Beyond Stamp Collecting: Evolutionary and Functional Genomics Advance Our Understanding of Cancer Biology

Joseph Lachance

In this issue of *Cancer Research*, Emami and colleagues leveraged genetic data from over 200,000 men of European descent to implicate rare alleles that are associated with prostate cancer. However, this study went beyond a simple description of statistical associations between genetic variants and cancer risk. Polygenic risk scores were applied to large cohorts from Kaiser Permanente and the UK Biobank, demonstrating the clinical utility of genetic predictors of disease risk. Furthermore, by placing their results in an evolutionary framework and integrating genetic information with functional data, the authors of this major study were able to bridge the gap between genome-wide association studies and the biological mechanisms underlying prostate cancer risk.

See related article by Emami et al., 1695

Nuclear physicist Ernest Rutherford famously claimed that "all science is either physics or stamp collecting." Although there is some debate among historians regarding the exact wording and meaning of this quote (1), it can be viewed as a criticism of descriptive science. Rutherford's claim may have been a bit harsh, but it does bring to attention the limitations of assembling long lists of biological phenomena, such as the top hits from genome-wide association studies (GWAS). Because of low statistical power, early GWAS were only able to identify common variants that were associated with complex diseases such as prostate cancer. However, as genotyping technologies have become more affordable and sample sizes have increased, the ability to identify rare cancer-associated alleles has greatly improved. This has resulted in an ever-growing catalog of disease associations (2), and it has become increasingly clear that prostate cancer is a highly polygenic disease. On the one hand, GWAS can be viewed as a series of successes (3), but these studies also present a challenge: how can we move beyond "mere" lists of disease-associated loci toward a greater understanding of cancer biology and improved clinical outcomes?

In this issue, Emami and colleagues analyzed over 14,000 prostate cancer cases and 194,000 controls of European descent from the Kaiser Permanente and UK Biobank datasets (4). They integrated genetic data from a custom microarray with large imputation panels to identify germline variants that are associated with increased risk of prostate cancer. This GWAS led to the discovery of three novel disease associations (rs557046152 at 8p12, rs555778703 at 4q31.21, and rs62262671 at 3p21.31). Two of the novel associations discovered here involved rare alleles with large effect sizes. The authors also applied a polygenic risk score built from 188 previously reported prostate cancer variants to the Kaiser Permanente and UK Biobank datasets. This genetic classifier was able to successfully distinguish between high- and low-risk individuals. Men whose polygenic risk scores placed them into the top 1% bin had ORs of 4.12 compared with men in the 40th to 60th percentiles. This finding suggests that polygenic risk scores for prostate cancer can be used to identify individuals who would benefit the most from early screening. Importantly, Emami and colleagues also examined the evolutionary history and functional characteristics of prostate cancer–associated variants.

Gene-based association tests can reveal aspects of cancer biology that are missed by traditional GWAS. Because of low statistical power, it is difficult to detect individual disease associations when alleles are rare. To surmount this statistical challenge, Emami and colleagues identified genes that contain multiple associations with prostate cancer (4). The *ILDR1* gene yielded a suggestive hit in the Kaiser Permanente cohort. *ILDR1* is expressed in the prostate, and it encodes a protein that contains an immunoglobin-like domain. Gene-based tests also revealed that *HOXB13* contains multiple rare variants that are associated with prostate cancer in the UK Biobank. This gene encodes a homeobox transcription factor that acts as a tumor suppressor, and a rare missense mutation in *HOXB13* (G84E, rs138213197) has previously been associated with prostate cancer (5). Intriguingly, rs138213197 was found to be in long-range linkage disequilibrium with other rare variants.

One major limitation of rare-variant association tests is that they are not robust to realistic evolutionary scenarios (6). Furthermore, purifying natural selection can cause disease-associated alleles to be found at low frequencies. To account for this, Emami and colleagues examined the evolutionary history of cancer-associated variants. Extended haplotype homozygosity, a hallmark of recent natural selection, was found near multiple prostate cancer–associated loci. Specifically, rs555778703 and rs138213197 had large integrated haplotype score (iHS) statistics (4). The prostate cancer–associated haplotype at *HOXB13* that includes rs138213197 has a recent evolutionary origin and it is largely restricted to Northern Europe. Adopting an evolutionary perspective can explain why some cancer-associated variants have a limited geographic range. Similarly, population genetics approaches can also shed light on why incidence and mortality rates vary across the globe. For example, neutral evolutionary processes and natural selection have contributed to elevated risks of prostate cancer in men of African descent (7).

How do germline variants actually contribute to differences in prostate cancer risk? To answer this question, Emami and colleagues integrated genetic information with gene expression data from healthy prostate tissue. This functional genomics approach combined expression quantitative trait loci (eQTL) findings with GWAS results. Eighty of the 188 previously reported and three novel prostate cancer variants analyzed in this study were in high linkage disequilibrium with a known prostate eQTL (4). The remaining disease-associated variants may modulate prostate cancer risk via other tissues. Alternatively, they

School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia.

Corresponding Author: Joseph Lachance, Georgia Institute of Technology, 950 Atlantic Dr., Atlanta, GA 30332. Phone: 404-894-0794; E-mail: joseph.lachance@biology.gatech.edu

Cancer Res 2021;81:1637–8

doi: 10.1158/0008-5472.CAN-21-0146

may have effects on gene expression that are only revealed in unhealthy prostate tissue. In addition, 32 prostate cancer variants were predicted to significantly alter transcription factor–binding sites. These findings highlight the importance of regulatory DNA to cancer biology. Of particular note, rs2680708 at 17q22 has a large predicted effect on binding of the androgen receptor (AR) transcription factor to chromatin. Other prostate cancer variants that interact with androgen-related transcription factors include rs2660753, rs742134, rs9625483, and rs62262671. Taken together, these results imply that the androgens play a key role in the etiology of prostate cancer.

As seen in this issue of *Cancer Research*, genome-wide association studies can be considered an important first step toward understanding the biological mechanisms behind prostate cancer risk. By examining the evolutionary history of disease-associated variants it is possible to infer why specific genetic variants are common or rare. An evolutionary perspective also enables driver mutations to be distinguished from passenger mutations, yielding a deeper understanding of tumor biology (8). In addition, evolutionary thinking can lead to novel therapeutic approaches, such as adaptive therapy (9). Integrating genetic information with gene expression, methylation, or other functional data helps answer how questions. Going forward, integrative 'omics approaches will continue to enhance studies of cancer biology (10).

Authors' Disclosures

No disclosures were reported.

Acknowledgments

This work was supported by an NIH grant (R35GM133727 to J. Lachance).

Received January 14, 2021; accepted January 14, 2021; published first April 1, 2021.

References

1. Fletcher T, Brownstein EM. Physics or stamp collecting? Pitfalls of the hierarchy of disciplines. Science Scope 2009;33:6.
2. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res 2017;45:D896–901.
3. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet 2017;101:5–22.
4. Emami NC, Cavazos TB, Rashkin SR, Graff RE, Tai CG, Mefford JA, et al. A large-scale association study detects novel rare variants, risk genes, functional elements, and polygenic architecture of prostate cancer susceptibility. Cancer Res 2021;81:1695–703.
5. Ewing CM, Ray AM, Lange EM, Zuhlke KA, Robbins CM, Tembe WD, et al. Germline mutations in HOXB13 and prostate-cancer risk. N Engl J Med 2012;366:141–9.
6. Uricchio LH, Zaitlen NA, Ye CJ, Witte JS, Hernandez RD. Selection and explosive growth alter genetic architecture and hamper the detection of causal rare variants. Genome Res 2016;26:863–73.
7. Lachance J, Berens AJ, Hansen MEB, Teng AK, Tishkoff SA, Rebbeck TR. Genetic hitchhiking and population bottlenecks contribute to prostate cancer disparities in men of African descent. Cancer Res 2018;78:2432–43.
8. Gomez K, Miura S, Huuki LA, Spell BS, Townsend JP, Kumar S. Somatic evolutionary timings of driver mutations. BMC Cancer 2018;18:85.
9. Gatenby RA, Silva AS, Gillies RJ, Frieden BR. Adaptive therapy. Cancer Res 2009;69:4894–903.
10. Karczewski KJ, Snyder MP. Integrative omics for health and disease. Nat Rev Genet 2018;19:299–310.