KnomePathways™: A Tool to View and Query Gene Interaction Networks in Individual Human Genomes

Adem Albayrak[1]
Knome, Inc.

Harish Mahadevan[2]
Knome, Inc.

James M. D'Augustine[3]
Knome, Inc.

Nathaniel M. Pearson[4]
Knome, Inc.

Abstract

Common diseases and other complex human phenotypes likely trace to interactions between variants in distinct genes. Spotting such interactions is hard, as each person's genome carries millions of distinctive variants, distributed among thousands of genes. To overcome this challenge, we present KnomePathways, a tool that overlays variants found in individual human genomes onto publicly annotated sets of interacting or co-expressed genes. Using a simple color scheme, rich underlying annotation, and powerful comparative querying, KnomePathways lets users quickly find gene-pairwise interactions that may be functionally disrupted – and, more broadly, networks of genes that are enriched for suspect variants in some genomes (e.g., cases) versus others (e.g., controls).

Keywords: Pathway visualization, Gene network, Variant shortlisting, Protein interaction, Whole genome analysis, Exome, Connected graph
Index terms: D.0 General

1. Introduction

The genetic basis of human disease tends to be complex, likely involving many interactions among rare and common sequence variants – especially those in genes. But finding functional interactions between variants in a given human genome is hard, given that each genome carries millions of distinctive variants, and that one must compare multiple phenotyped genomes in order to identify distinctive interactions that govern phenotype. To meet this challenge, we have built a Flex 4-based tool, KnomePathways (Figure 1), to show thousands of gene interaction and co-expression networks through the distinctive lens of a given human genome, by coloring genes in those networks to reflect the class(es) of sequence variants that they carry in that genome. The tool can query networks by gene name, gene- or network-associated phenotype, and comparative criteria that distinguish networks by what kinds of sequence variants they carry in one subset of studied genomes versus another. Networks shown by the tool derive from Human Protein Reference Database [HPRD] data and a special commercial license of the Broad Institute's Molecular Signatures Database (MSigDB) [1].

2. Background

KnomePathways displays publicly annotated sets of genes as graphs comprising nodes (genes) connected by edges (interactions among gene products, as reported in HPRD, KEGG, Reactome, and MSigDb). Nodes are color-classed by a simple scheme to specify the most functionally suspect kind of variant found in that gene in that genome:

- **Gray gene**: each copy encodes a protein identical to that encoded by the human reference genome, and carries no phenotype-implicated synonymous or noncoding variant (per Knome's database of curated publications).
- **Yellow gene**: carries a missense variant predicted to preserve protein function (relative to reference-version of protein), or a phenotype-implicated synonymous or noncoding variant, and no more functionally suspect variant (see below).
- **Orange gene**: carries a missense variant predicted to alter protein function, but no more functionally suspect variant.
- **Red gene**: carries a heterozgous nonsense, frameshift, or splice variant, but no homozygous or second such variant.
- **Red-ringed black gene**: carries a homozygous nonsense, frameshift, or splice variant, or more than one heterozygous such variant.
- **Green-haloed gene**: carries a novel variant that defines the main color (yellow, orange, red, or red-ringed black).

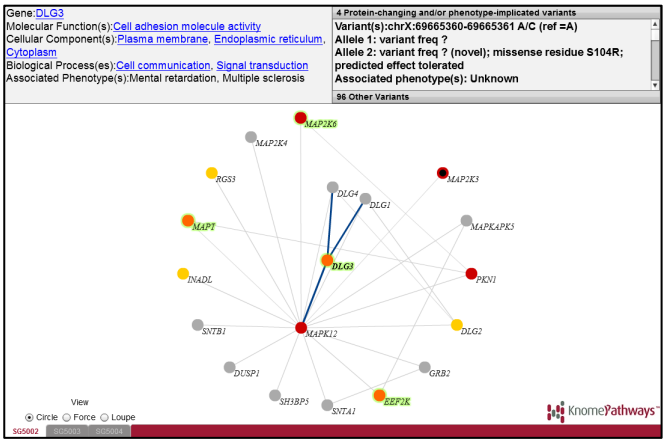

Figure 1: KnomePathways *MAPK12* focal gene network in one of three studied genomes (see tabs at bottom left)

Users can highlight genes in a network to see which genes of a given color carry a particular kind of color-definitive variant. Although the most functionally suspect variant class determines the color of any given gene in a given genome, all other called reference-mismatching variants are detailed in the subject-specific text box at the top right of the user interface by selecting a given

1. aalbayrak@knome.com 2. hmahadevan@knome.com 3. jdaugustine@knome.com 4. npearson@knome.com

gene. A variant not seen before in our variant catalog is considered novel, and the gene carrying it will appear with a glowing green halo.

KnomePathways includes an interface for finding networks by constituent gene name, associated phenotype, or annotated network name. And an advanced search feature lets users find networks carrying at least $k > 0$ sufficiently suspect variants in each of a chosen set of studied genomes, but not in another chosen set. This feature can help users shortlist networks enriched for potentially phenotype-relevant variants in disease cases versus healthy controls, tumor genomes versus healthy tissue genomes, good versus poor responders to a particular drug, etc.

There are three modes for viewing a pathway, all built using the software library Flare[2], written in ActionScript. Circle view surrounds a chosen gene with its first- and higher- (by double-clicking) order protein interaction neighbors, showing interactions among the latter too. Force view puts a sum force on each node proportional to its edge count. Loupe view locally magnifies user-chosen subnetworks, and is especially useful for picking out nodes in dense networks while retaining overall perspective. Users studying multiple genomes can toggle among genome-specific views of a given network by selecting tabs associated with a given subject.

3. Source Data

We use HPRD and MsigDB data to define interactions among genes, and leverage these by letting the user search for any given gene and view a *focal gene network* comprising all genes reported to directly interact with the gene of interest. MSigDB contains five major collections: *positional gene sets* which correspond to human chromosomes; *curated gene sets* from various domain experts, pathway databases (except for BioCarta), and publications in PubMed; *motif gene sets* where genes were shown to share cis-regulatory motifs across species; *computational gene sets* where genes had some association with cancer and *Gene Ontology (GO) gene sets* where genes contained matching GO terms. By combining the data in HPRD and MsigDB, keyed by HPRD identifiers (due to inconsistencies in gene names),we can record whether a gene in a given MSigDB network interacts with another gene in that network, and the user can double-click on any gene to add the rest of its focal gene network to the view, including interactions among added genes and between added genes and those originally shown focal gene network.

We integrate the data from these networks with richly annotated data from the output of the Knome Genome Analysis Platform (kGAP). For studies of a given phenotype, Knome curates relevant published research., bolstering our reference database. We also report allele frequency data, and algorithmic predictions (SIFT) of whether an amino acid substitution likely affects protein function. KnomePathways is, to our knowledge, the first tool to richly functionally annotate and visualize putatively interacting human gene sets through the lens of thorough individuated genome data..

4. Future Work

The current release of KnomePathways uniformly weights edges between genes, and does not report edge directionality. Future releases will provide more data to the user, such as interaction type, regulatory directionality, and strength of evidence.

The power of comparative analysis among multiple genomes is immense, and further refinements to KnomePathways will enhance comparative querying (making it sensitive to network size, etc.), and comparative viewing of networks. Further features will let KnomePathways interface more closely with other Knome and public genomics tools.

References

[1] Molecular Signatures Database (MSigDB), Subramanian, Tamayo, et al. (2005, PNAS 102, 15545-15550)

[2] Flare http://flare.prefuse.org/

[3] SIFT http://sift.jcvi.org/

Acknowledgements

We thank Alexander Baumann for a thoughtful revision.

1. aalbayrak@knome.com 2. hmahadevan@knome.com 3. jdaugustine@knome.com 4. npearson@knome.com