

POSTER PRESENTATION

Open Access

Towards automated metabolome assembly: application of text mining to correlate small molecules, targets and tissues

P Moreno^{*}, KV Jayaseelan, C Steinbeck

From 6th German Conference on Chemoinformatics, GCC 2010
Goslar, Germany. 7-9 November 2010

How many species are there on Earth? Single celled organisms aside a few million are known. How many species had their genomes sequenced so far? Around two thousand. Approaching the era of a new genome sequence per week, it is fair to wonder: How many metabolomes have been compiled? The metabolome could be considered the ultimate phenotypic expression of the cell, and yet, we barely have one [1]. The metabolome refers to the complete set of small molecules (< 1500Da) present on a biological sample or organism [2]. Among all omics, metabolomics is probably the most unique to each individual in terms of the variation of its elements. Here we define Automated Metabolome Assembly (AMA) as a set of techniques to predict the metabolome of an organism based on the complete set of boundary information available (gene sequence, proteomics, bibliomics, etc.). As a first step towards Automated Metabolome Assembly we report the implementation of a text mining resource based on the existing EBI text mining infrastructure [3] to address the problem of finding co-occurrences of chemical entities, proteins, organisms, and tissues/cell types terms. We created a workflow based on a database holding 365 million of relations between these terms (proteins, metabolites, organisms and tissues) and PubMed citations, obtained from the whole PubMed collection up to September 2009. Dictionaries of terms for each kind of entity were generated from different ontologies. All known metabolites present in liver were obtained from the latest version of HMDB. The text mining results were compared to this reference set. Close to 90% of

the reference set shows co-occurrence of our liver-related tissue ontology terms with the respective metabolite names, demonstrating that this text-mining workflow can form an important building block for a comprehensive system for metabolome prediction.

Published: 19 April 2011

References

1. Wishart D: Hmdb: the human metabolome database. *Nucleic Acids Res* 2007, **35**(Database issue):D521.
2. Wishart D: Current progress in computational metabolomics. *Briefings in Bioinformatics* 2007, **8**:279.
3. Rebholz-Schuhmann D, Arregui M, Gaudan S, Kirsch H, Jimeno A: Text processing through web services: calling whatizit. *Bioinformatics* 2008, **24**:296.

doi:10.1186/1758-2946-3-S1-P19

Cite this article as: Moreno et al.: Towards automated metabolome assembly: application of text mining to correlate small molecules, targets and tissues. *Journal of Cheminformatics* 2011 **3**(Suppl 1):P19.

Publish with **ChemistryCentral** and every scientist can read your work free of charge

"Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge."

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
<http://www.chemistrycentral.com/manuscript/>



ChemistryCentral

* Correspondence: pmoreno@ebi.ac.uk
Cheminformatics and Metabolism Group, European Bioinformatics Institute (EBI), Cambridge, CB10 1SD, UK