

Trends in Genetics

Forum

Every gene can (and possibly will) be associated with cancer

João Pedro de Magalhães D^{1,*}

A PubMed analysis shows that the vast majority of human genes have been studied in the context of cancer. As such, the study of nearly any human gene can be justified based on existing literature by its potential relevance to cancer. Moreover, these results have implications for analyzing and interpreting largescale analyses. topic in biological and biomedical sciences. However, the huge amount of data gathered concerning cancer means that there is much more information concerning genes associated with cancer than for any other disease or process. Human biases in the way genes are studied can confound systematic analyses, such as network analyses [4,5]. Here, I explore the extent of such biases in cancer associations and their implications.

An analysis of cancer-related publications

Of the 17 371 human genes with at least one paper in PubMed, 15 233 (87.7%) also have at least one paper mentioning cancer (see [6] for methods). Only three (*SLC26A5*, *PRPH2*, and *CRYZ*) out of 4186 genes with over 100 publications do not have a publication mentioning cancer.

Interestingly, these three genes play specific roles in the eye and hair cells, tissues that are not common sources of cancer. Likewise, genes with at least one paper in PubMed but no paper mentioning cancer (n = 2138 genes, out of 17 371 genes in)total) are enriched for olfactory receptors and antimicrobial defenses (see the supplemental information online); these are processes that, even though some olfactory receptors can be differentially expressed in cancer [7], are not commonly associated with cancer. An incredible 24.4% of all publications associated with genes in PubMed mention cancer (Figure 2). While co-occurrence of a gene and cancer in a publication is not by itself evidence of a causal association, these numbers notwithstanding illustrate the huge amount of studies relating the vast majority of human genes to cancer.

Cancer as the most studied biomedical topic

Cancer is one of the most common diseases of modern times. In industrialized countries, cancer affects roughly one in two people at some point during their lives [1] and cancer incidence and mortality is expected to continue increasing given the ageing populations worldwide [2]. Not surprisingly, cancer attracts a huge amount of research funding from government, private, and philanthropic sources [3]. At the time of writing, over 4 million of the over 30 million publications in PubMed mention cancer. For comparison, roughly 350 000 publications mention stroke. As of 2020, over 200 000 papers are published each year mentioning cancer (Figure 1). Compared with other common diseases, like heart or neurodegenerative diseases, cancer is also seemingly more straightforward to study, given the wide availability of materials, like cell lines. In other words, the experimental methods necessary to study cancer seem to have lower technical limitations compared with many other disease scenarios. As such, cancer is the most widely studied



Figure 1. Publications per year mentioning cancer, according to PubMed (https://pubmed.ncbi. nlm.nih.gov/). Stroke is shown for comparison purposes.

Trends in Genetics





Percentage of cancer publications per gene

Figure 2. Percentage of publications per gene mentioning cancer, according to PubMed (https://pubmed.ncbi.nlm.nih.gov/). Genes with ten or more publications ($n = 11\,866$) are ranked from lowest to highest percentage of publications mentioning cancer.

Implications for interpreting large-scale studies

In light of these results, clearly the vast majority of genes that have been studied have also been studied in the context of cancer. The study of nearly any human gene can be justified (e.g., in grant applications) based on existing literature by its potential relevance to cancer. When interpreting results from genome-wide studies and high-throughput approaches, that typically yield large gene lists, it is very likely to find associations with cancer (>99% chance for results with three or more genes). Cancer publication biases can also impact on network analyses, such as protein–protein interactions, that are influenced by the

number of studies of each protein [6]. Understanding the reasons for biases in large-scale analyses and correcting for them is of growing importance to increase the value of insights and predictions [4,5,8].

In conclusion, researchers should be aware of the strong biases towards the study of genes in the context of cancer when discussing their results and interpreting the works of others. In genetics and genomics, literally everything is associated with cancer. If a gene has not been associated with cancer yet, it probably means it has not been studied enough and will most likely be associated with cancer in the future. In a scientific world where everything and every gene can be associated with cancer, the challenge is determining which are the key drivers of cancer and more promising therapeutic targets.

Acknowledgments

Work in my lab is supported by grants from the Wellcome Trust (208375/Z/17/Z), LongeCity and the Biotechnology and Biological Sciences Research Council (BB/R014949/1). I am also grateful to Ines Lopes for assistance with retrieving data from PubMed and making the figures.

Declaration of interests

The authors have no interests to declare.

Supplemental information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.tig. 2021.09.005.

 1 Integrative Genomics of Ageing Group, Institute of Ageing and Chronic Disease, University of Liverpool, Liverpool, UK

*Correspondence:

Trends in Genetics

jp@senescence.info (J.P. de Magalhães). https://doi.org/10.1016/j.tig.2021.09.005

© 2021 Published by Elsevier Ltd.

References

- Ahmad, A.S. et al. (2015) Trends in the lifetime risk of developing cancer in Great Britain: comparison of risk for those born from 1930 to 1960. Br. J. Cancer 112, 943–947
- de Magalhaes, J.P. (2013) How ageing processes influence cancer. Nat. Rev. Cancer 13, 357–365
- Schmutz, A. et al. (2019) Mapping the global cancer research funding landscape. JNCI Cancer Spectr. 3, pkz069
- Fernandes, M. et al. (2016) Systematic analysis of the gerontome reveals links between aging and age-related diseases. *Hum. Mol. Genet.* 25, 4804–4818
- Schaefer, M.H. et al. (2015) Correcting for the study bias associated with protein-protein interaction measurements reveals differences between protein degree distributions from different cancer types. *Front. Genet.* 6, 260
- Lopes, I. et al. (2021) Gene size matters: an analysis of gene length in the human genome. Front. Genet. 12, 559998
- Weber, L. *et al.* (2018) Olfactory receptors as biomarkers in human breast carcinoma tissues. *Front. Oncol.* 8, 33
- Stoeger, T. *et al.* (2018) Large-scale investigation of the reasons why potentially important genes are ignored. *PLoS Biol.* 16, e2006643