As genome-wide association studies rapidly identify genetic loci across implicated loci (or GRAIL), that uses statistical text-mining genetic findings (Iossifov et al., 2008) to identify key pathways and biological processes suggested for a broad range of phenotypes, investigators are critically focused on identifying common function among genes near associated SNPs across a wide range of phenotypes including height (Lango Allen et al., 2010), rheumatoid arthritis (Raychaudhuri et al., 2009b), Crohn’s disease (Franke et al., 2010) and cancer (Beroukhim et al., 2010). While the GRAIL statistical approach calculates the statistical significance of the number and strength of functional similarity across loci, it does not concisely illustrate functional similarities in an intuitive fashion that reveals the underlying biology. Our goal was to provide a visualization that allowed users to see more clearly the underlying genes and biological functionality driving the GRAIL statistical scores.

VIZ-GRAIL: visualizing functional connections across disease loci
Soumya Raychaudhuri

Divisions of Genetics and Rheumatology, Brigham and Women’s Hospital, Partners Center for Personalized Genomic Medicine, Boston, MA 02115 and Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA
Associate Editor: Jonathan Wren

ABSTRACT
Motivation: As disease loci are rapidly discovered, an emerging challenge is to identify common pathways and biological functionality across loci. Such pathways might point to potential disease mechanisms. One approach is to look for functionally related genes across loci and to then assess if that degree of similarity is more than might be expected by chance (Raychaudhuri et al., 2009a). We have separately described a computational strategy, Gene Relationships Across Implicated Loci (GRAIL), to identify whether pair-wise gene relationships defined using PubMed text similarity are enriched across loci. Here, we have implemented VIZ-GRAIL, a software tool to display those relationships and to depict the underlying biological patterns.

Results: Our tool can seamlessly interact with the GRAIL web site to obtain the results of analyses and create easy to read visual displays. To most clearly display results, VIZ-GRAIL arranges genes and genetic loci to minimize intersecting pair-wise gene connections, VIZ-GRAIL can be easily applied to other types of functional connections, beyond those from GRAIL. This method should help investigators appreciate the presence of potentially important common functions across loci.


Contact: soumya@broadinstitute.org

Supplementary Information: Supplementary methods and data are available at Bioinformatics online.

Received on March 5, 2011; revised on April 2, 2011; accepted on April 4, 2011

1 INTRODUCTION
As genome-wide association studies rapidly identify genetic loci for a broad range of phenotypes, investigators are critically focused on identifying key pathways and biological processes suggested by genetic findings (Iossifov et al., 2008; Lage et al., 2007; Perez-Iturbe et al., 2007; Rossin et al., 2011; Wang et al., 2007). We have separately described a computational strategy, Gene Relationships Across Implicated Loci (or GRAIL), that uses statistical text-mining strategy to rapidly identify genes across multiple loci that are similar to each other, and to then assess if that degree of similarity is more than might be expected by chance (Raychaudhuri et al., 2009a). The approach depends on pairs of related genes using 525 000 PubMed article abstracts identified using word similarity metrics. GRAIL has now been applied to prioritize SNPs for replication or to demonstrate common function among genes near associated SNPs across a wide range of phenotypes including height (Lango Allen et al., 2010), rheumatoid arthritis (Raychaudhuri et al., 2009b), Crohn’s disease (Franke et al., 2010) and cancer (Beroukhim et al., 2010). While the GRAIL statistical approach calculates the statistical significance of the number and strength of functional similarity across loci, it does not concisely illustrate functional similarities in an intuitive fashion that reveals the underlying biology. Our goal was to provide a visualization that allowed users to see more clearly the underlying genes and biological functionality driving the GRAIL statistical scores.
As a realistic example, we plot the literature-based similarity across the figure (see Supplementary Material). Briefly, we define an objective function that calculates the total burden of intersections, weighing intersections between thicker connections more heavily. Then we iteratively chose random loci with at least one gene with an intersecting connection, and then we try manipulating the arrangement by either (i) moving the locus to each of the different positions in the circle, (ii) swapping the locus with every other locus in the circle or (iii) inverting different segments of the circle starting from that locus and ending at other positions. At each iteration, we chose the manipulation that most reduces the total number of intersecting connections and update the arrangement iteratively. Once the loci have been arranged, then genes within each of the loci are permuted to reduce the number of total intersections.

3 EXAMPLES

We present examples of VIZ-GRAIL runs in Figure 1. The files used to create this figures are provided in the Supplementary Material. Figure 1A and B presents an illustrative example. In this case, there is a single optimal solution without any intersecting connections. In Figure 1A, the regions and genes are plotted without arranging to minimize intersections—the display looks jumbled and it is difficult to see any clear patterns. VIZ-GRAIL is able to find the optimal arrangement in 184 iterations run on a personal laptop in <1 h (Fig. 1B); the connections between genes and loci are much more clear. As a realistic example, we plot the literature-based similarity across 34 known rheumatoid arthritis (RA) risk loci, implicating a total of 132 genes (Raychaudhuri, 2010). After using VIZ-GRAIL to arrange genes and loci to minimize intersections, related genes are clearly seen, for example the genes involved in the IL2 pathway (IL2, IL2RA and IL2RB), as well as the CD28-CTLA4 pathway.

ACKNOWLEDGEMENTS

We acknowledge the generous institutional support from Brigham and Women’s Hospital and the Broad Institute. We also acknowledge the feedback from Mark Daly, Andre Frank and the GRAIL user community.

Funding: National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIH-NIAMS) Development Award (1K08AR055688).

Conflict of Interest: none declared.

REFERENCES


