

Molecular epidemiology of pregnancy using omics data: advances, success stories, and challenges

Authors: Ali Rahnavard^{1,2,*}, Ranojoy Chatterjee^{1,2}, Hui Wen^{1,2}, Clark Gaylord^{1,2}, Sabina Mugusi³, Kevin C Klatt⁴, Emily R. Smith^{5,6,*}

Affiliations:

¹Computational Biology Institute, Milken Institute School of Public Health, The George Washington University, Washington, DC 20052 USA.

²Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health, The George Washington University, Washington, DC 20052 USA.

³Department of Clinical Pharmacology, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania.

⁴Nutritional Sciences & Toxicology, University of California, Berkeley, CA 94720, USA.

⁵Department of Global Health, The Milken Institute School of Public Health, The George Washington University, Washington, DC 20052 USA.

⁶Department of Exercise and Nutrition Sciences, The Milken Institute School of Public Health, The George Washington University, Washington, DC 20052 USA.

*Correspondence to rahnavard@gwu.edu and emilysmith@gwu.edu

Omics literature processing: We used the PUBMED database[1] on February 25th, 2022, and to find original researches, we filtered for “*Journal Article*” and used Clinical Trial (Phase I, II, III, and IV) Controlled Clinical Trial, Journal Article, Multicenter Study, Observational Study, and Randomized Controlled Trial. We also filtered species and language by only selecting “human” and “english”. The search terms that were associated with each category are given below in **STable 1**. For investigating literature on pregnancy omics, we added “AND Pregnancy” to our regular expression search.

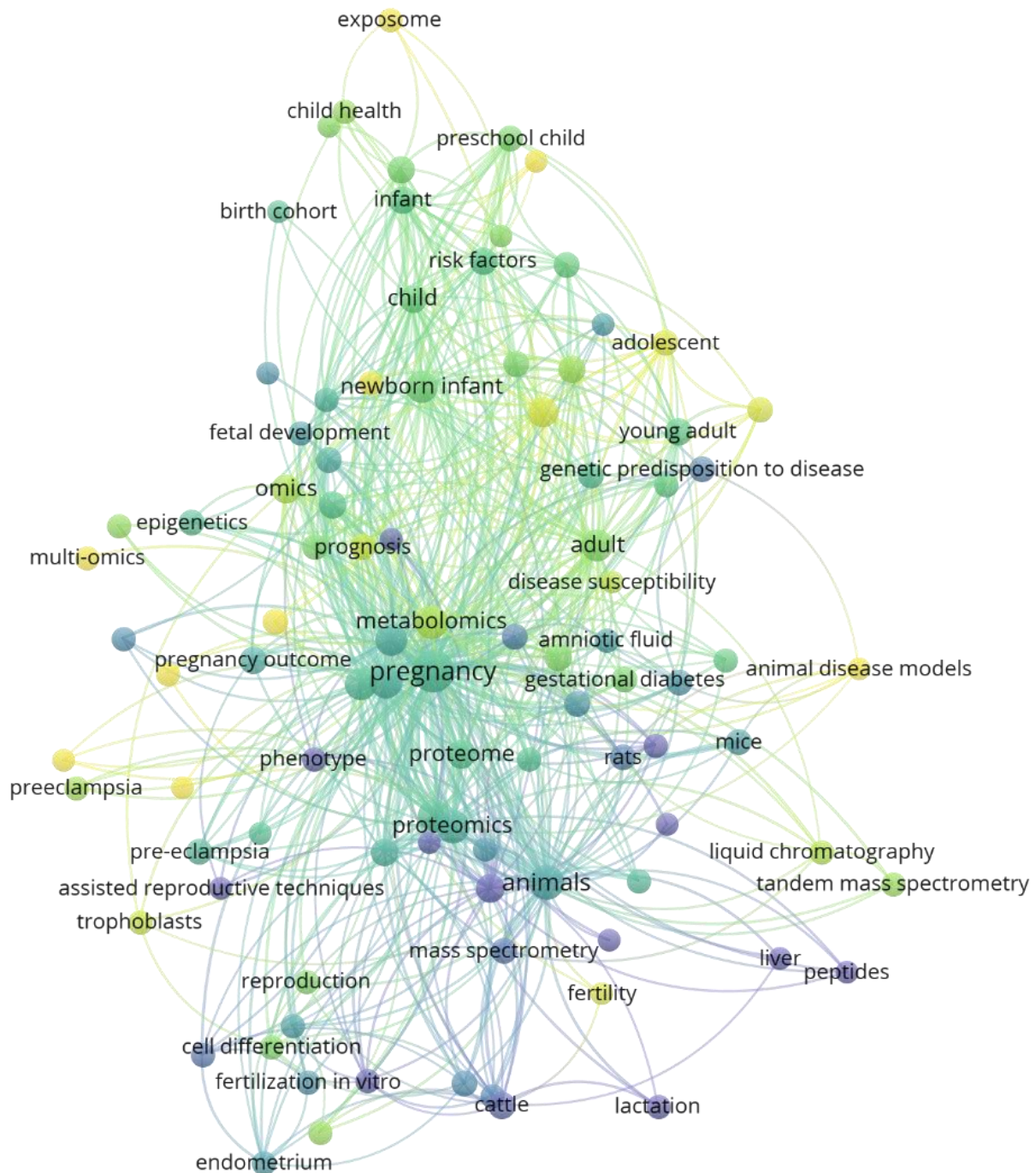
Omics	Studies		
Metabolomics	(metabolom*) AND (NMR)	(metabolom*) AND (GC-MS)	(metabolom*) AND (LC-MS)
Metagenomics	(metagenom*) AND (16S)	(metagenom*) AND (18S)	(metagenom*) AND (Shotgun)
Genomics	(genom*) AND (Microarray)	(genom*) AND (NGS)	(genom*) AND (qPCR)
Epigenomics	(epigenom*) AND (ATAC-SEQ)	(epigenom*)AND (ChIp-SEQ)	(epigenom*) AND (Hi-C)
Proteomics	(proteom*) AND (LC-MS)	(proteom*) AND (NMR)	(proteom*) AND (GC-MS)

Viromics	(virom*) AND (NGS)	(virom*) AND (HTS)
----------	-----------------------	--------------------

STable 1: Search string used to capture publications for each omic.

Network analysis and data mining of literature:

Expression search: 1) (OMICS) AND (PREGNANCY) resulted in 219 papers (**SFig. 1**) colored by year of publication and 2) OMICS search results for 18,502 papers. Omics literature had 39,489 keywords 214 of them co-occurred with the omics keyword at least 100 times. Our network analysis using VOSviewer[2] identified 6 clusters (each specified with different colors (**SFig. 2**)). Papers filtered for categories including Clinical Trial, Clinical Trial, Phase I, Clinical Trial, Phase II, Clinical Trial, Phase III, Clinical Trial, Phase IV, and only Journal Article, Randomized Controlled Trial, and language English, which resulted in 18,502 papers.



SFigure 1: Current important scientific directions of omics utilization in pregnancy research. Abstract of 219 papers with expression search of omics and pregnancy, and condition we process and from 1385 extracted Scientific keywords from abstracts that occurred with pregnancy and omics we show 97 keywords with least 5 co-occurrence. The co-occurrence link with other keywords also is measured and shown as links between keywords. Colors represent year of publication. The network analysis was performed by VOSviewer[2]. Nodes are keywords that are linked by edges for their co-occurrence. Edges reflect the number of co-co-occurrence of keywords in publications used in the analysis. Each color refers to a cluster of keywords that co-occurred in publications.

