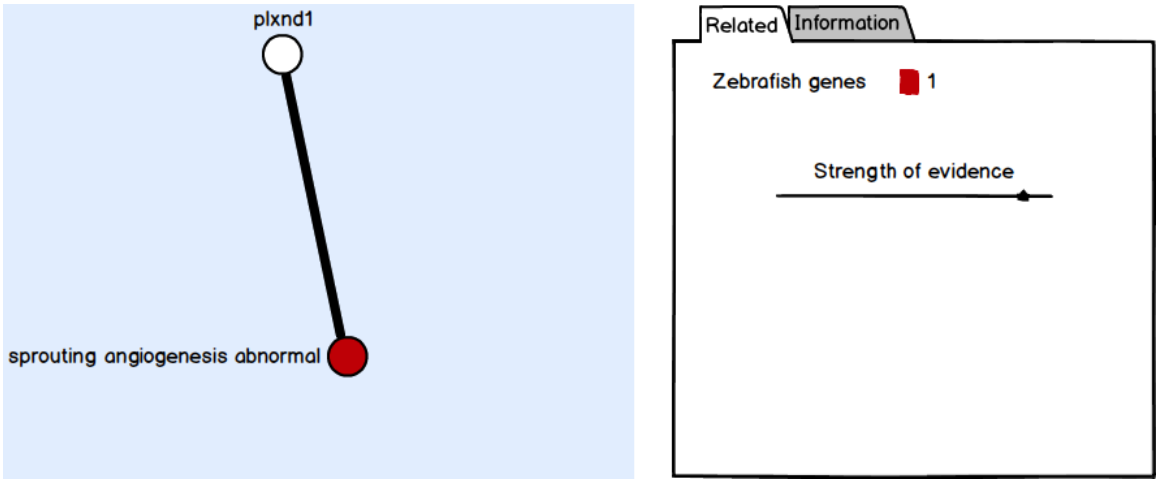

Figure 10: Moving the slider to the right hides all but the one node with the strongest association with the selected node.

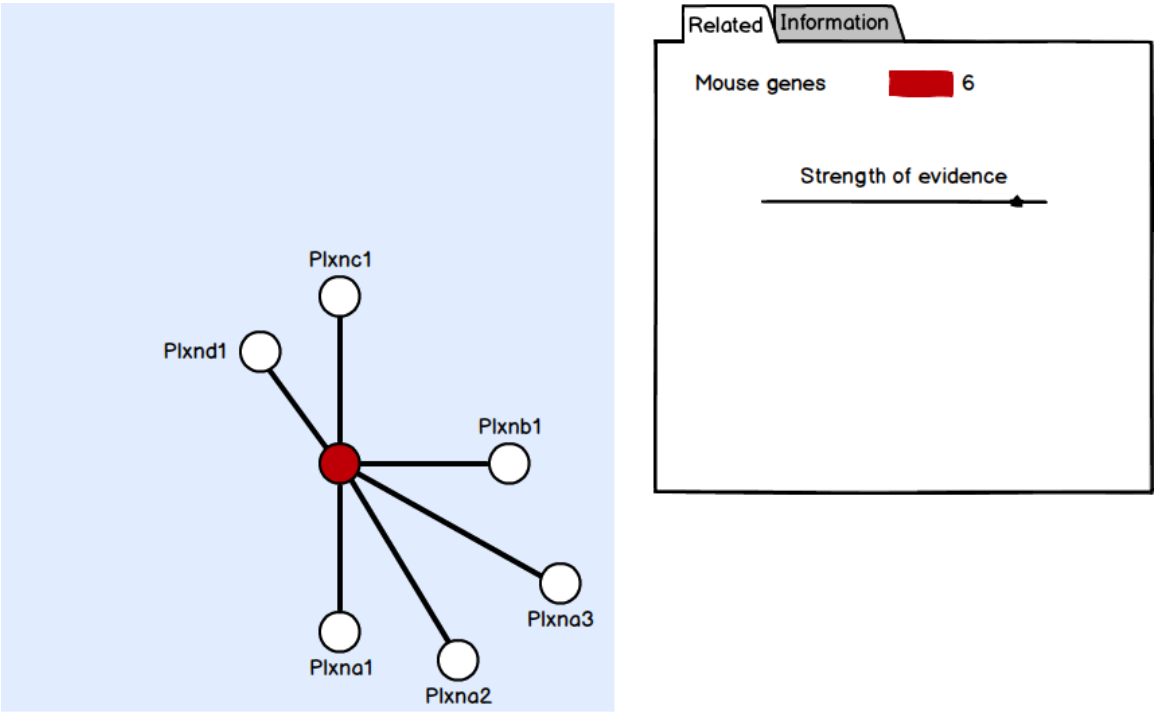

Figure 11: A family of mouse Plexin genes is shown.

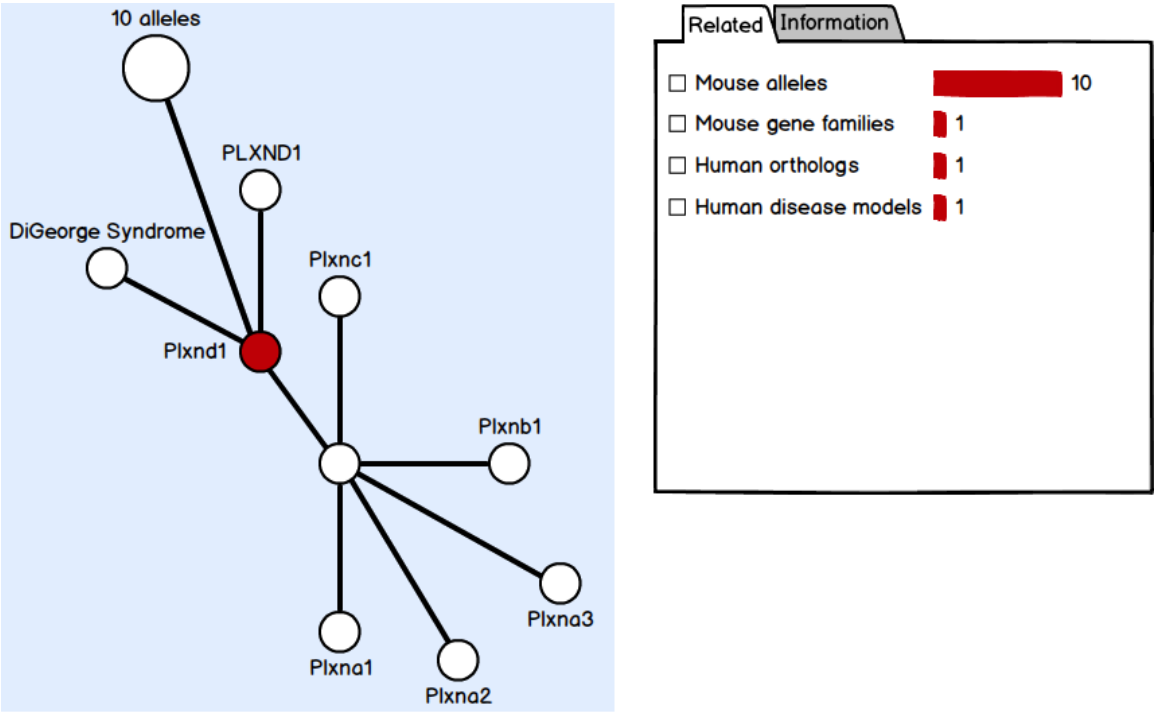

Figure 12: The Plxnd1 gene is selected and related nodes of varying types are added to the display. The ten alleles (variants) associated with the gene are grouped as a metanode, which is larger than the nodes representing single entities.

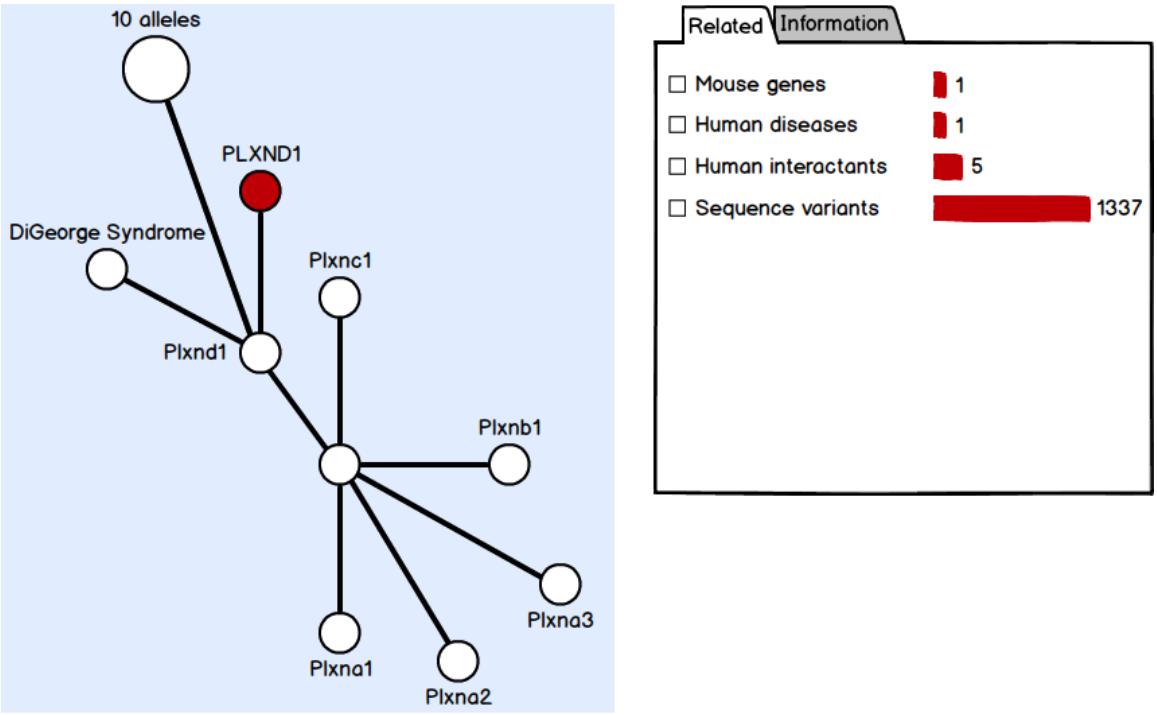

Figure 13: The human ortholog of Plxnd1 is selected.

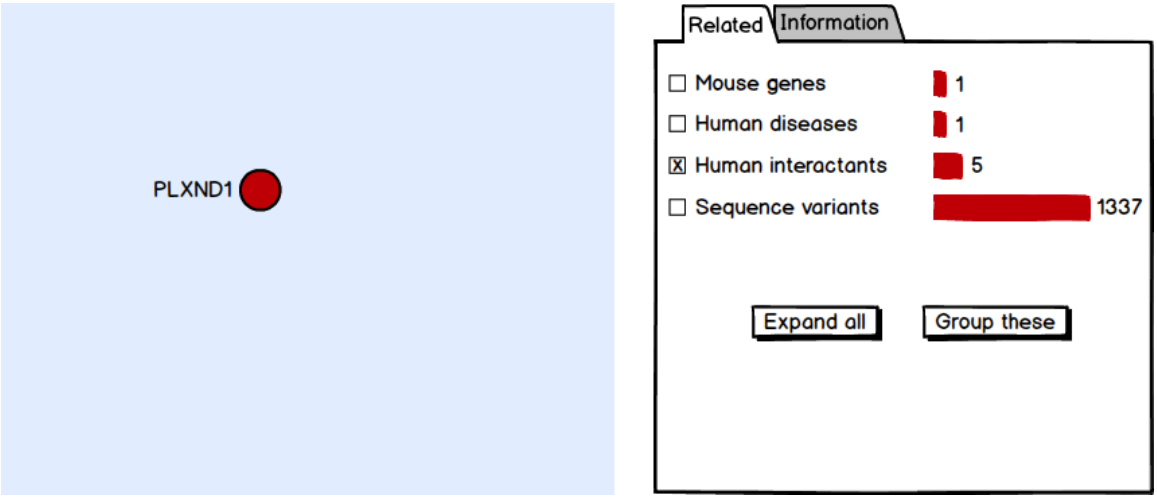

Figure 14: The display is cleared of all but the selected human gene, PLXND1. A box next to "Human interactants" in the "Related" tab is checked, and buttons enable users to add to the display all such items either as single nodes ("Expand all") or as one metanode ("Group these").

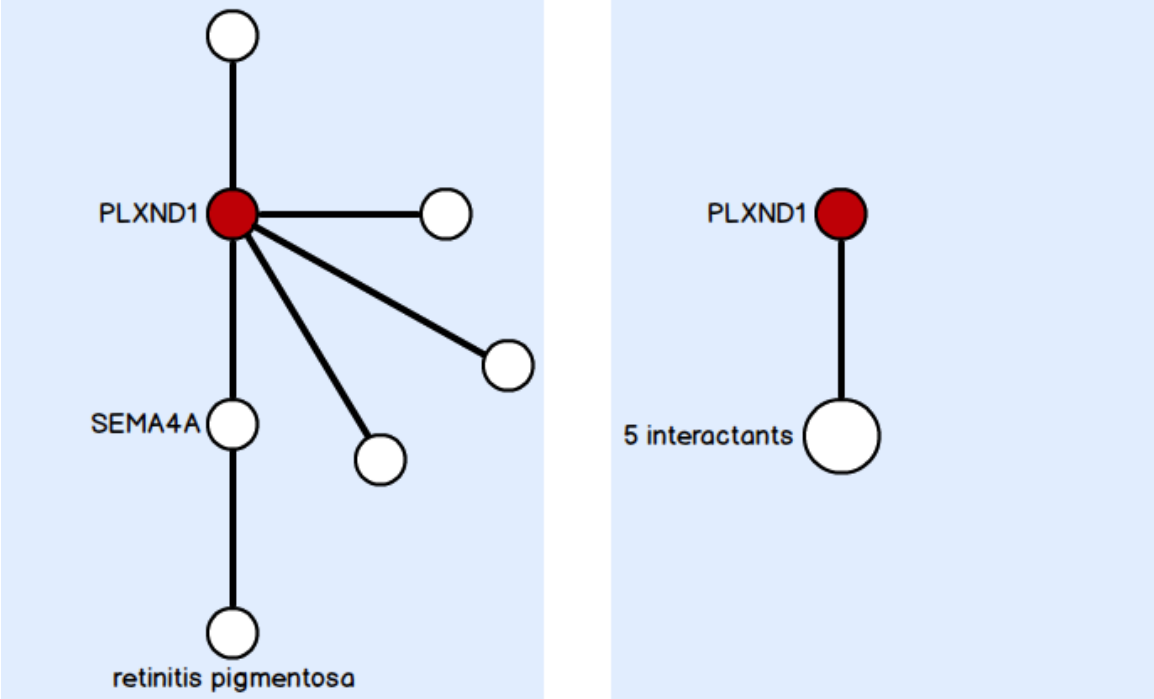

Figure 15: At left, PLXND1's five genetic interactants are displayed individually (only one, SEMA4A, is labeled), and a disease related to SEMA4A is added to the display. At right, the five interactants are displayed alternatively as one metanode.

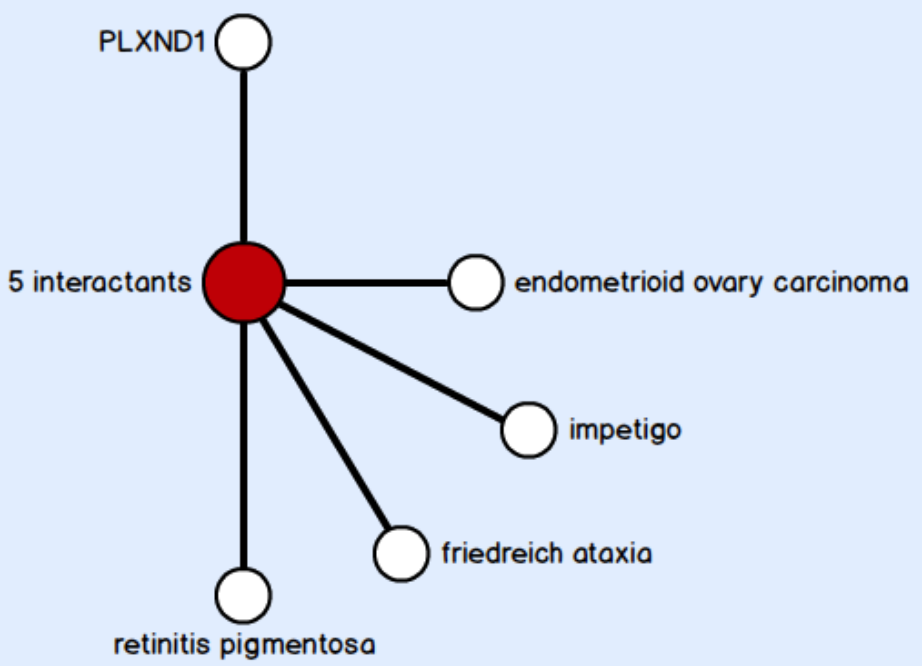

Figure 16: All diseases related to any of PLXND1's genetic interactants are added to the display.

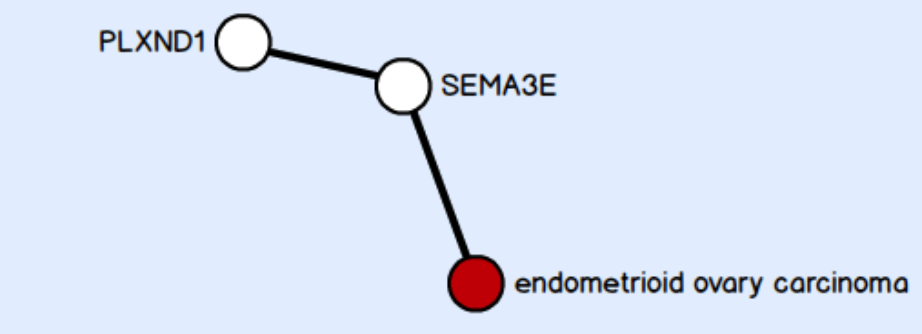

Figure 17: A specific disease is selected and the path linking it back to the original gene is visualized, revealing the specific interactant that is implicated in the disease.

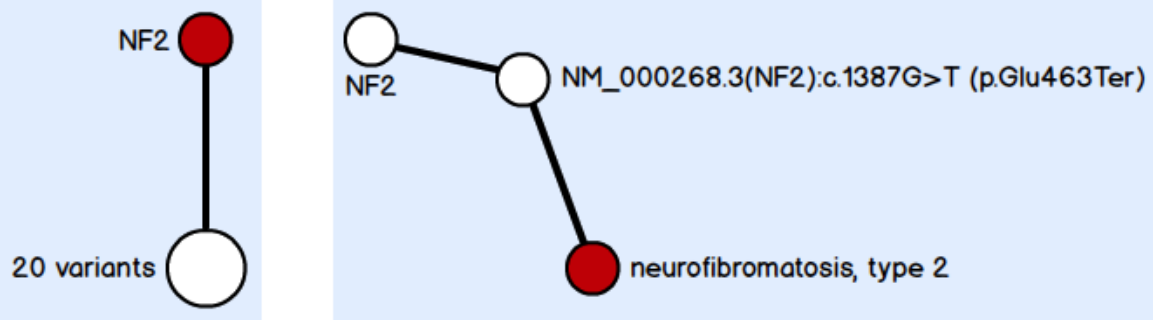

Figure 18: At left, the human NF2 gene is shown with its related variants, visualized as a metanode. At right, a disease and the path linking it back to the original gene are shown, revealing the specific variant that is implicated in the disease.

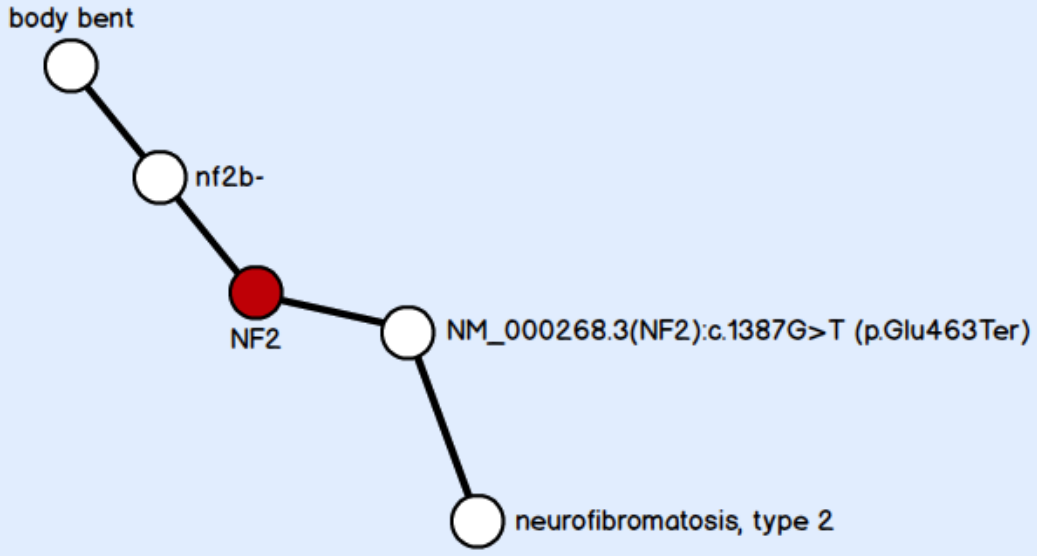

Figure 19: Exploration reveals zebrafish entities related to NF2. A mutation of the *nf2b* ortholog and its related "bent body" phenotype are added to the display.

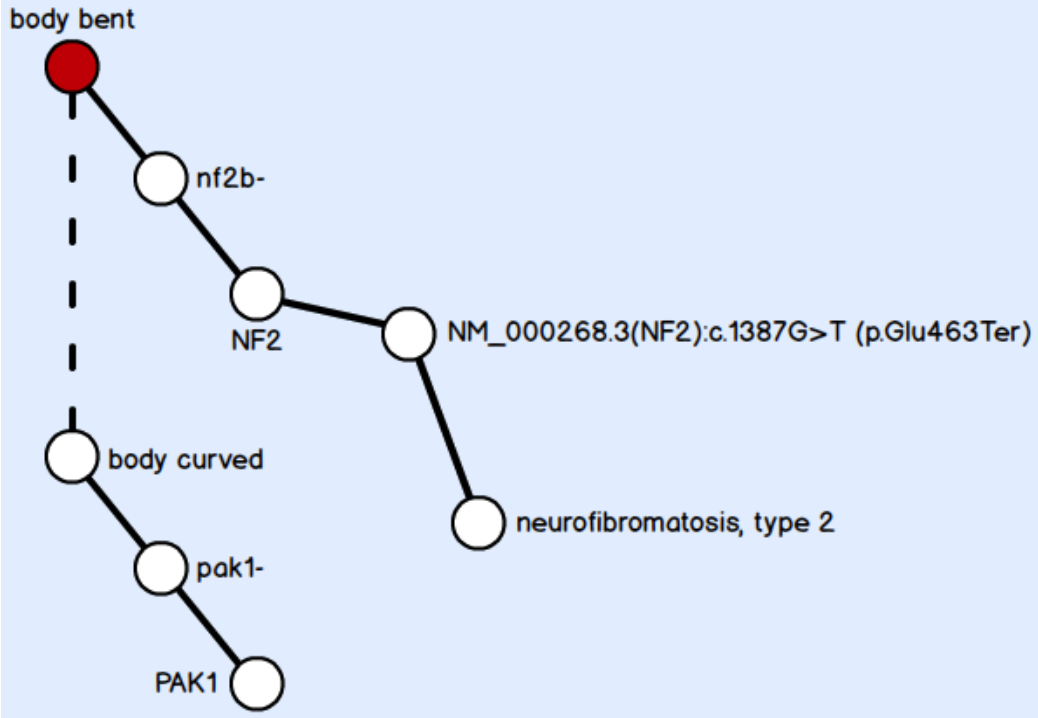

Figure 20: A similar zebrafish phenotype is added to the display (linked by a dashed line) along with a related mutation and human ortholog, PAK1.

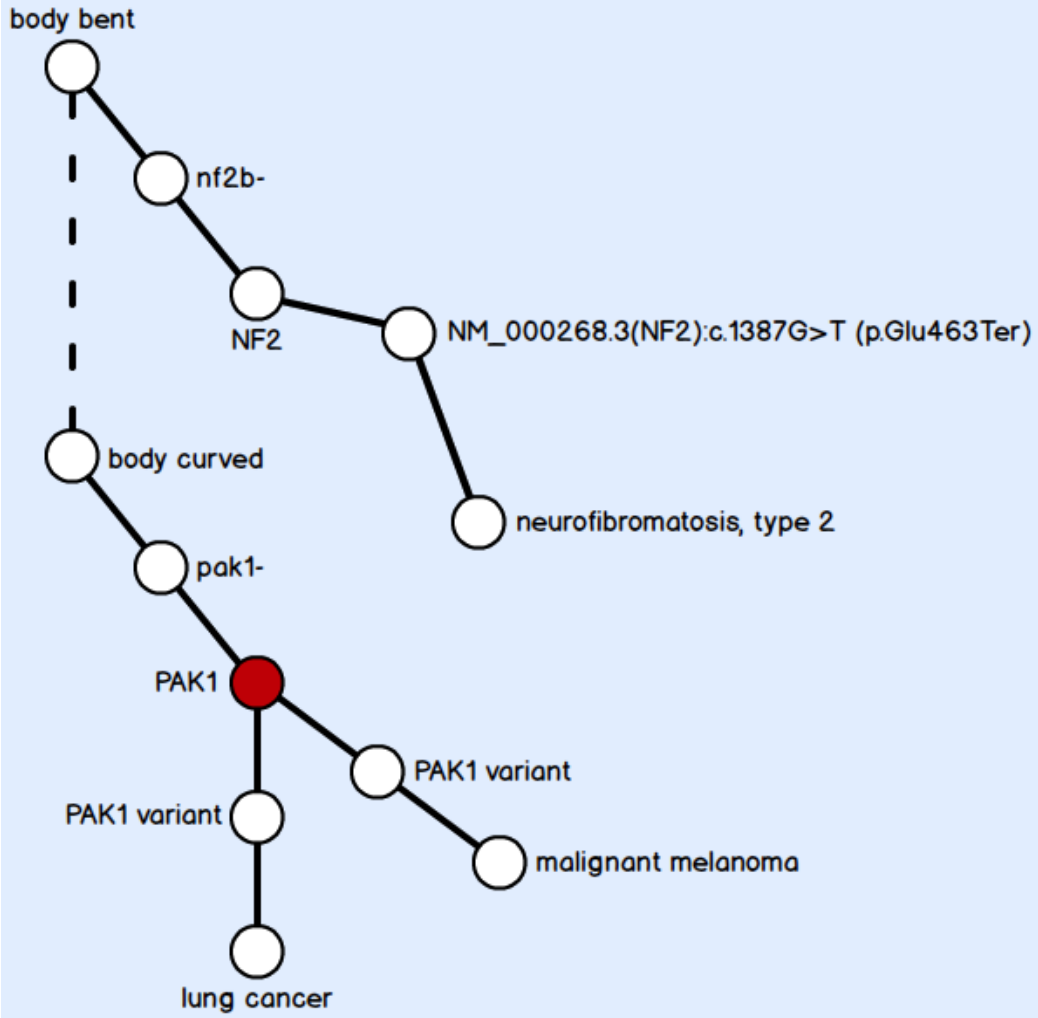

Figure 21: Subsequent expansion of the graph reveals diseases related to PAK1. Descriptions of the various diseases (not shown) and the nature of the zebrafish phenotypes may suggest to the user that some genetic alteration affecting an interaction between NF2 and PAK1 is involved.

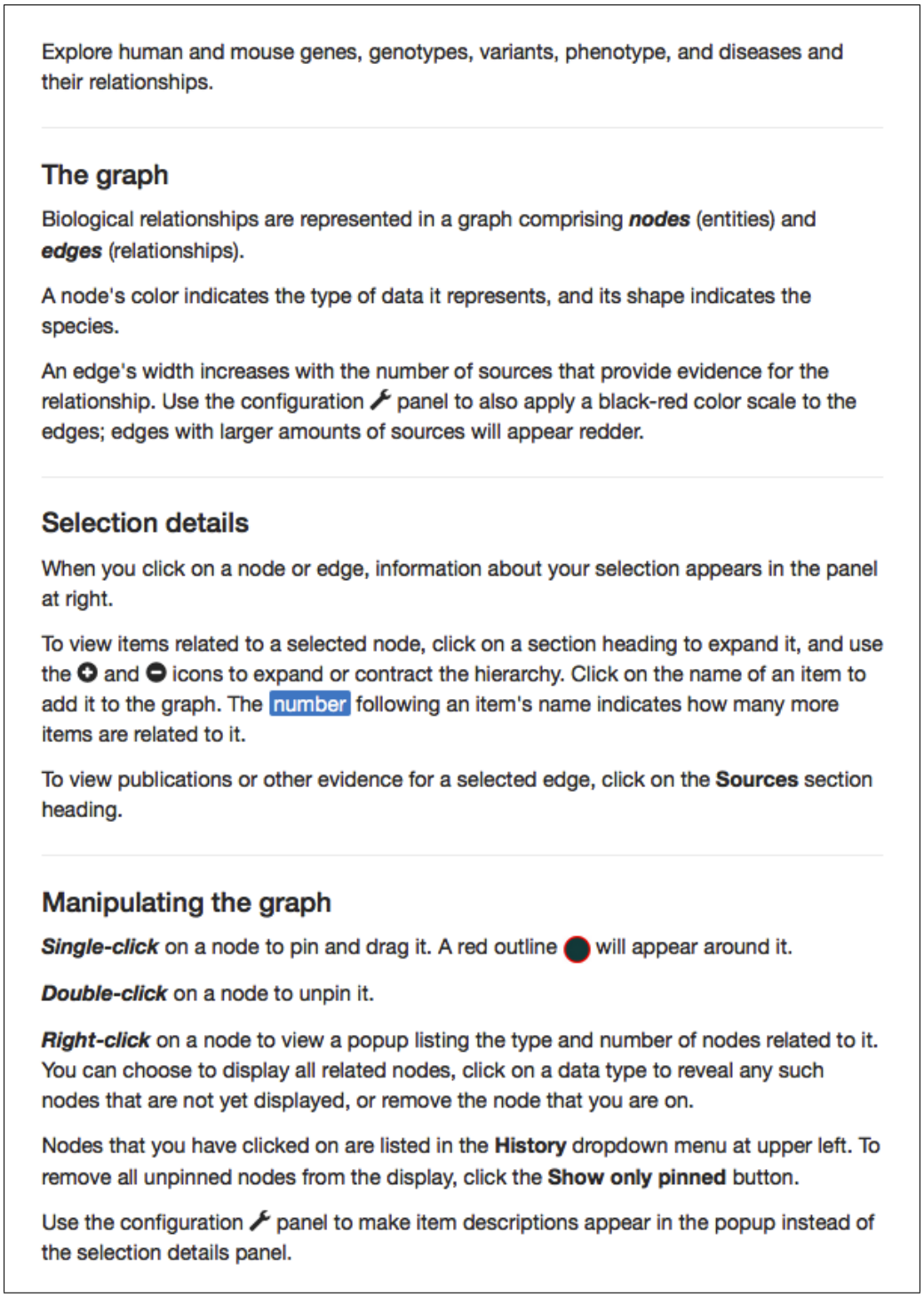

Figure 22: The functional prototype's help dialog window.

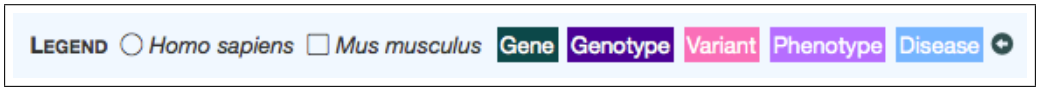

Figure 23: An expandable legend defining the node shapes and colors in the graph.

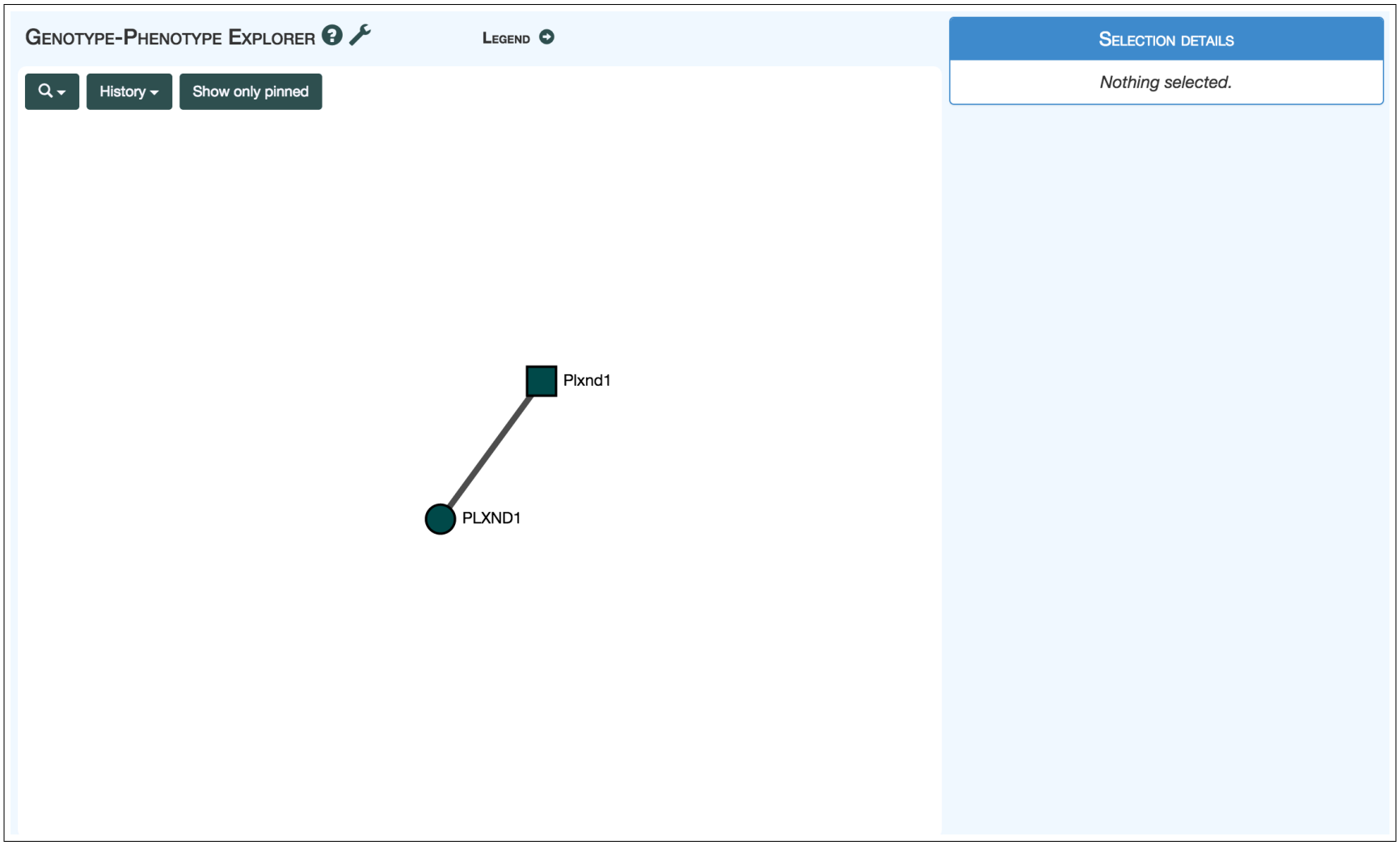

Figure 24: The first in a series of screenshots of the functional prototype. A force-directed graph, initially showing the human and mouse orthologs of a plexin gene, is displayed at left and a contextual detail panel at right.

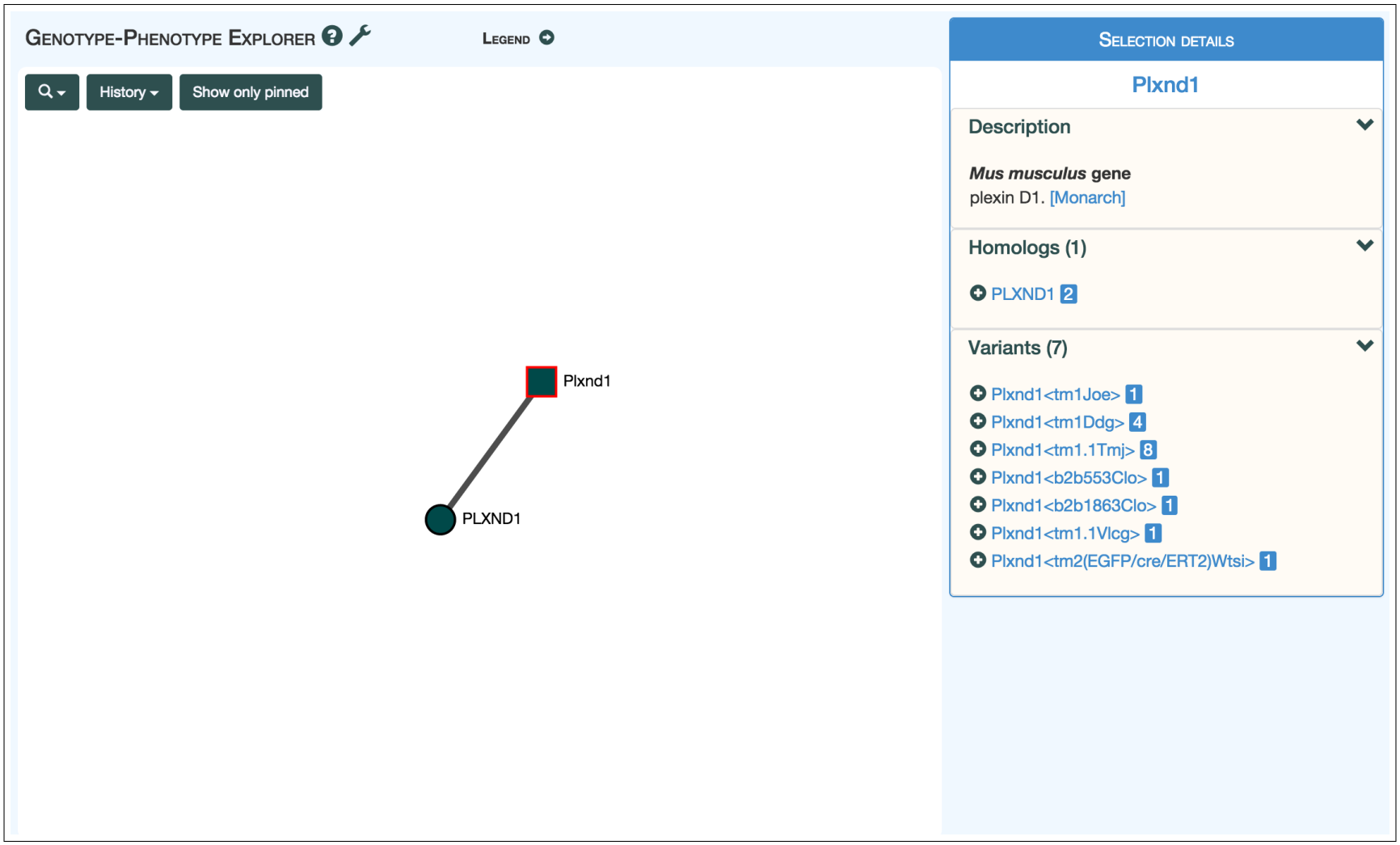

Figure 25: The node representing the mouse Plxnd1 gene (red stroke) is pinned to the display. The "Selection Details" panel reveals a description of the gene and expandable lists of related items.

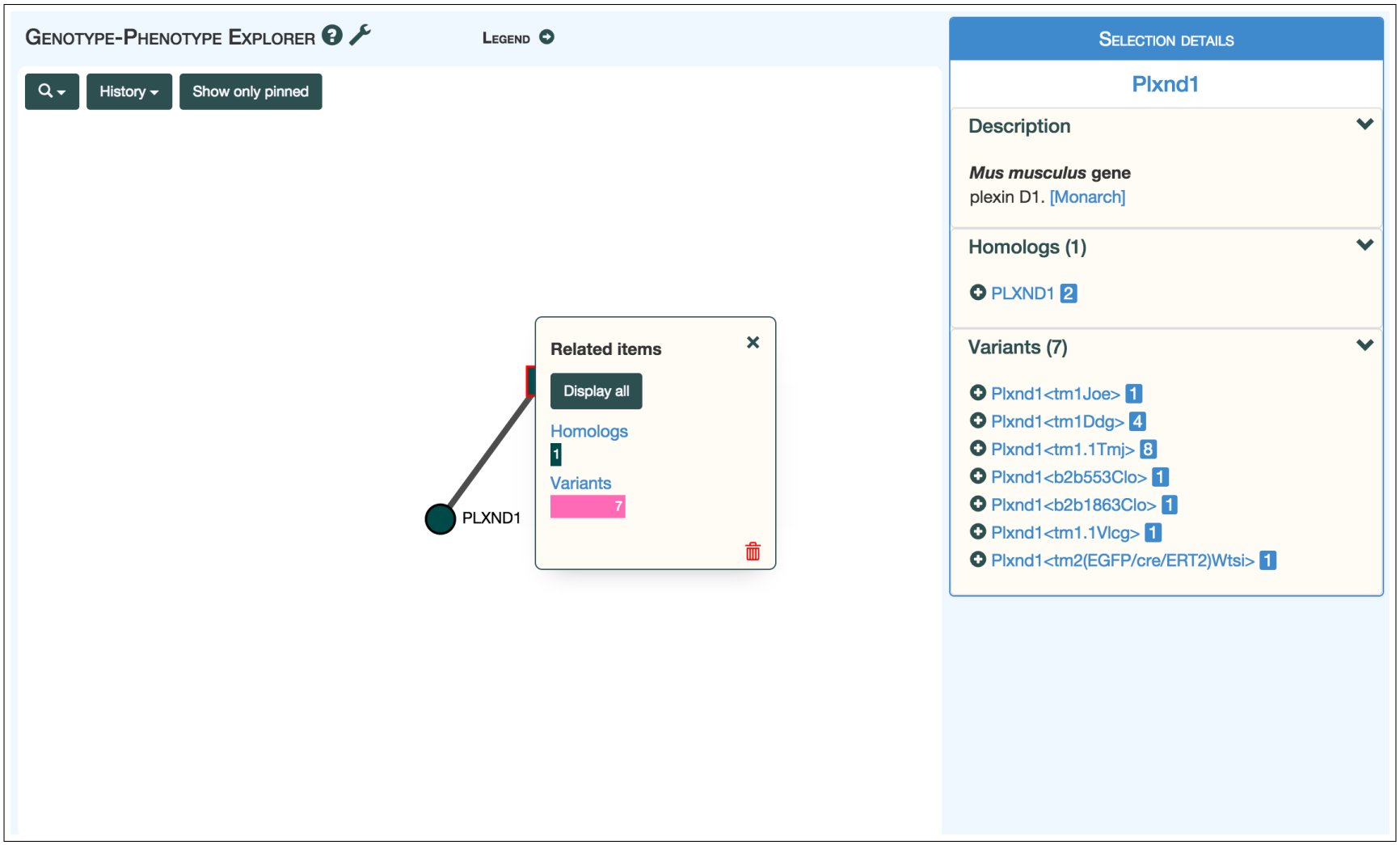

Figure 26: A right-click popup on the Plxnd1 node reveals options for adding related items to the graph, either all at once or by a specific data type. The node can also be removed from the graph by clicking the trash can icon.

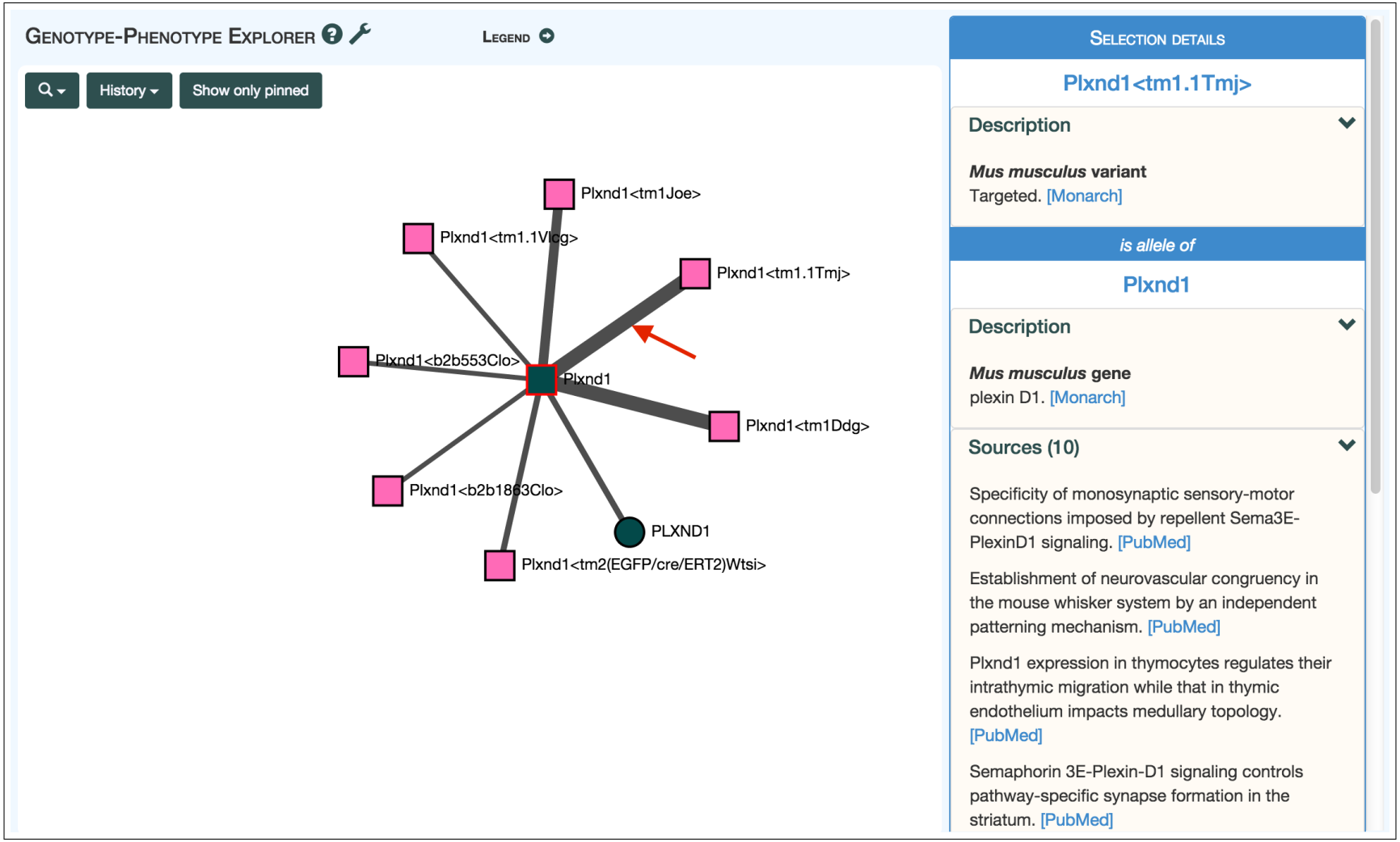

Figure 27: All Plxnd1 variants (pink fill) are added to the graph. Clicking on the edge (arrow) linking Plxnd1 and the tm1.1Tmj variant fills the "Details" panel with information about the relationship and a list of sources that support it.

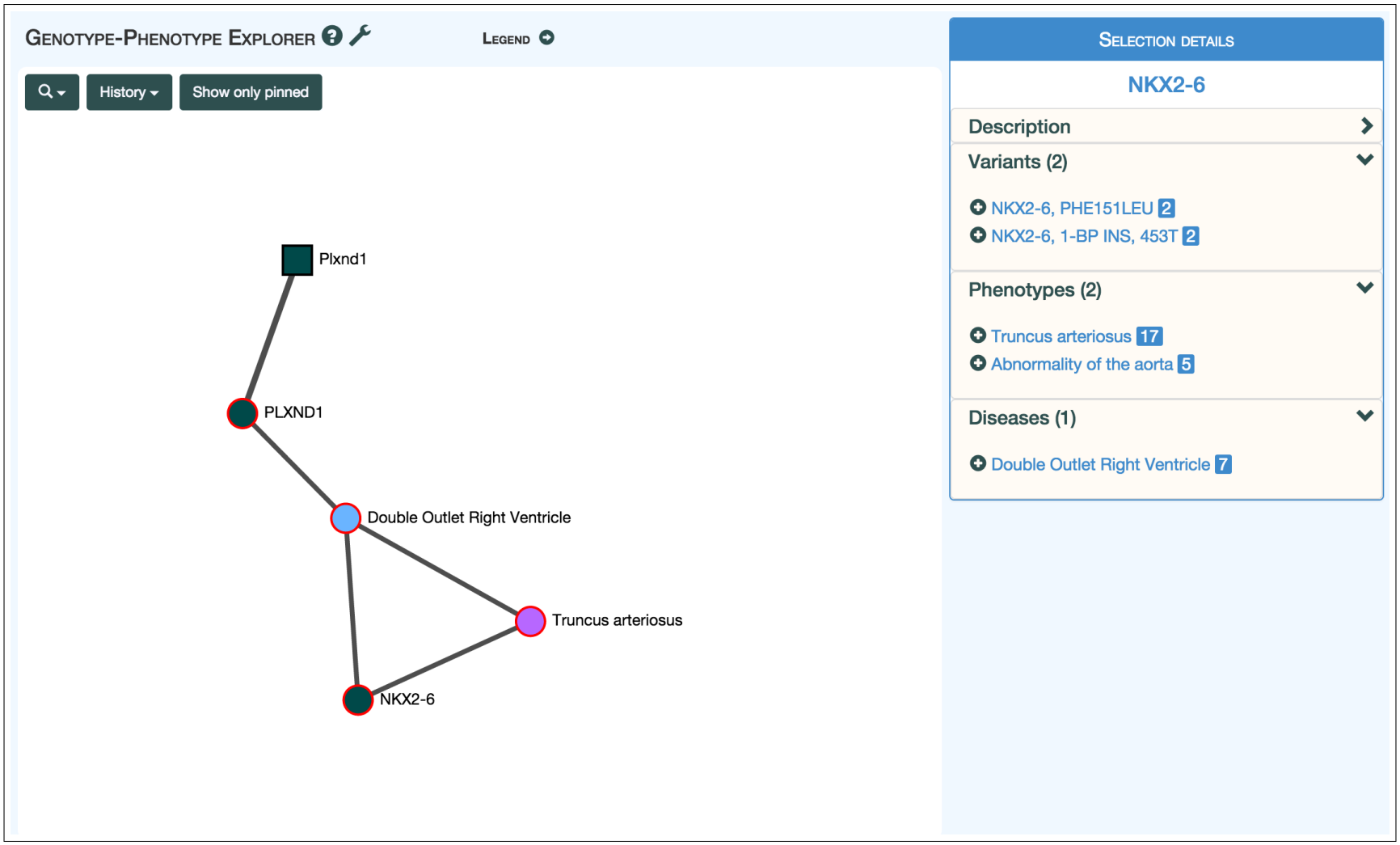

Figure 28: Individual items have been added to the graph by clicking on their labels in the "Details" panel. Adding the NKX2-6 gene reveals its relationship to a human phenotype and disease.

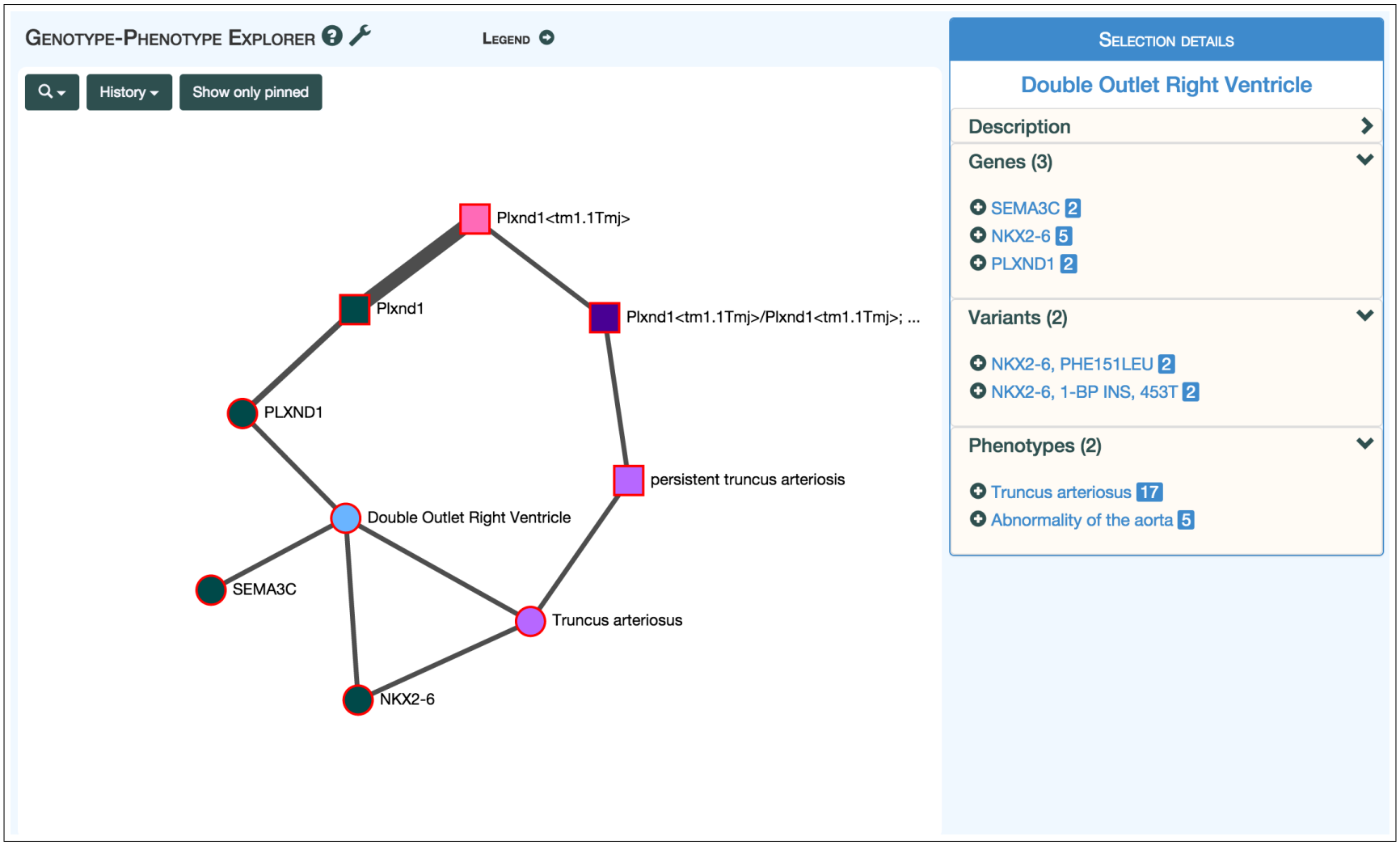

Figure 29: Subsequent exploration reveals examples of every data type that the prototype supports, showing multiple paths between genes (dark green fill) and a disease (light blue fill).

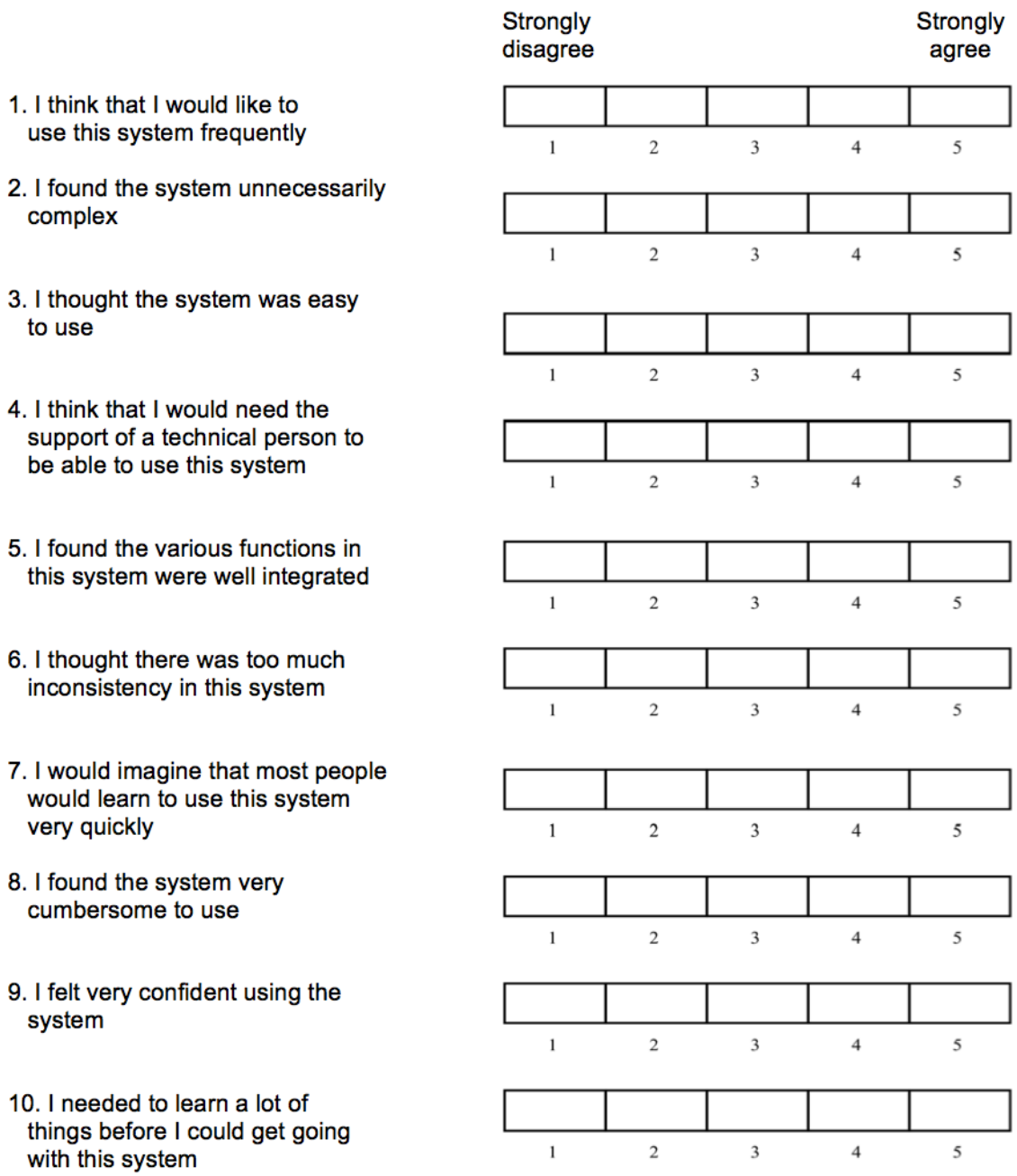

Figure 30: The System Usability Scale, developed by the Digital Equipment Corporation [21].

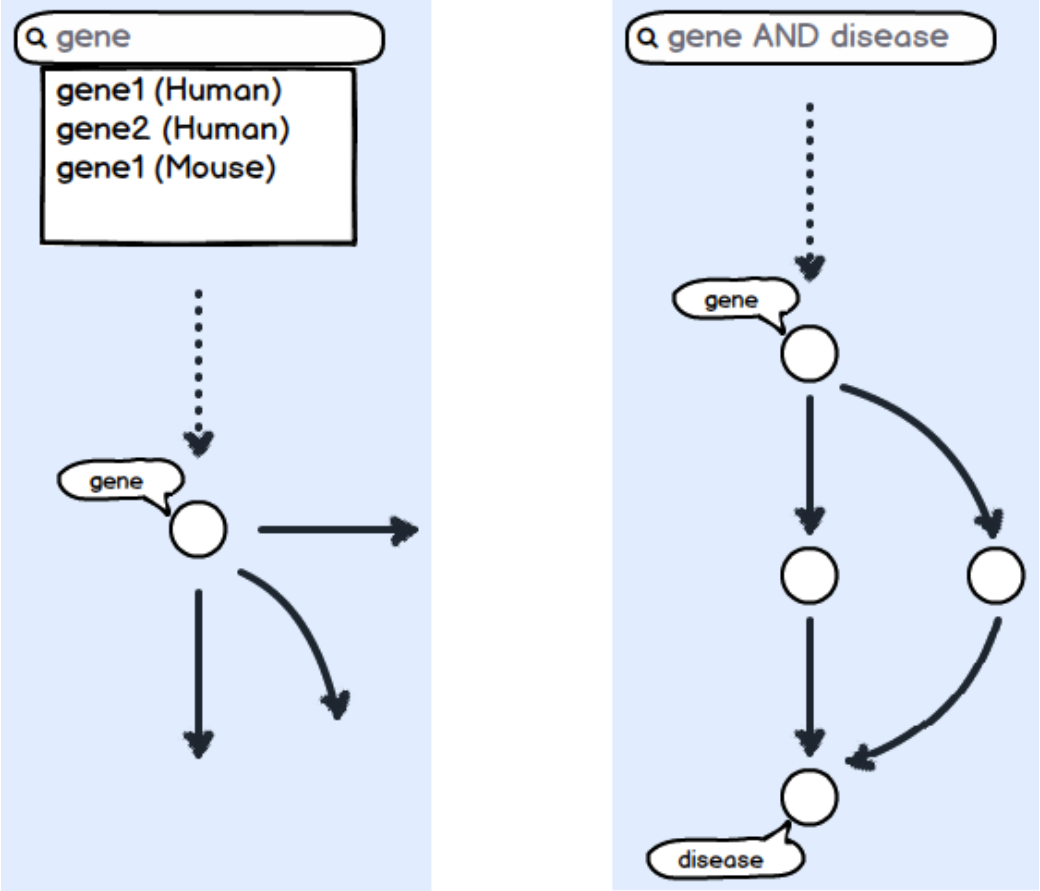

Figure 31: Sketches of two simple user search scenarios. At left, the user searches for a gene, aided by an auto-complete menu, and places a node on the display that provides a starting point for exploring related items. At right, she alternatively queries genes and diseases, placing two nodes on the display, with all possible paths between the two being revealed.

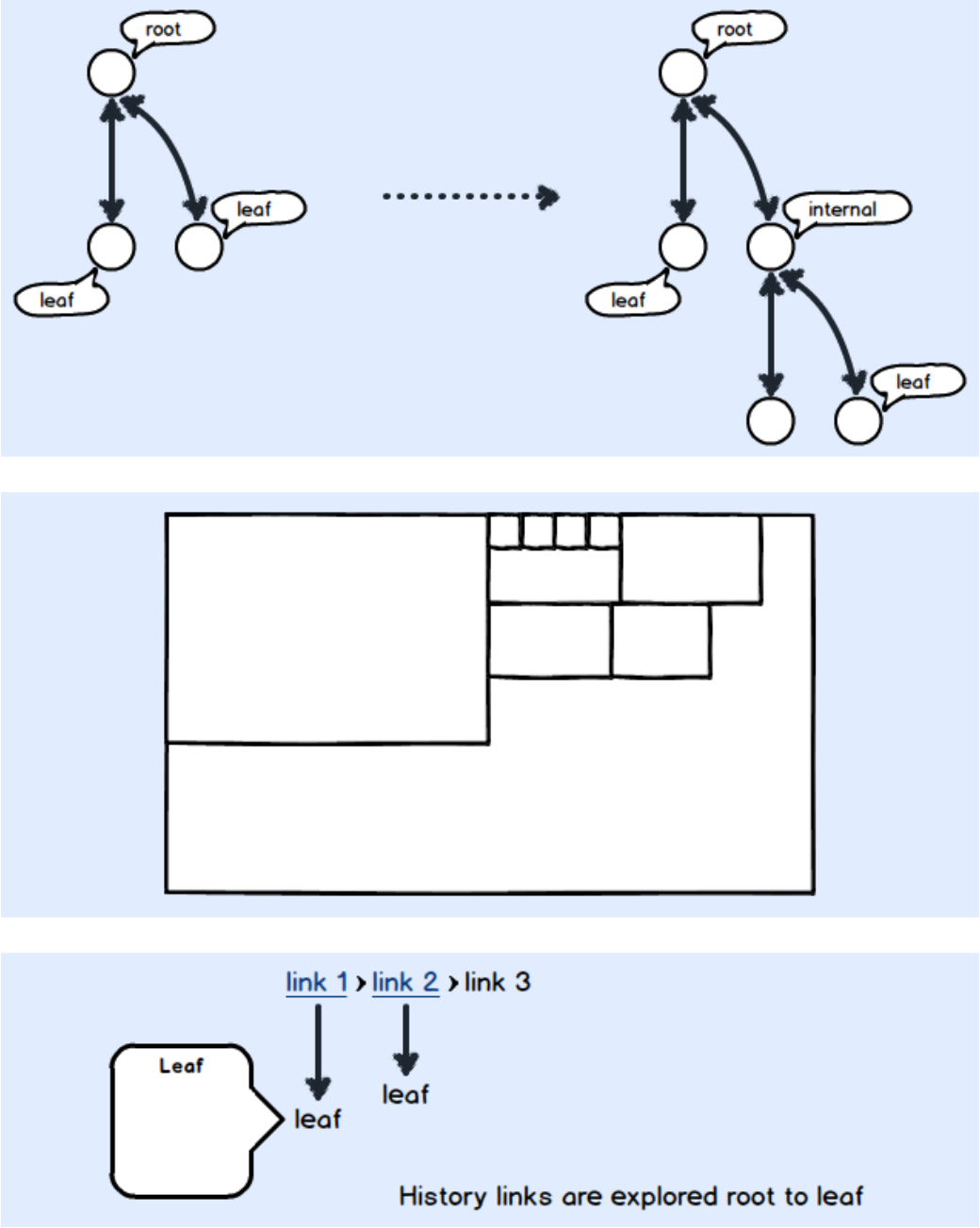

Figure 32: Sketches of alternative graph-based views: a tree view of entities and relationships (top), a treemap consisting of nested rectangles the relative areas of which let users compare values (middle), and a breadcrumb trail allowing the user to inspect each visited link from root to leaf.

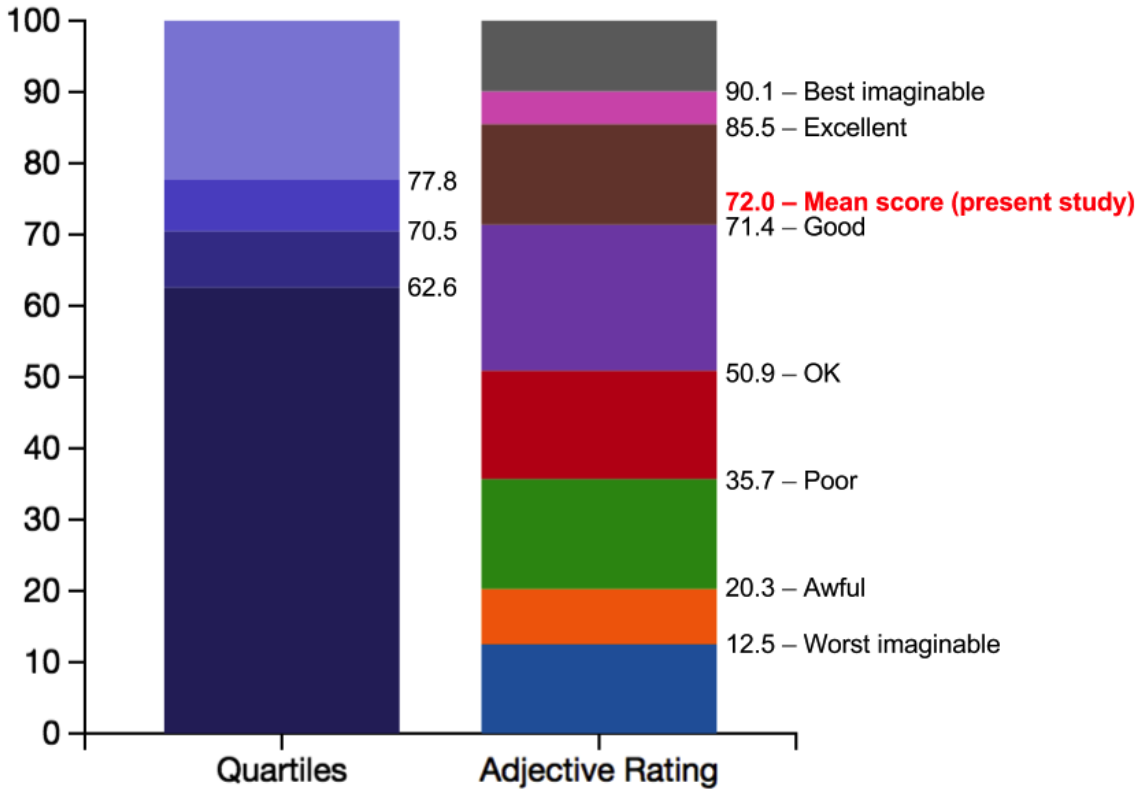

Figure 33: Quartiles for System Usability Scale study mean scores (≈3500 surveys in 273 studies) and a corresponding adjective scale developed by Bangor, et al. (2009) [48]. The mean score (6 surveys) obtained from the present study is shown in red.

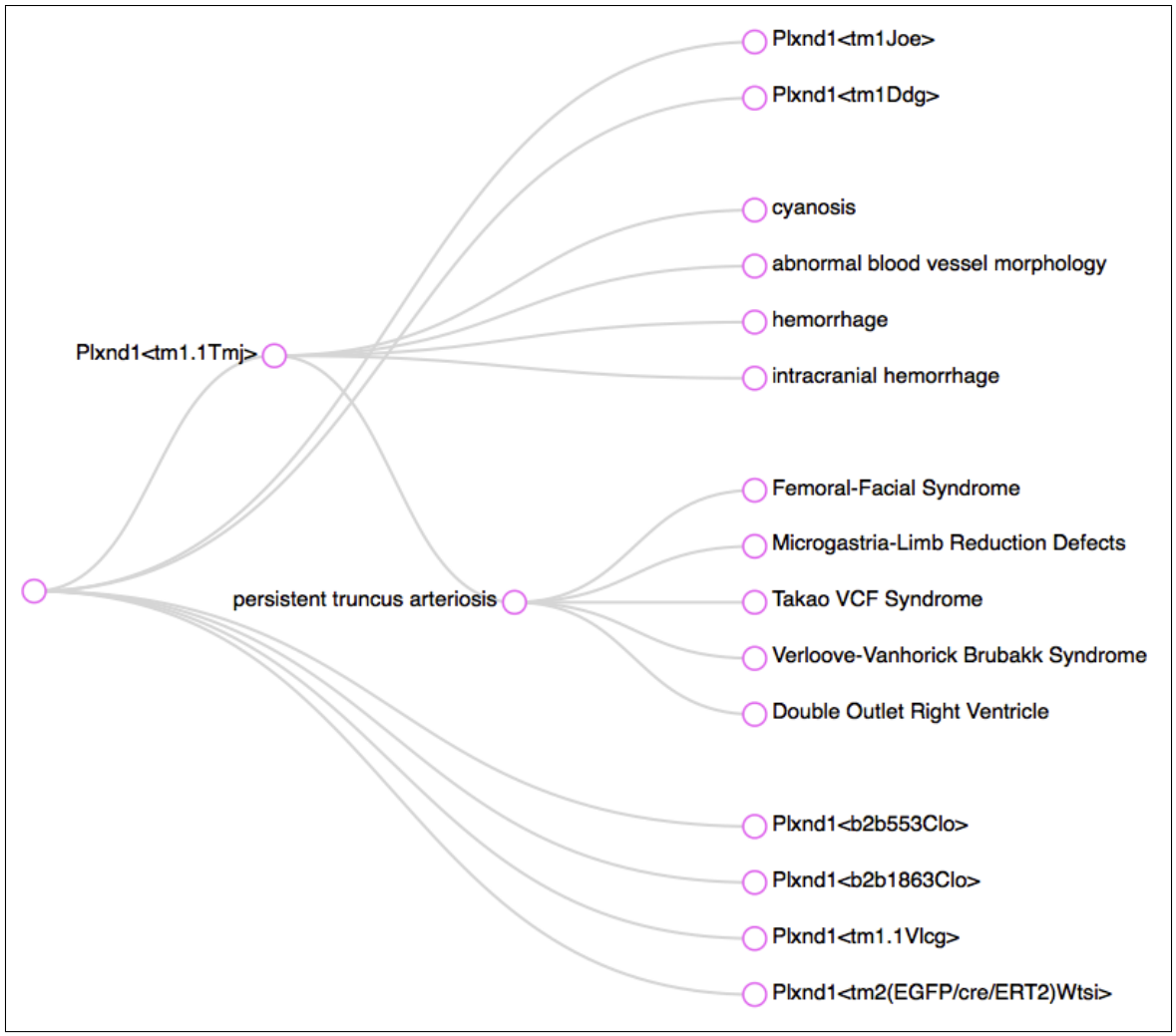

Figure 34: A mocked-up D3 cluster layout showing links between a mouse variant (Plxnd1$^{\mathrm{tm1.1Tmj}}$) and related phenotypes and diseases.

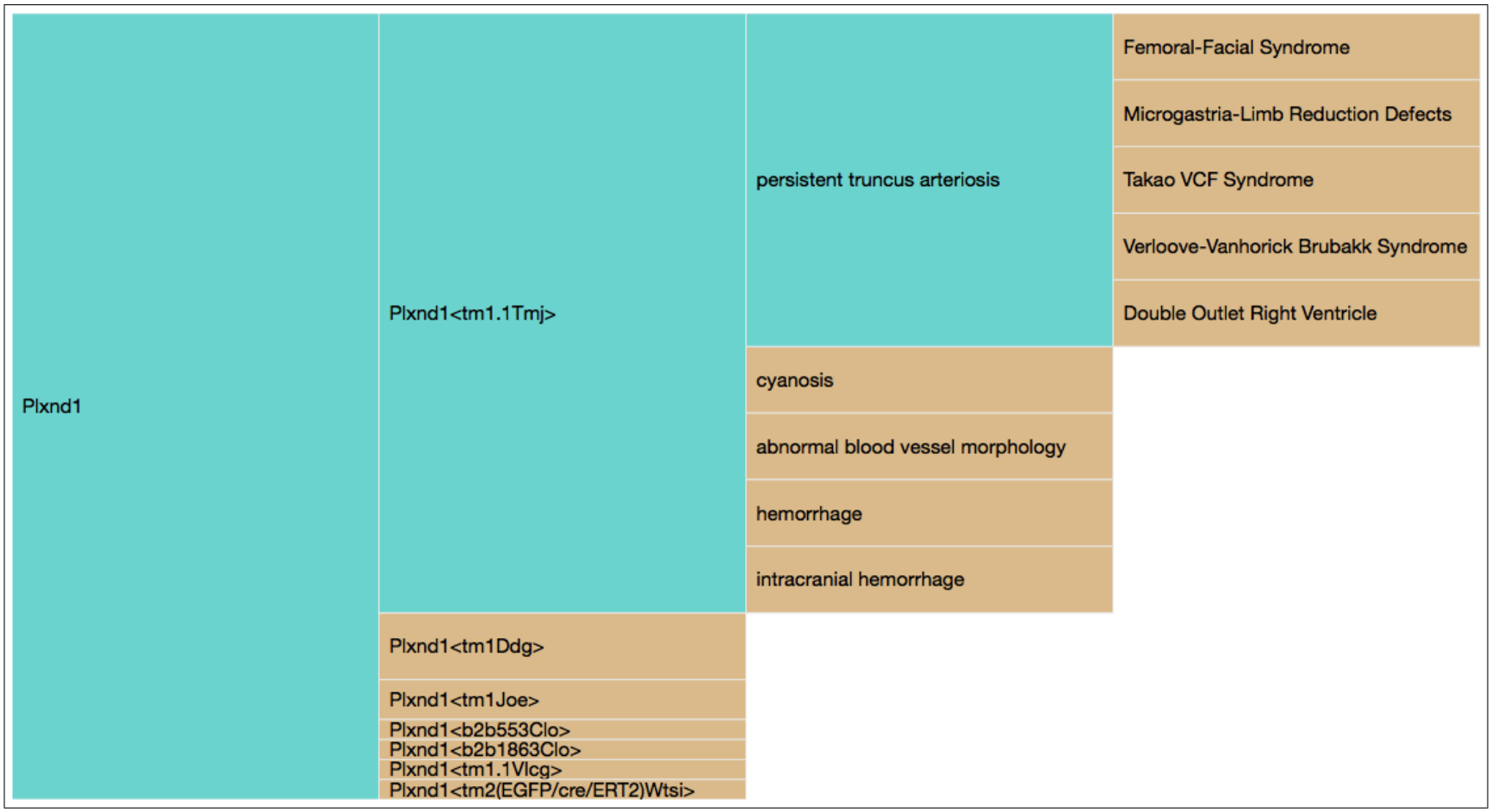

Figure 35: A mocked-up D3 zoomable partition layout showing associations between a mouse gene (Plxnd1) and variants and related phenotypes and diseases. Users can click on rectangles to zoom in and out of the display. Rectangle heights can be made proportional to some dimension of the underlying data.

Tables

Table 1: Iterative interface design and usability evaluation plan.

Iteration	Activities
I. Mockups. Scenario building, brainstorming, requirements gathering, and informal evaluation of low-fidelity prototypes. One hour session with each of six users, working individually.	Introduction – 10 min. Brief overview of Monarch Initiative. Simple storyboarding.
	Task A – 15 min. Angiogenesis cancer phenotype general use case. Several nodes of the same data type are linked to a node of interest (e.g., zebrafish genes related to a zebrafish phenotype) and the user wishes to filter them. She sees the type and strength of evidence for these associations and applies a filter based on this information.
	Task B – 15 min. Angiogenesis cancer phenotype general use case. The user has found a few zebrafish or laboratory mouse genes of interest. Many different types of nodes are linked to them (i.e., there are several different relationship types) and she needs to decide which path through the knowledge graph to explore next. She selects a node and sees the number of linked nodes for each relationship type. She applies a filter based on relationship type, and opts to show nodes of a particular related type as a metanode.
	Task C – 15 min. Cell proliferation phenotype general use case. The user sees that a single gene of interest is linked to many diseases. She wishes to gain insight into how the gene is associated with one of those diseases. For example, the gene may have many interactants, and mutations in these interactants may be pathological. The user explores the various links between genes, interactants, genetic variants, and diseases.
	Debriefing – 5 min. Follow-up remarks.

Continued on next page

Iteration	Activities
II. Functional prototype. Formal evaluation of interactive web-based high-fidelity prototype; requirements gathering for future iterations of a Monarch application. 50 min. sessions with each of the same six users, working individually.	Think-aloud evaluation – 30 min. Concrete use case involving angiogenesis genes, genotypes, phenotypes, and diseases. Task A – The user starts with a small set of human and mouse genes and displays genetic mutations linked to one of them. She finds out the extent to which each variant has been studied and reported in the scientific literature. She follows a reference to one of the publications and then returns to the knowledge graph. She finds out how many nodes (of any type) are linked to each of the variant nodes. She wishes to backtrack and remove the variants from view. She then views only the variant which has the most additional nodes linked to it in the knowledge graph. Task B – Starting with a human gene, the user explores the knowledge graph, selectively visualizing related diseases, phenotypes, genes, and genotypes in humans and mice. Task C – The user configures how the interface displays descriptions and indicates strength of association between nodes.
	Written evaluation – 10 min. Survey the user's experience with the System Usability Scale (SUS) and collect additional remarks that he or she wishes to add.
	Debriefing – 10 min. Oral interview. Follow up on SUS questions. Discuss needs for future iterations of the tool (beyond the present study).

Table 2: Questions for the users.

Requirements
• How do you want to use the system?
• What kinds of queries do you want to perform?
Tasks and work
• What are your thought processes as you navigate the interface?
• Do you understand how various sections of the interface are related?
• Are the interface's visualization features intuitive?
• Can you do the specific tasks that you are prompted to do?
• Does the system effectively present and communicate biological information?
• Are the relationships between different kinds of data meaningful?
• Is it easy to access related biomedical resources that are outside the system?
• Does the system enhance your ability to identify and solve biological problems?
Heuristics
• Is the system engaging, enjoyable, and easy to learn?
• Does the system follow real-world conventions?
• Does the system tolerate errors and allow backtracking?
• Is the system consistent and follow standards?
• Can you recognize how to perform a task based on doing previous tasks?
• Is the amount of time that it takes to accomplish a task satisfactory?
• Is the system visually appealing?
• Is the system's content appropriate for your knowledge and training?
• Can you easily find the system's help resource and is it useful?

Table 3: Data and relationship types supported by the functional prototype.

Data type	**Species**	**Relationship**	**Related to**
Gene	*human, mouse*	in orthologous relationship to	gene
		has allele	variant
		has phenotype	phenotype
		is associated with	disease
Genotype	*mouse*	has part	variant
		has phenotype	phenotype
Variant	*human, mouse*	is allele of	gene
		is part of	genotype
		is associated with	disease
Phenotype	*human, mouse*	is associated with	gene
		is associated with	genotype
		is similar to	phenotype
		is associated with	disease
Disease	*human*	is associated with	gene
		is associated with	variant
		is associated with	phenotype

Table 4: User feedback on the mockups.

	User					
The user...	**C1**	**C2**	**S1**	**S2**	**R1**	**R2**
• Understood the meaning of nodes.	◆	◆	◆	◆	◆	◆
• Understood the meaning of metanodes.	◆	◆	◆	◆	◆	
• Was unsure of the meaning of metanodes.						◆
• Understood the meaning of edges.	◆	◆	◆	◆	◆	◆
• Interpreted edge width as indicative of evidence.	◆	◆		◆		◆
• Was misled by varying edge widths.	◆		◆	◆	◆	◆
• Understood organization of the panel.	◆	◆	◆	◆	◆	◆
• Understood connections between graph and panel.	◆		◆	◆	◆	◆
• Was uncertain of graph-panel connections.		◆				
• Grasped concept of filtering nodes with sliders.		◆	◆	◆	◆	◆
• Was unsure of how sliders would be used.	◆					
• Was confused by some text labels.	◆	◆	◆	◆	◆	
• Thought data in the graph lacked provenance.						◆
• Made biological interpretation of gaps in data.					◆	
• Was uncertain how to interpret gaps in data.		◆				◆

Key:

• Favorable response.

• Unfavorable or critical response.

All tables:

C1 and **C2** are individual clinicians.

S1 and **S2** are individual students.

R1 and **R2** are individual researchers.

Table 5: Feature requests emanating from discussion of the mockups.

	User					
Interface feature	**C1**	**C2**	**S1**	**S2**	**R1**	**R2**
Filtering on one or more node types.	◆				◆	
Simplified panel layout.	◆					
Clearer, more detailed contextual information.				◆	◆	◆
Breadcrumbs/history of visited nodes.				◆	◆	◆
Grouping/clustering nodes.		◆	◆	◆	◆	◆
Hiding or removing unwanted nodes.	◆	◆	◆			
Contextual popup on right-click or hover.	◆		◆			◆
Hierarchical/ontological supplemental display.		◆	◆	◆		
Configurable node shapes.	◆	◆	◆	◆	◆	◆
Configurable node textures.				◆	◆	
Configurable node colors.		◆	◆	◆	◆	◆
Configurable edge styles.			◆	◆		
Configurable edge colors.		◆		◆		
Configurable evidence types in edges.	◆					
Configuring interface to focus on select species.	◆	◆				
Showing/hiding overlays of particular information.			◆			
Undoing an action.	◆					
Indicating richness of potential paths.		◆				
Genetic pathway information.				◆	◆	
Alternatives to the force-directed graph layout.				◆		

Table 6: Candidate user interface features for graph visualization.

Interactive features	**Included?**
Exploring related nodes.	Yes.
Filtering related nodes by data/relationship type.	Yes.
Manually adding nodes to the graph one at a time.	Yes.
Pinning and unpinning nodes.	Yes.
One-click removal of all unpinned nodes.	Yes.
Breadcrumbs/history of visited nodes.	Partially.
Dragging/repositioning nodes in the graph.	Partially.
Contextual popup on right-click or hover.	Partially.
Hierarchical/ontological supplemental display.	Partially.
User configuration or preferences panel.	Partially.
Filtering related nodes by strength of association.	No.
Collapsing nodes; dragging and dropping one node into another.	No.
Automated grouping of similar nodes, e.g. by species.	No.
Applying filters to all displayed nodes.	No.
Advanced filtering widgets, e.g. sliders.	No.
Showing/hiding overlays of particular information, e.g. species.	No.
Zooming and panning.	No.
Undoing an action.	No.
Saving the visualization and interaction history.	No.

Visual/content features	**Included?**
Contextual information about selected node or edge.	Yes.
Distinguishing different data types and species.	Yes.
Relative strength of association between nodes.	Partially.
Indicating richness of potential paths.	Partially.
Displaying multiple nodes as a metanode.	No.
Suggesting alternative paths between two nodes.	No.
Genetic pathway information.	No.
Tree and balloon views as alternatives to force-directed graph.	No.

Table 7: User feedback on the functional prototype.

	User					
User response	**C1**	**C2**	**S1**	**S2**	**R1**	**R2**
• Search strategies need to be better understood.				◆		
• Legend misleadingly appeared to be interactive.	◆	◆	◆	◆	◆	◆
• Nodes need more contextual information.	◆			◆		
• Meaning of edge width was intuitive.		◆	◆	◆		
• Edge width had multiple interpretations.	◆				◆	◆
• Relationships between data types easy to discern.	◆	◆	◆	◆	◆	◆
• Data connections and paths were informative.	◆					◆
• Contextual information and links were easy to find.	◆	◆			◆	
• Graph-panel connections were unintuitive.		◆				◆
• Relationship information in panel was unintuitive.		◆				◆
• Item counts in hierarchical list were unintuitive.	◆		◆			
• Publication lists lacked organization.			◆	◆		◆
• Adding/removing nodes was generally intuitive.	◆	◆	◆	◆	◆	◆
• The right-click node popup was intuitive.	◆				◆	
• Popup and panel functions were inconsistent.		◆		◆		
• Mouse-click conventions were unintuitive.	◆	◆		◆		
• Feedback necessary to prevent repeat user actions.		◆		◆		
• Interface needs to use tooltips more liberally.			◆		◆	

Key:

• Favorable response.

• Unfavorable or critical response.

Table 8: Feature requests emanating from prototype evaluation.

	User					
Interface feature	**C1**	**C2**	**S1**	**S2**	**R1**	**R2**
◦ Video/animation in the help resource.	◆	◆	◆	◆		◆
◦ A robust search interface.		◆		◆	◆	
• Fully functional breadcrumbs/history.			◆	◆	◆	
• Undoing an action.		◆	◆		◆	◆
◦ Zooming and panning.				◆		◆
◦ Saving the visualization and interaction history.		◆				◆
• More robust user configuration.				◆		
◦ Suggesting alternate paths through the graph.	◆					
◦ Listing of the knowledge graph's non-visible nodes.				◆		
• Alternatives to the force-directed graph layout.				◆		

Key:

◦ New request.

• Repeats a request from earlier user sessions.

Table 9: System Usability Scale scores.

	User						
Statement	**C1**	**C2**	**S1**	**S2**	**R1**	**R2**	**Mean**
1	5	4	4	2	4	4	• 3.4
2	2	2	1	1	1	3	• 1.7
3	4	5	4	4	4	4	• 4.0
4	2	1	3	1	1	2	• 2.0
5	4	5	4	4	5	4	• 4.4
6	1	2	3	2	1	3	• 2.6
7	4	5	3	4	3	3	• 4.1
8	2	1	2	3	1	2	• 2.7
9	3	4	3	4	2	4	• 4.1
10	3	2	2	1	2	4	• 3.4

Key:

- Clearly favorable mean response.
- Slightly favorable mean response.
- Slightly unfavorable mean response.