Development of text-mining solutions to facilitate lipid metabolism interpretation in Genome-Scale Metabolic Models

Adriano Silva[1], João Ribeiro[1], and Emanuel Cunha[1]

University of Minho

Abstract. The abstract should briefly summarize the contents of the paper in 15–250 words.

Keywords: First keyword · Second keyword · Another keyword.

1 Introduction

1.1 Context and motivation

Following the new era of digital development, in the past decades, the whole science is experiencing a great evolution and growth starting with the investment in education and facilities and ultimately leading to the innovation or the creation of a new "value". The exponential growth of the science produced can be measured by the number of articles and citations that are been made through these years[2, 4]. This is self-evident in the field of omics where the advancement in sequencing techniques and the cost-effectiveness allowed the generation of the so-called Big Data that almost grew by forty times in the last decade [5, 4].

Data generated by the multi-omics is exceptionally useful in the understanding of the different cell components, as well as in the interactions between them. Here Systems Biology plays an important role in the unveiling of the cellular secrets that are needed to survive and thrive [4, 6]. Taking advantage of the multi-omics data gathered, and trying to comprehend the metabolism of a determined organism, the construction of genome-scale metabolic models (GSM) took the spotlight. Allowing the seeking for new combinations in the cell network of interactions, this tool as showned useful not only to deepen the knowledge of an organism but also to the benefit of the humankind [4, 3].

The growing amount and complexity of data generated by multi-omics aligned with the increasing need to have easily accessible, well structured and cured data have brought to light some challenges. GSM approaches integrate all known metabolic information to predict the cellular metabolism and eventually trying to optimize the output of a compound with some interest [4, 3]. It is therefore imperative to integrate in GSM that can allow the bests results. But in fact, the lack of biochemical and structural information in conventional databases is a problem [3].

In the particular case of Lipids, their representation is often generic with the lack of structural and biochemical information in the commonly used databases[1]. For this reason, GSM reconstruction does not consider the missing information that is vital to the understanding of the lipids metabolism network. It is in this niche that a tool that is capable to integrate specific information of lipids like BOIMMG acts.

1.2 Objective

Objetivo proposto para este projeto.

2 State of art

2.1 Omics big data

O que é o conceito de omica.

2.2 GSM

explicar o que é GSM

2.3 Lipid problem

Explicar o que é um lipido? e aprofundar o problema.

2.4 Graph database

Explicar o que são bases de dados em grafos onde se enquadra a boimmg.

Please note that the first paragraph of a section or subsection is not indented. The first paragraph that follows a table, figure, equation etc. does not need an indent, either.

Subsequent paragraphs, however, are indented.

Sample Heading (Third Level) Only two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.

Sample Heading (Fourth Level) The contribution should contain no more than four levels of headings. Table 1 gives a summary of all heading levels.
Displayed equations are centered and set on a separate line.

$$x + y = z \tag{1}$$

Please try to avoid rasterized images for line-art diagrams and schemas. Whenever possible, use vector graphics instead (see Fig. 1).

Table 1. Table captions should be placed above the tables.

Heading level	Example	Font size and style
Title (centered)	**Lecture Notes**	14 point, bold
1st-level heading	**1 Introduction**	12 point, bold
2nd-level heading	**2.1 Printing Area**	10 point, bold
3rd-level heading	**Run-in Heading in Bold.** Text follows	10 point, bold
4th-level heading	*Lowest Level Heading.* Text follows	10 point, italic

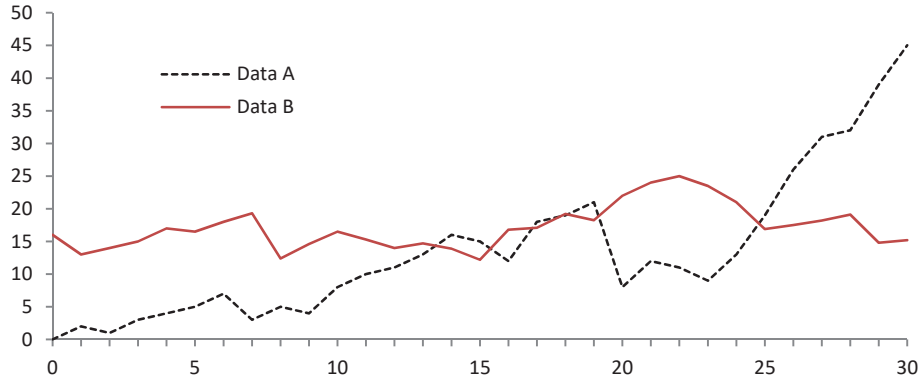

Fig. 1. A figure caption is always placed below the illustration. Please note that short captions are centered, while long ones are justified by the macro package automatically.

Theorem 1. *This is a sample theorem. The run-in heading is set in bold, while the following text appears in italics. Definitions, lemmas, propositions, and corollaries are styled the same way.*

Proof. Proofs, examples, and remarks have the initial word in italics, while the following text appears in normal font.

For citations of references, we prefer the use of square brackets and consecutive numbers. Citations using labels or the author/year convention are also acceptable. The following bibliography provides a sample reference list with entries for journal

References

1. Hnin W. Aung, Susan A. Henry, and Larry P. Walker. Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. *Industrial Biotechnology*, 9:215, 8 2013.
2. Lutz Bornmann and Rüdiger Mutz. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66:2215–2222, 11 2015.
3. Changdai Gu, Gi Bae Kim, Won Jun Kim, Hyun Uk Kim, and Sang Yup Lee. Current status and applications of genome-scale metabolic models.

4. Anurag Passi, Juan D. Tibocha-Bonilla, Manish Kumar, Diego Tec-Campos, Karsten Zengler, and Cristal Zuniga. Genome-scale metabolic modeling enables in-depth understanding of big data. *Metabolites*, 12, 1 2021.
5. Jeffrey A Van Santen, Satria A Kautsar, Marnix H Medema, and Roger G Linington. Microbial natural product databases: Moving forward in the multi-omics era hhs public access.
6. Yawen Zou and Manfred D. Laubichler. From systems to biology: A computational analysis of the research articles on systems biology from 1992 to 2013. *PLOS ONE*, 13:e0200929, 7 2018.