BMC Systems Biology

BioMed Central

Poster presentation

Integration and analysis of heterogeneous microarray data sources for supporting drug target identification in atherosclerosis

Anyela Camargo*[1] and Francisco Azuaje[1,2]

Address: [1]University of Ulster at Jordanstown, School of Computing and Mathematics, Shore Road, Newtownabbey, Co. Antrim, BT37 0QB, Northern Ireland, UK and [2]Systems Biology Research Group, University of Ulster

Email: Anyela Camargo* - av.camargo@ulster.ac.uk

* Corresponding author

This abstract is available from: http://www.biomedcentral.com/1752-0509/1?issue=S1

Background

Atherosclerosis is one of the major causes of morbidity and mortality in industrialised countries. Despite the introduction of new pharmacological drugs, this tendency continues to grow as the world changes food habits that fit into people's life styles. Atherosclerosis is an inflammatory disease in which high concentrations of cholesterol accumulate around the wall of blood vessels [1]. The implementation of large-scale *in vitro* and *in silico* research is fundamental to discover significant patterns and pathways involved in the disease progression. This research integrates and analyses two heterogeneous microarray data sets. The study led to the identification of genes, biological processes and pathways that may be used to determine the progression of coronary artery disease (CAD) in humans. Figure 1 illustrates the integrative data mining procedure implemented in this study.

Results

Two heterogeneous data sets obtained from the GEO (Gene expression Omnibus): Aortic stiffness (AS) and human coronary artery disease (CAD) studies were analysed and integrated. After normalisation, scaling and harmonisation, the data were analysed upon two different approaches. The first approach focused on uncommon genes, i.e. those included in AS but not in CAD. The second study focused on the expression patterns of common genes shared by both data sets. The latter analyses yielded a list of significantly differentiated expressed genes. To verify the potential biological significance of the results the genes were furthered assessed based on their involvements in different biological processes as defined by GO-driven annotations and published papers. The lists of significant genes from each study were ranked based on their relevance encoded in public, external functional databases. Additionally, text mining allowed the identification of a list of documents relating such significant genes to the disease. Many of the genes identified in this study proved to have strong relations with atherosclerosis. Some genes are relevant to disease control, severity and progress. For instance, the study stresses the roles of key genes (e.g. *TNFRSF1B, MAP2K1*) and pathways linked to the expression of antimicrobial peptides defensins, which may be associated with inflammation and lipid accumulation in atherosclerosis. The study also identified key biological patterns and genes related to "programmed cell death" and "apoptosis", which describe disease state and degree of degeneration.

Conclusion

This investigation generated a list of genes and biological processes that can be strongly associated with processes relevant to atherosclerosis. Some of the genes highlighted (Figure 1) may be directly related to the disease progression and control. This study shows how the large-scale, computational integration of heterogeneous microarray data sets, functional annotation databases and published literature may support the identification and assessment
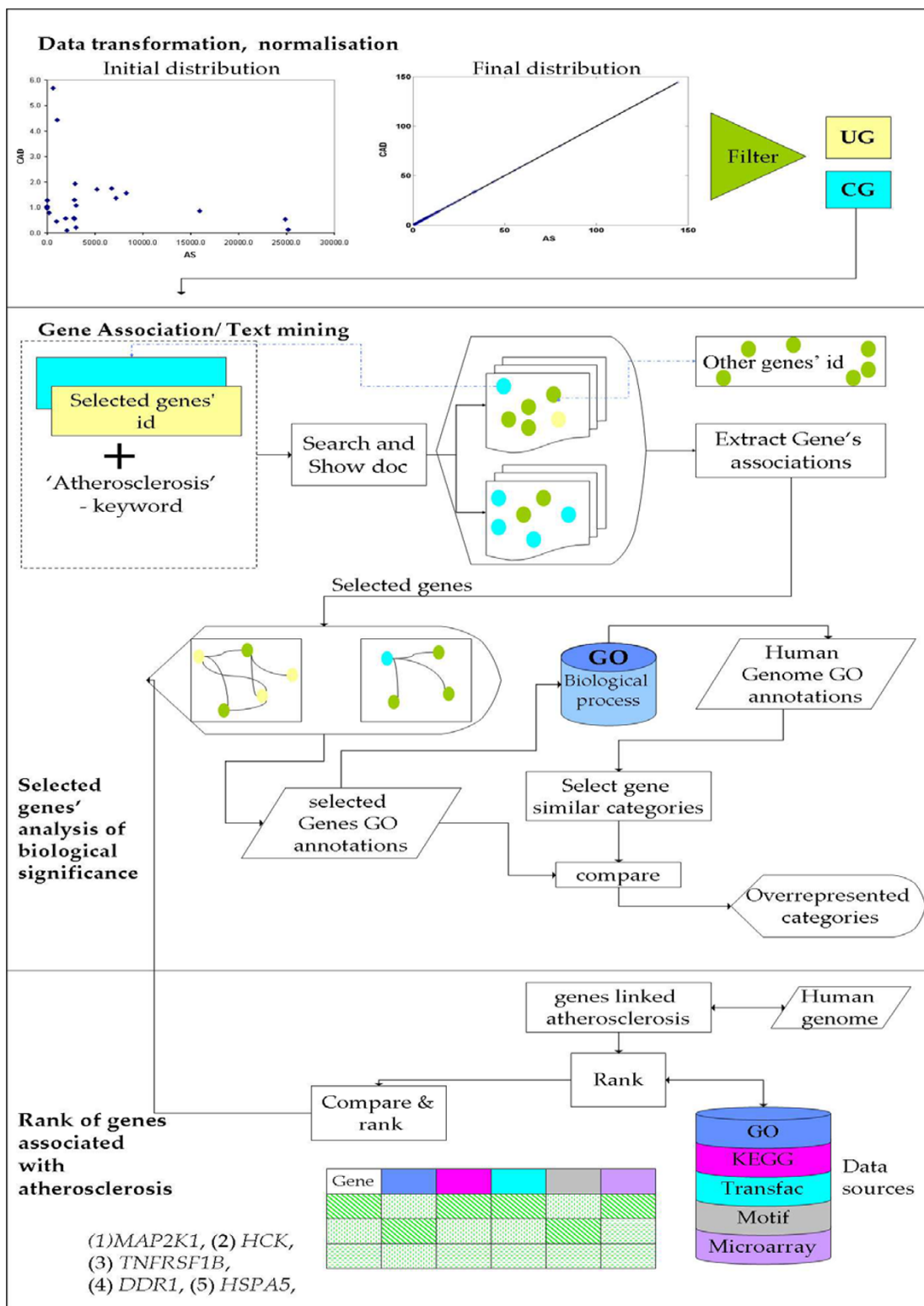
Figure 1
Data mining methodology for the integration and analysis of microarray data relevant to understand the progression of human atherosclerosis.

of potential therapeutic targets. It also demonstrates how integrative data mining may allow scientists to recover essential patterns and unknown relationships that may be overlooked when single studies were carried out in the first place. In this particular case, a set of representative disease-related genes were detected, which are suggested as testable hypotheses in relation to their roles in CAD progression.

References

1. Ross R: **Atherosclerosis – an inflammatory disease.** *N Engl J Med* 1999, **340(2):**115-26.