Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230349

Electromagnetic Fields Literature Analysis for Precision Medicine

Christopher HAWTHORNE [a*], Donal HARRISON [a*], Eva McEvoy [a*] and Guillermo LOPEZ CAMPOS [a,1]
[a] *Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast,UK*
ORCiD ID: Guillermo Lopez Campos https://orcid.org/0000-0003-3011-0940

Abstract. During the last century technological advances have increased the number of anthropogenic electromagnetic fields (EMFs) and therefore human exposures. In this work we have mined from more than 30,000 EMF-related publications the genes, diseases and molecular mechanisms associated with the exposure to six different subsets of EMFs. Results show 3653 unique disease MeSH terms and 9966 unique genes identified of which only 4340 genes are human. Overall, our approach highlights the molecular aspects of the increasing exposure to EMFs.

Keywords. Bioinformatics, Exposome, Electromagnetic fields, text mining, precision medicine

1. Introduction

The exposome [1] has become an increasingly relevant area of research during the last decade as the environmental counterpart of the genome to understand and explain the health and disease status of an individual. Among the different environmental sources of exposures, technological advances during the last century have caused an increase in the exposure to anthropogenic electromagnetic fields (EMFs). These are nowadays ubiquitous and are originated from multiple sources such as electrical wiring, appliances or telecommunications among others and have been object of an increasing regulatory and social scrutiny regarding their potential effects on human health. Recent controversy about the effects of mobile communications (and particularly 5G technologies) are a clear example of the relevance of research in this area and the increasing volume of publications. In this context, biomedical informatics and text mining offers the opportunity to analyze the corpus of literature associated with EMFs and extract information related with genes and diseases to investigate the molecular aspects related to EMF exposures. Here we present an in-silico analysis of EMF-related literature aiming to extract and compare molecular information and disease.

* Equal contributors

[1] Corresponding Author: Guillermo Lopez Campos, Wellcome-Wolfson Institute for Experimental Medicine, E-mail: g.lopezcampos@qub.ac.uk.

2. Methods

Bibliographic information for the six different EMF subcategories (direct current, electric currents, low frequency, mobile telecommunications, power frequencies and radio frequency) was extracted from the EMF-Portal (https://www.emf-portal.org/en) literature database in ".ris" format and processed to annotate selected documents with their PubMed identifiers (PMIDs). Then were incorporated to an MS Access database where they were further filtered to retain only articles in English. A second stage used PubTator [2] annotations to annotate the selected documents with genes and proteins, and diseases (MeSH terms). These annotations were further filtered to keep only human genes and proteins. These genes and proteins were then used in an additional analytical step, using gene set enrichment analyses to infer diseases and pathways (adj. p-value <0.05) associated with the different subcategories of EMFs, calculated using Enrichr [3].

3. Results and Discussion

A total of 32,986 English documents were retrieved, of which only 25,632 (77.7%) had a PMID and 11,667 (45.6% of PMID annotated documents) were uniquely associated to a single EMF subcategory. Results from PubTator annotations are presented in Table 1. In addition to disease inference, a total of 227 different KEGG pathways were associated as potential mechanisms affected by the EMFs.

Table 1. Results from text mining and gene set enrichment analyses

EMF category	No. of Documents	No. of Genes/proteins	No. of diseases (Found as MeSH)	No. Diseases (Inferred from DisGeNet)
Direct Current	6100	4683	2613	6075
Electric Current	6409	4747	2782	6103
Low Frequency	14732	7270	3317	6713
Mobile Comms	3301	2124	1362	5417
Power Frequency	4776	2824	1989	6200
Radio Frequency	7984	3917	2151	6713

Surprisingly, a large number of unique genes (9966) and diseases (3653) are found to be related with the different subcategories. However, of all the genes identified only 4340 have been annotated as human genes. There is an important redundancy in the genes and diseases found from the text and in the inferred diseases from the gene lists, this is not entirely unexpected as the corpus displays redundancy in the annotation of the different EMF subcategories. Overall, our approach has identified molecular elements related with the exposure to EMFs and these preliminary results require further biological in-depth analysis and will serve as the foundational element for a related graph knowledgebase.

References

[1] Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev. 2005;14: 1847–50

[2] CH Wei, A Allot, R Leaman, Z Lu PubTator central: automated concept annotation for biomedical full text articles. Nucleic Acids Res. 2019 Jul; 47(W1) W587-W593

[3] Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013 April; 14:128