Advances in Clinical Medical Research and Healthcare Delivery

Volume 1
Issue 1 *Inaugural Issue*

Article 2

2021

Knowing What We Know: Leveraging Community Knowledge Through Automated Text-Mining

Justin Gardner
Rochester Institute of Technology, jmg7932@g.rit.edu

Jonathan Tory Toole
Rochester Institute of Technology, jttihst@rit.edu

Hemant Kalia
The Spine Center at Ridgeway, Rochester Regional Health, Rochester, NY, USA, drhemantkalia@gmail.com

Garry Spink Jr.
Behavioral Health Services, Rochester General Hospital, Rochester, NY, USA, garry.spink@rochesterregional.org

Gordon Broderick
Rochester General Hospital, gordon.broderick@rochesterregional.org

Recommended Citation

Gardner J, Toole JT, Kalia H, Spink G, Broderick G. Knowing What We Know: Leveraging Community Knowledge Through Automated Text-Mining. *Advances in Clinical Medical Research and Healthcare Delivery*. 2021; 1(1). doi: 10.53785/2769-2779.1006.

ISSN: 2769-2779

Knowing What We Know: Leveraging Community Knowledge Through Automated Text-Mining

Keywords

text mining, psychoneuroimmunology, inflammation, chronic pain, addiction, computational model, regulatory logic, homeostasis

Cover Page Footnote

The authors would also like to thank Drs. Chris Cheadle, Hongbao Cao, Dicle Hasdemir, Anton Yuriev, Philipp Anokhin, Umesh Nandal and Thibault Geoui of Elsevier BV for their technical assistance and many helpful discussions.

Knowing What We Know: Leveraging Community Knowledge Through Automated Text-Mining

Justin Gardner*[1,2], J. Tory Toole[1,2], Hemant Kalia[3], Garry L. Spink, Jr.[4], and Gordon Broderick[1]

[1] Center for Clinical Systems Biology, Rochester General Hospital, Rochester, NY, USA

[2] Institute of Health Sciences and Technology, Rochester Institute of Technology, Rochester, NY, USA

[3] The Spine Center at Ridgeway, Rochester Regional Health, Rochester, NY, USA

[4] Behavioral Health Services, Rochester General Hospital, Rochester, NY, USA

*** Correspondence:**
Justin Gardner
jmg7932@rit.edu

Keywords: text mining, psychoneuroimmunology, inflammation, chronic pain, addiction, computational model, regulatory logic, homeostasis

To the Editor: Much of the knowledge contained in the peer-reviewed literature has, until recently, been infeasible to extract comprehensively and systematically due to inherent time limitations in scale associated with a manual literature review as well as the inherent researcher bias in the selection of articles and the interpretation of their content.[1,2] However, the advent of the latest natural language processing (NLP) engines hail a new era of efficient, robust and large-scale extraction of high-quality information from peer-reviewed publications, one where published expert-informed interpretations of clinical and experimental data can be used to inform analyses in concert with numerical data. Our group, and others, are now actively leveraging such technologies to construct comprehensive knowledge graphs that aggregate and reconcile documented interactions between molecular, cellular, physiological and behavioral constructs into cohesive, mechanistically rigorous, descriptions of illness that can be interrogated in the broader context of a dynamic and networked biology. [3,4]

Under an ongoing research collaboration between Rochester General Hospital's (RGH) Center for Clinical Systems Biology and Elsevier Life

Sciences, functional relationships linking elements of biology are being extracted from full text in a growing corpus of over 8 million peer-reviewed publications using the MedScan[5] natural language processor (NLP) engine, and archived in Elsevier's Knowledge Graph (Elsevier, Amsterdam) database.[6] These are being used within the Pathway Studio (Elsevier, Amsterdam) to construct executable computer models of the molecular, cellular and biobehavioral signaling that underlies a broad range of illness. In recent work, we recovered 188 documented elements of bio-behavioral feedback and feedforward regulation linking 44 neurotransmitters, immune mediators, and behavioral constructs into a network model capturing the potential mechanistic underpinnings and psychoneuroimmunology (PNI) supporting chronic pain, persistent depression and addictive behavior (see **Figure 1**). These were supported by a total of 21,533 peer-reviewed articles with at least one and up to 1,168 citations supporting each interaction.

Intuitively we expected that the bulk of prior knowledge would describe interactions between markers expressed within the same level of biology, i.e. interactions between two physiological markers or two behavioral constructs. However, results of our text-mining did indeed indicate that over half of the 188 documented interactions (128 in 16,574 citations) connected two physiological markers, only a small minority of these were reported at the behavioral level (6 in 9 citations) (**Table 1**). Interestingly, in what was thought to be the least well-documented layer of biology we recovered 35 interactions where a physiological marker regulated a behavioral construct (3,688 citations), and conversely 19 interactions where a behavioral construct regulated a physiological marker (1,262 citations). Moreover, the strength of the underlying statements was such that a majority of these relationships (137 of 188) were classified as regulatory interactions (119 in 10,739 citations), with roughly 1 in 6 satisfying the more stringent conditions of direct regulation (18 in 1,510 citations). Indeed, fewer than 1 of 3 (51 of 188) interactions were considered more associative in nature (36 co-expressions in 9,150 citations, 15 weaker quantitative changes in 134 citations).

Examples such as this illustrate the growing capability of automated text-mining for reliably extracting community knowledge of regulatory biology from peer-reviewed literature with a thoroughness and at a scale previously unimaginable. Furthermore, cursory results presented here suggest that while network analytical approaches may not yet be widely applied in behavioral studies, PNI is making strong inroads as evidenced by a growing body of literature reporting interactions between not just brain but also mind and body. We expect that effectively harnessing such community-wide information in a systematic and computationally accessible manner will offer clinicians the ability to more efficiently integrate best practice knowledge into clinical decision-

making. Likewise, we expect these technologies to accelerate the rate of knowledge generation itself by highlighting gaps in our current understanding and hence informing on more cost-effective and informative study designs, as well as informing novel possible treatments for complex presentations.

Conflict of Interest

None declared.

Author Contributions

Conceived and designed the manuscript: JG, GB. Contributed to the analysis: JG, JTT, GB. Drafted the manuscript: JG. Supervised manuscript writing: GB. Provided supervision and expertise on chronic pain and clinical applications: HK, GS. Provided supervision and expertise on computational modelling techniques: GB.

Funding

This work was supported by Rochester Regional Health in conjunction with Elsevier BV (Amsterdam) under a collaborative research sponsorship (Broderick, PI) and the US Department of Defense Congressionally Directed Medical Research Programs (CDMRP) http://cdmrp.army.mil/ award W81XWH-15-1-0582 CDMRP (Broderick PI/ Whitley Partnering PI) and under Peer Reviewed Medical Research Program (PRMRP) award W81XWH-19-1-0804 (Broderick - PI; Sethi - Partnering PI).

Mandatory Disclaimer

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of Defense, Rochester Regional Health, or Elsevier BV.

Acknowledgments

The authors would also like to thank Drs. Chris Cheadle, Hongbao Cao, Dicle Hasdemir, Anton Yuriev, Philipp Anokhin, Umesh Nandal and Thibault Geoui of Elsevier BV for their technical assistance and many helpful discussions.

Data Availability Statement

Any and all data used is publicly available.

References

1. Chou, L., Chang, K., & Puspitasari, I. (2020). Drug abuse research trend investigation with text mining. *Computational and Mathematic Methods in Medicine*. https://doi.org/10.1155/2020/1030815

2. Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text mining in big data analytics. *Big Data and Cognitive Computing, 4*(1). Doi: 10.3390/bdcc4010001

3. Sedghamiz, H., Morris, M., Craddock, T. J., Whitley, D., & Broderick, G. (2019). Bio-ModelChecker: Using bounded constraint satisfaction to seamlessly integrate observed behavior with prior knowledge of biological networks. *Frontiers in Bioengineering and Biotechnology, 7*, 1-13. doi: 10.3389/fbioe.2019.00048

4. Toole, J. T., Rice, M. A., Craddock, T. J., Nierenberg, B., Klimas, N. G., Fletcher, M. A., Zysman, J., Morris, M., & Broderick, G. (2018). Breaking away: The role of homeostatic drive in perpetuating depression. *Methods in Molecular Biology, 1781*, 121-144. doi: 10.1007/978-1-4939-7828-1_8

5. Novichkova, S., Egorov, S., and Daraselia, N. (2003). MedScan: A natural language processing engine for MEDLINE abstracts. *Bioinformatics, 19*(3), 1699-1706.

6. Kamdar, M. R., Stanley, C. E., Carroll, M., Wogulis, L., Dowling, W., Deus, H. F., & Samarasinghe M. (2020). Text snippets to corroborate medical relations: An unsupervised approach using a knowledge graph and embeddings. *AMIA Joint Summits on Translational Science, 2020*, 288-297. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7233036/

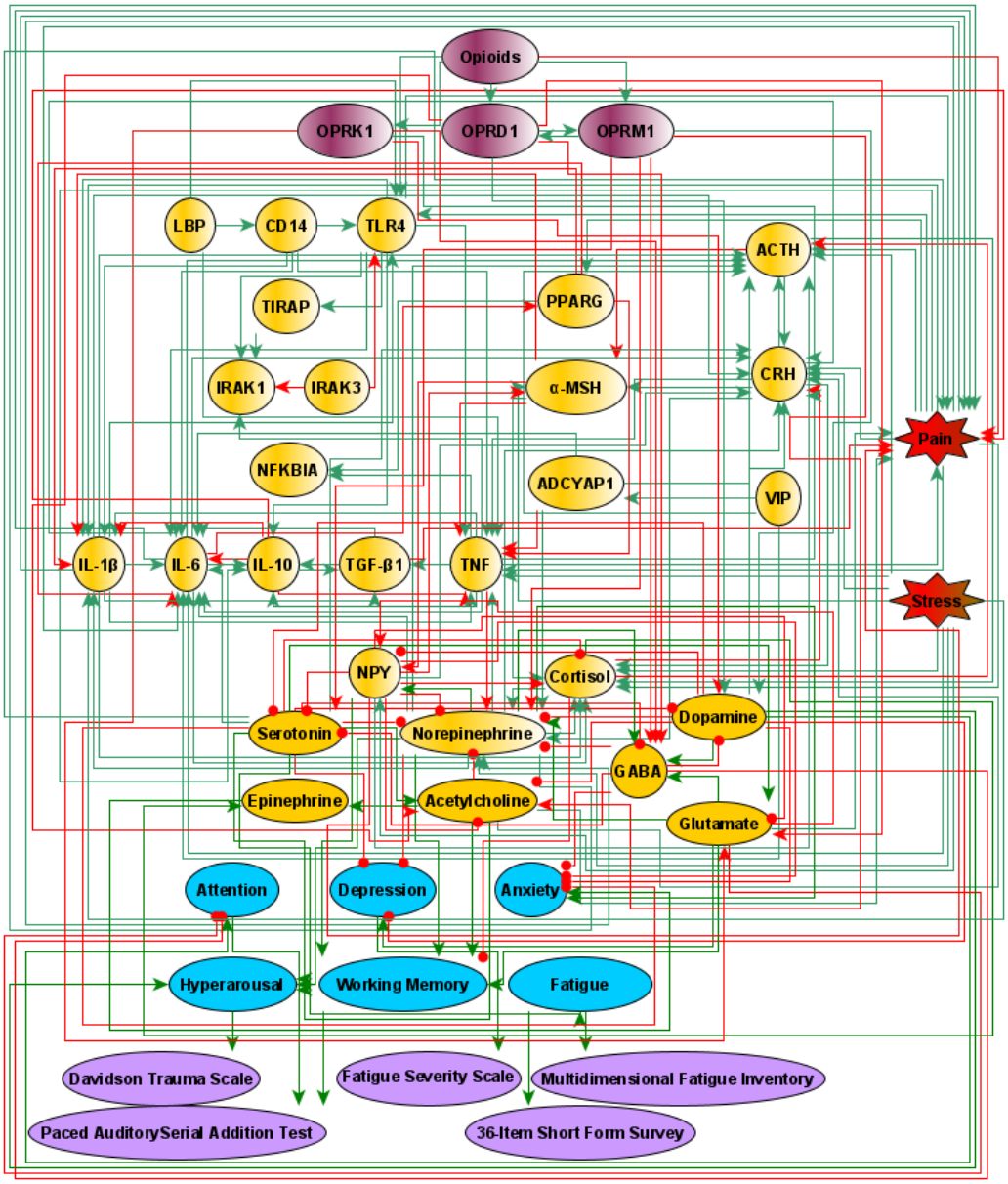

Figure 1. Circuit diagram depicting the resultant text-mined feedback and feedforward interactions between specified neurotransmitters, immune mediators, and behavioral constructs. Colored arrows indicate directed regulatory interactions, with green arrows indicating that the source mediator upregulates the target mediator, and red arrows indicating the source mediator downregulates the target mediator.

Interactions	Type	Frequency	Number of Citations
Biological to Biological	Regulation	79	6,422
	Direct Regulation	16	1,378
	Expression	33	8,774
	Quantitative Change	0	0
Biological to Behavioral	Regulation	35	3,688
	Direct Regulation	0	0
	Expression	0	0
	Quantitative Change	0	0
Behavioral to Biological	Regulation	5	629
	Direct Regulation	2	132
	Expression	3	376
	Quantitative Change	9	125
Behavioral to Behavioral	Regulation	0	0
	Direct Regulation	0	0
	Expression	0	0
	Quantitative Change	6	9

Table 2. Types of interactions with associated frequency and number of supporting citations.