From Genes to Pathways and Back Again: A Visual Analytical Approach

Bryant Dang BS[1], Justin A. Drake BS[1], Gowtham Bellala PhD[2], Suresh K. Bhavnani PhD[1]
[1]Inst. for Translational Sciences, UTMB; [2]Hewlett Packard Laboratories, Palo Alto, CA

Abstract

A critical goal of biomarker discovery is to infer biological pathways based on co-occurring genes. Unfortunately, most gene-pathway databases provide very basic features such as Boolean searching that do not fully support the comprehension and exploration of the complex overlap between genes and pathways. We therefore repurposed an existing visual analytical system to guide in the search of pathways based on genes, and to enable the identification of additional genes from the narrowed set of pathways. A pilot study with a domain expert suggested that while the above two functionalities were useful, there was a need for a more integrated search and exploration approach that enabled (1) pathway narrowing based on inputted genes, (2) gene discovery from narrowed pathways, (3) gene narrowing based on inputted pathways, and (4) pathway discovery from narrowed genes.

Introduction

Despite the increasing need of researchers to comprehend the relationship between genes and pathways, current online databases provide very basic search and exploration capabilities to support such complex tasks. For example, a researcher typically starts with a set of genes, and attempts to narrow down the pathways that contain those genes. However, the researcher might find additional genes from the narrowed set of pathways, leading to yet other pathways and their genes. Given the complex nature of this search and discovery process, we posed the question: *Can a visual analytical system enable the rapid comprehension of the relationship between genes and pathways?*

Method

We repurposed an existing visual analytical system called <u>M</u>ining <u>A</u>nd <u>I</u>nterpretation of <u>D</u>iagnostic <u>N</u>etworks (MAIDN)[1], to enable searching a gene-pathway database of 1267 genes and 217 pathways extracted from BioCarta. As shown in Figure 1, MAIDN provides (1) an **input pane** which creates a dynamically generated list of genes ranked by their ability to eliminate the remaining pathways (e.g., the user has pointed to the top ranked gene JAK3), (2) an **undo pane** which lists the inputted genes (e.g., MAPK3, SHC1, ELK1, CSNK1A, and JAK1) that can be unselected, (3) a **visualization pane** which provides "at-a-glance" the pathways that are eliminated (white dots), and candidate pathways (black dots with labels) that contain all the inputted genes, and progressively move into smaller rings towards the inner red circle, and (4) an **information pane** that displays the full list of genes for a selected pathway (pink dot). We explored the usefulness of MAIDN by observing a domain expert search for pathways based on genes implicated in rickettsioses, followed by a semi-structured interview to obtain feedback on the system.

Results and Conclusion

The domain expert was successful in inputting a set of genes in the input pane to narrow down the candidate pathways, and in clicking in the visualization pane on a pathway node to explore all the genes it contained. The latter functionality he noted was difficult to do in systems like BioCarta. However, he was unable to rapidly identify the maximal set of gene overlap among the candidate pathways, which had to be done by clicking on each pathway and remembering the genes they contained. Furthermore, he noted a need to narrow down genes based on pathways. The pilot study therefore helped to pinpoint four use cases that together appear critical for searching and exploring gene-pathway databases: (1) **pathway narrowing** based on inputted genes, (2) **gene discovery** from narrowed pathways, (3) **gene narrowing** based on inputted pathways, and (4) **pathway discovery** from narrowed genes. These use cases suggest the need for a tight integration of features that help narrow and expand candidate genes and pathways, which should enable a two-way exploration and comprehension of genes and pathways.

Acknowledgements

References

1. Bhavnani, S.K., et al. Rapid Identification of Toxic Chemicals during Emergencies: Integrating Search with Visual Analytics. *Proceedings of AMIA'11*.
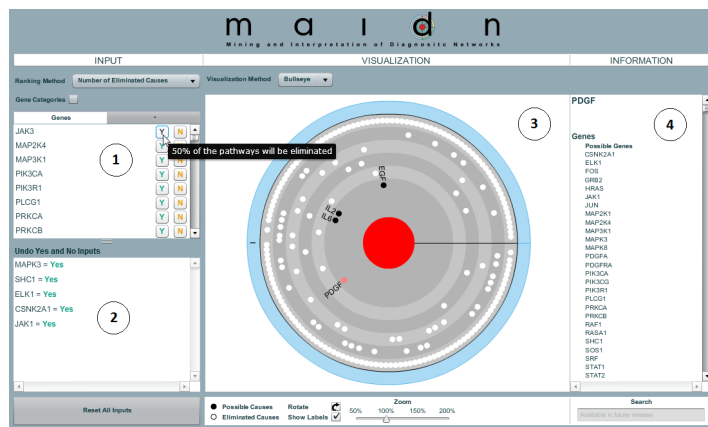
Figure 1. MAIDN integrates (1) guided input of genes, (2) viewing and undoing of inputs, (3) visualization of pathways with different numbers of genes that match the inputs placed in concentric rings, and (4) listing of all genes in a selected pathway (pink node).