MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

Phenotype Analysis for Insights into Complex Inheritance Diseases: a Case Study

Yunli Wang[a], Jesse Li-Ling[b]

[a]*Institute for Information Technology, National Research Council, Canada*
[b]*Department of Medical Genetics, China Medical University, China*

Abstract

Systematic analysis of phenotypic data may provide important clues for delineating developmental mechanisms as well as functional genetic networks. In this study, 68 entries involving congenital pancreatic abnormalities were selected from OMIM. Clinical synopses data were extracted and standardized. Data mining of processed data has suggested that, anatomically, major features of syndromes involving pancreatic defects seems to congregate along the midline of the human body, which conforms to John Opitz's theorem of developmental field defects. We conclude that utilization of phenotypic data can facilitate delineation of developmental as well as genetic mechanisms underlying various types of congenital anomalies.

Keywords:

pancreas; syndrome; phenotype; data mining

Methods

Phenotypic data could be used for obtaining insights into complex diseases. Various phenotypes have usually been considered as isolated features, and studies on high-level phenotypes were rare. Cantor and Lussier (2004) had attempted to use self-organizing maps and hierarchical clustering for clinical phenotypes in order to achieve better understanding for genotypes. However, since the original categories provided in the OMIM entries were not rationalized, it is difficult to judge the usefulness of clustering results. More recently, van Driel et al. proposed an approach for measuring similarities between phenotypes, and demonstrated they are correlated with the protein sequence, protein motifs, and functional annotation. Although this result appears to be promising, meaningful patterns for a particular syndrome still cannot be obtained.

In this study, we explored using data mining techniques for analyzing clinical data derived from congenital syndromes involving pancreatic malformations. 68 pancreatic malformations entries were retrieved from OMIM (as by Nov 1st 2005). After removal of those that are non-congenital in nature, 23 entries were derived. The listed clinical synopses of these entries were converted into 55 categories manually.

The correlation between individual phenotypes among pancreatic malformations was analyzed using statistically significant association rules [3]. The results were also compared with hierarchical clustering and K-means clustering.

Results

Correlations between phenotypes were obtained, among them, kidney and liver has the strongest correlation. This result was consistent with results from hierarchical clustering and k-means clustering. Correlation of the common features among the selected syndromes seems to suggest that, developmentally, pancreas has a close relation with adjacent organs such as the kidney and the liver, whilst the overall pattern of above syndromes conforms to John M Opitz's theorem on the developmental field defect, that the midline structure of the body are most frequently involved among congenital syndromes.

Conclusion

Phenotype analysis may facilitate functional genomics and delineation of genotype ~ phenotype correlations. Data mining for a particular feature can be used for inferring closeness between genes and delineating genetic networks. Our future work will include development of a system that can automatically translate retrieved data into a comparable format, as well as utilization of laboratory experiment to validate the derived results.[1]

References

[1] Cantor MN, Lussier YA. Mining OMIM for insight into complex diseases, Medinfo 2004, pp. 753-757.
[2] van Driel MA, Bruggeman J, Vriend G, Brunner HG, Leunissen JA. A text-mining analysis of the human phenome, *European Journal of Human Genetics*, 2006, 14: 535-542.
[3] Brin S, Motwani R, Silverstein C. Beyond market baskets: generalizing association rules to correlations, Proceedings ACM SIGMOD International Conference on Management of Data, SIGMOD 1997, pp.265-276.

1 This study has been sponsored by a grant (to J. L.) from the National Science Foundation of China (No.60574040).

Address for correspondence

Yunli Wang, Address: 46 Dineen Drive, Fredericton, NB, E3B
9W4, Canada. Phone: 506-444-0552
Email: Yunli.Wang@nrc-cnrc.gc.ca.