

## Poster B-26

### NUDLE: A Tool for Inferring Nutrient Disease Relationships from Metabolic Networks



#### Authors:

Anoop Kumar ( *Department of Computer Science, Tufts University, Medford, MA, USA* )

Jennifer S. Buell ( *Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, USA* )

Lenore J. Cowen ( *Department of Computer Science, Tufts University, Medford, MA, USA* )

Inga Peter ( *Institute for Clinical Research and Health Policy Studies, Tufts-New England Medical Center, Boston* )

**Short Abstract:** We have developed NUtrient-Disease Link Estimator (NUDLE), a tool that infers nutrient disease relationships using data from metabolic, genomic and disease databases. It constructs metabolic networks and allows the inference of relationships between genes, diseases and nutrients from distance properties within the networks.

#### Long Abstract:

##### Introduction

Associations between nutrients and diseases have long been recognized. The availability of genomic, disease and metabolic data combined with computational technologies creates a possibility of inferring nutrient gene interactions and relating them to diseases.

There is a variety of publicly available databases that present information on nutritional, genomic, and metabolic diseases. Metabolic data is available in databases such as MetaCyc [1] and KEGG [2]. The quantity of information available on metabolic pathways is increasing leading to problems of spatial and temporal computational complexity for analysis. Data models usually implemented as in-memory structures containing information from these databases for analysis are well studied [4]. Several groups have applied graph theoretic approaches to analyze the metabolic data [5]. Metabolic Pathways have been analyzed using distance metrics and linear programming, [6] and PETRI nets [7]. Gene associated information is available at Entrez Gene. Information pertaining to diseases and their connections with genes is available in OMIM [3]. However, these heterogeneous sources of information have not previously been integrated into a single resource with a focus on the interactions between nutrients and diseases. We are developing a tool that uses data available in metabolic, genomic and disease databases, annotates nutrients in the networks and infers nutrient-gene disease relationships.

The relational database represents compounds and reactions from KEGG [8] in a bipartite graph. A link connects a compound to a reaction if that compound was part of that reaction and no links were made among same types of nodes. Since most of the reactions in KEGG database that were part of the bipartite graph were bidirectional the links were not directed. A directed bipartite graph can be constructed in case all reactions are not bi-directional.

Graphs were searched through a depth of 5 reactions starting from a nutrient, where 5 was

arbitrarily chosen to capture major nutrient conversions while not overwhelming the user with excessive information. The problem of traversing shortcuts through highly connected molecules, such as H<sub>2</sub>O, NADP, and ATP was avoided by deleting these highly connected compounds that are involved in hundreds of reactions and thus cannot be assumed to be unique or specific intermediates between two reactions. Lists of all the enzymes that catalyze the reaction are listed along with its distance from root nutrient. Genes in HUGO [9] that mapped to the Enzyme Commission (EC) number of enzymes in KEGG assigned to the nutrients.

Diseases related to the genes in these reactions were obtained from OMIM [10] and linked to the nutrients. Finally, a table for each nutrient along with one or more enzymes, genes, diseases and the depth from origin, i.e., the location of nutrient on graph, is presented. The depth of an enzyme or the disease is given by the number of reactions on the metabolic graph between given nutrient and the enzyme a reaction takes part in. Minimum depth is one for the enzyme that catalyzes the reaction in which the nutrient is a reactant.

## Results

A list of nutrient disease interactions was created by an expert nutritionist to verify the relations generated by NUDLE. The proposed method generated all nutrient-disease associations from the expert list and generated hypothesis for additional relationships. An example of such an additional relationship with support in the literature is a putative relationship between folate and cancer, inferred from pursuing the one-carbon metabolism cycle, where folate is involved in DNA methylation and that a disruption in one-carbon metabolism may result in aberrant DNA. All the relationships inferred along with the related genes and enzymes are available online at (<http://www.nutgen.org/ismb06/NutrientEnzyme.html>). An easy-to- navigate display with a comprehensive search mechanism will soon be available for human browsing.

## References

- [1] Karp, P.D., Riley, M., Paley, S.M., Toole, A.P., (2002) The MetaCyc Database, *Nucleic Acids Research*, 20
- [2] Kanehisa, M., Goto, S., Kawashima, S., Nakaya, A., (2005) The KEGG databases at GenemoNet. *Nucleic Acids Research*, 30
- [3] Karp, P.D., Paley, S.M., Romero, P., (2002) The Pathway Tools Software, *Bioinformatics*, 18
- [4] McKusick, V.A., Mendelian Inheritance in Man. A Catalog of Human Genes and genetic Disorders. The large number of different resources, their heterogeneity, complexity, and inaccessibility make it difficult for instructors and students to actively utilize innovative information in these emerging fields. Karp, P.D., Paley, S.M., Romero, P., (2002b) The Pathway Tools Software, *Bioinformatics*, 18
- [5] Croes, D., Couche, F., Wodak, S. J., Helden, J.V., (2005) Metabolic PathFinding: inferring relevant pathways in biochemical networks, *Nucleic Acids Research*, 33
- [6] Simeonidis, E., Rison, S.C., Thorntom, J.M, Bogle, I.D., Papageorgiou, L.G, (2003) Analysis of metabolic networks using pathway distance metric through linear programming, *Metabolic Engineering*, 5

- [7] Kuffner, R., Zimmer, R., Lengauer, T., (2000) Pathway analysis in metabolic databases via differential metabolic display (DMD), Bioinformatics, 16
- [8] KEGG - <ftp://ftp.genome.jp/pub/kegg/>
- [9] HUGO - [http://www.gene.ucl.ac.uk/nomenclature/data/gdlw\\_index.html](http://www.gene.ucl.ac.uk/nomenclature/data/gdlw_index.html)
- [10] OMIM - <ftp://ftp.ncbi.nih.gov/repository/OMIM/>