Gene 830 (2022) 146534

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Editorial

A retrospective evaluation of a decade of Gene Wiki Reviews and their impact

1. Background

The Gene Wiki Reviews are a series of review articles published in a partnership between GENE and the Gene Wiki Initiative. Launched in 2013 (Su et al., 2013), the series initially aimed to improve the accessibility of gene-centric information by incentivizing experts to share their knowledge in Wikipedia as one of the world's most widely-consulted, openly-accessible encyclopedias. The participating authors were invited to submit two gene-specific review articles. The first submission is a peer-reviewed "article of record" written for researchers and published in Gene. The second a "living article" written for lay audiences and published in Wikipedia that would continue to evolve. In this manner, the series would contribute to closing the knowledge gap between the human genome project and publicly available reliable information about human genes. Within three years of its launch, the Gene Wiki Reviews explored a thematic focus on cardiovascular diseases, developed ways to link the articles and Wikipedia entries published via the series, and surveyed authors to understand the impact of the series on its contributors (Tsueng et al., 2016). *RNA Biology* and *PLOS Computational Biology* ran similar initiatives. The Gene Wiki Reviews project was endorsed with limited funding from Elsevier (publisher of Gene) until 2021 ('8 year benchmark') (Fig. 1), and the project will evolve further as an academic volunteer effort with continued involvement of the authors. After nearly a decade of progress, we provide a retrospective overview of what the series has accomplished.

2. Impact of the series

Since the start of the series, there have been 96 review articles published along with 125 corresponding gene-specific Wikipedia articles updated. Collectively, the Gene Wiki Reviews have been cited 6412 times, viewed over 165,414 times, and accumulated 95,685 downloads. The series' most cited review that is available from PubMed Central is "*Epithelial sodium channel (ENaC) family: Phylogeny, structure–function, tissue distribution, and associated inherited diseases*" (Hanukoglu and Hanukoglu, 2016). The most viewed review available in PubMed Central is "*FBN1: The disease-causing gene for Marfan syndrome and other genetic disorders*" (Sakai et al., 2016). A virtual special issue listing is available here: https://www.sciencedirect.com/journal/gene/special-issue/10VRTD18BMK.

Collectively, the Wikipedia articles updated in the series have been viewed more than 4.9 million times since the start of the series, with the collective size of the articles growing from about 1 million characters to nearly 4 million characters. Average monthly page views for Wikipedia articles updated via the series have changed from 370 in the six months prior to an update, to 698 in the six months after an update. Wikipedia articles in this series grew by a median of about 20,000 characters. The greatest increase in article length was 118,000 characters. A list of Wikipedia articles that were improved by the series can be found at https://en.wikipedia.org/wiki/Category:Wikipedia_articles_with_corresponding_academic_peer_reviewed_articles (Fig. 2).

3. Rare disease focus

The Gene Wiki Reviews were distinct and timely in reflecting the shifting landscape of informational needs, specifically as rare disease communities helped drive the political discourse on precision medicine in 2015 (Saltonstall, 2015; House of Representatives of the 114th Congress, 2015; Handelsman, 2015). In the United States, a rare disease afflicts less than 200,000 people. Roughly 10% or 30 million people in the U.S. are estimated to be afflicted with one or more of the estimated 7000 rare diseases. Although the human genome project was declared complete nearly a decade before the start of the Gene Wiki Review series, its impact on the rare disease community was only beginning as the price of genomic sequencing dropped. The urgency for gene-specific information to be made widely and publicly available increased as genome sequencing began providing diagnoses for previously undiagnosed patients.

The early exploration of a thematic (Cardiac Disease) Gene Wiki Review series laid the groundwork to focus on rare diseases as rare disease groups raised public support for various precision medicine and orphan drug initiatives in 2016. Since 2016, the Gene Wiki Reviews have thus intentionally sought authors for rare disease genes for which the corresponding Wikipedia articles were rare or non-existent. Reproducibility concerns and orphan drug initiatives to find treatments for rare diseases have steered conversations from open access for research papers to open access for data. Data management mandates from funders and open data mandates from journals have made it clear that this level of transparency will be needed for science moving forward. In anticipation of this shift, and following the evolution of the Gene Wiki Initiative from Wikipedia to Wikidata, the Gene Wiki Reviews have included a Wikidata component since 2018.

The Gene Wiki Reviews were always open for submission for disease-agnostic genes as long as they met a few basic criteria: 1. The corresponding Wikipedia article for that gene was either a stub article, woefully inadequate, or missing altogether; 2. There was sufficient literature to be reviewed (at least 30–70 primary research articles about the gene). This literature availability criterion proved challenging, especially for rare diseases associated with relatively unknown genes. Rare disease genes were prioritized based on the number of families

https://doi.org/10.1016/j.gene.2022.146534

Available online 4 May 2022

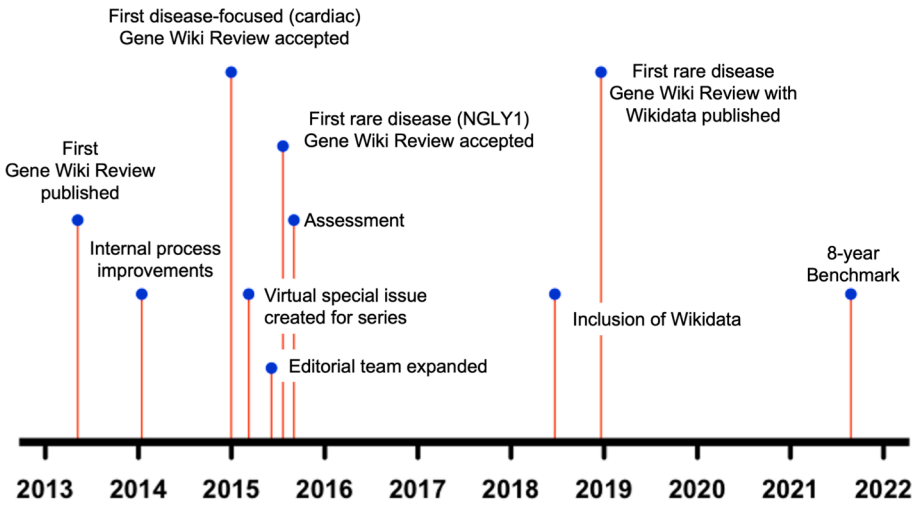

Fig. 1. Timeline of major events in the Gene Wiki project.

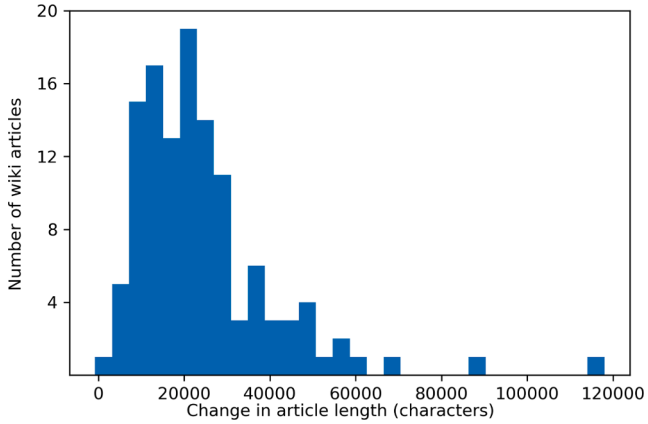

Fig. 2. The growth of the wiki articles.

affected, and the availability of literature and experts on those genes (University of Washington, 2021). Out of 96 Gene Wiki Reviews published, 19 were rare disease genes resulting in the expansion of 19 rare disease gene-specific articles in Wikipedia. Collectively, these Wikipedia articles have been viewed more than 215,000 times since 2016.

4. Contributions to open data

As members of rare disease communities shared their struggles to understand a diagnosis, they advocated and fostered much-needed conversations about the ethics of accessibility in publicly funded research and highlighted gaps in existing open access models. Open-access discussions frequently centered around tax-payers as stakeholders resulting in mandates and compliance (Butler, 2010; Jones et al., 2013; Schekman et al., 2012), re-use (Neylon, 2012; Schekman et al., 2012), features and limitations of publication models (Rankin and Franklin, 2004; Schekman et al., 2012), and the academic publishing industry (Oren, 2008; Schekman et al., 2012). The impact of open access remained unclear as information on the interest and use of research articles by the general public was limited (Davis and Walters, 2011,). New research paradigms driven by rare disease patients, families, and advocates cleared the doubt that the public would use openly accessible research articles (Wheeler, 2014; Might and Wilsey, 2014; Snow, 2016).

When the open-access movement shifted from ensuring accessibility of scholarly articles to encouraging data to be findable, accessible, interoperable, and reusable (FAIR), so too did the Gene Wiki Initiative's focus from improving Wikipedia to improving Wikidata. Wikidata is a freely accessible, machine and human-readable, open knowledge base. Multiple research groups have worked to import biomedical data into Wikidata, enabling it to serve as a biomedically relevant knowledge graph (Waagmeester et al., 2020). The Gene Wiki Reviews began exploring Wikidata improvements as part of the series in 2018. Out of 96 Gene Wiki Reviews published, 15 reviews included data submissions for improving Wikidata resulting in the addition of 49 new entities and 250 statements to Wikidata. Relative to the general statements imported to Wikidata from other databases, the curated content added via the Gene Wiki Reviews is smaller in number but much more specific. For example, 13,000 gene-disease associations were imported into Wikidata from Uniprot via bots by the Su research group, but these associations do not always specify how a gene relates to a disease, or the level of evidence that supports this statement. In contrast, statements curated by the series include the type of genetic association, methods of analysis and further specific information.

5. Conclusions

Ten years after we started the Gene Wiki project, we can conclude that our pioneering effort has succeeded in its goal of involving scientists to author and edit Wikipedia gene records backed up by scholarly articles on common and rare diseases. The Gene Wiki project has built a legacy of ~100 articles and contributed to Wikidata. We anticipate that the project will continue to grow new branches with greater scholarly community involvement to advance high-quality, openly accessible information about genes, diseases, and biomedical sciences in general.

References

Butler, D., 2010. US seeks to make science free for all. Nature 464 (7290), 822–823. https://doi.org/10.1038/464822a. PMID: 20376119.

Davis, P.M., Walters, W.H., 2011. The impact of free access to the scientific literature: a review of recent research. J. Med. Libr. Assoc. 99 (3), 208–217. https://doi.org/10.3163/1536-5050.99.3.008. PMID: 21753913; PMCID: PMC3133904.

Handelsman, J., 2015. Precision Medicine: Improving Health and Treating Disease. [online] whitehouse.gov. 2015 Available at: https://obamawhitehouse.archives.gov/blog/2015/01/21/precision-medicine-improving-health-and-treating-disease [Accessed 2 December 2021].

Hanukoglu, I., Hanukoglu, A., 2016. Epithelial sodium channel (ENaC) family: Phylogeny, structure-function, tissue distribution, and associated inherited diseases. Gene 592, 235–238. https://doi.org/10.1111/febs.13840. PMID: 27580245.

House of Representatives of the 114th Congress, 2015. H. Rept. 114-190 - 21ST CENTURY CURES ACT Committee Report. [online] Congress.gov. 2015 Available at: https://www.congress.gov/congressional-report/114th-congress/house-report/190/1 [Accessed 2 December 2021].

Jones, R., Dambha, H., Hull, C., 2013. Open access publishing: important changes for the BJGP. Br. J. Gen. Pract. 63 (609), 181.

Might, M., Wilsey, M., 2014. The shifting model in clinical diagnostics: how next-generation sequencing and families are altering the way rare diseases are discovered, studied, and treated. Genet. Med. 16 (10), 736–737. https://doi.org/10.1038/gim.2014.23. Epub 2014 Mar 20. PMID: 24651604.

Neylon, C., 2012. Science publishing: Open access must enable open use. Nature 492 (7429), 348–349. https://doi.org/10.1038/492348a. PMID: 23257864.

Oren, G.A., 2008. The Crisis in Scholarly Publishing: Open Access to the Rescue? J. Neuro-Ophthalmol. 28 (1), 1–4. https://doi.org/10.1097/WNO.0b013e3181678618.

Rankin, J.A., Franklin, S.G., 2004. Open access publishing. Emerg. Infect. Dis. 10 (7), 1352. https://doi.org/10.3201/eid1007.040122. PMID: 15338570; PMCID: PMC3323340.

Sakai, L.Y., Keene, D.R., Renard, M., De Backer, J., 2016. FBN1: The disease-causing gene for Marfan syndrome and other genetic disorders. Gene 591 (1), 279–291.

Saltonstall, P., 2015. Comments on the 21st Century Cures Act. [online] National Organization for Rare Disorders. 2015 Available at: https://rarediseases.org/wp-content/uploads/2015/06/NORD-Comments-on-the-21st-Century-Cures-Act.pdf [Accessed 2 December 2021].

Schekman, R., Watt, F., Weigel, D., 2012. Launching eLife, Part 2. Elife 13 (1), e00365. https://doi.org/10.7554/eLife.00365. PMID: 23240092; PMCID: PMC3510471.

Snow, K., 2016. Families Isolated By Rare Genetic Conditions Find New Ways To Reach Out. [online] NPR. 2016 Available at: https://www.npr.org/sections/health-shots/2016/06/05/480373533/families-isolated-by-rare-genetic-conditions-find-new-ways-to-reach-out [Accessed 2 December 2021].

Su, A.I., Good, B.M., van Wijnen, A.J., 2013. Gene Wiki Reviews: Marrying crowdsourcing with traditional peer review. Gene 531 (2), 125. https://doi.org/10.1016/j.gene.2013.08.093.

Tsueng, G., Good, B.M., Ping, P., Golemis, E., Hanukoglu, I., van Wijnen, A.J., Su, A.I., 2016. Gene Wiki Reviews—Raising the quality and accessibility of information about the human genome. Gene 592 (2), 235–238.

University of Washington, 2021. Candidate genes in MyGene2. [online] Mygene2.org. 2021 Available at: https://mygene2.org/MyGene2/genes [Accessed 2 December 2021].

Waagmeester, A., Stupp, G., Burgstaller-Muehlbacher, S., Good, B.M., Griffith, M., Griffith, O.L., Hanspers, K., Hermjakob, H., Hudson, T.S., Hybiske, K., Keating, S.M., Manske, M., Mayers, M., Mietchen, D., Mitraka, E., Pico, A.R., Putman, T., Riutta, A., Queralt-Rosinach, N., Schriml, L.M., Shafee, T., Slenter, D., Stephan, R., Thornton, K., Tsueng, G., Tu, R., Ul-Hasan, S., Willighagen, E., Wu, C., Su, A.I., 2020. Wikidata as a knowledge graph for the life sciences. Elife 17 (9), e52614. https://doi.org/10.7554/eLife.52614. PMID: 32180547; PMCID: PMC7077981.

Wheeler, D., 2014. When the Diagnosis Is Rare, Parents May Know More Than Professionals. [online] Motherlode Blog. 2014 Available at: https://parenting.blogs.nytimes.com/2014/10/28/when-the-diagnosis-is-rare-parents-may-know-more-than-professionals/ [Accessed 2 December 2021].

Andre J. van Wijnen*
Department of Biochemistry, University of Vermont, Burlington, VT, USA

Erica Golemis
Fox Chase Cancer Center, Philadelphia, PA, USA
Lewis Katz School of Medicine, Temple University, Philadelphia, PA, USA

Israel Hanukoglu
Ariel University, Ariel, Israel

Stephen Kwok-Wing Tsui
The Chinese University of Hong Kong, Hong Kong

Eric Hu, Sabah Ul-Hasan, Janet Joy, Andrew I. Su, Ginger Tsueng
The Scripps Research Institute, La Jolla, CA, USA

* Corresponding author.
E-mail address: vanwijnen.andre@mayo.edu (A.J. van Wijnen).