



Published in final edited form as:

*Cancer Discov.* 2017 April ; 7(4): 354–355. doi:10.1158/2159-8290.CD-17-0192.

## Exploring the link between the germline and somatic genome in cancer

Paul Geeleher<sup>1</sup> and R. Stephanie Huang<sup>1,\*</sup>

<sup>1</sup>Section of Hematology/Oncology, Department of Medicine, The University of Chicago, USA

### Main Text

Most cancer genomics research is focused on somatic events, such as acquired mutations; but increasing evidence suggests that inherited germline genetic variation also plays a key role in cancer risk (1,2) pharmacogenomics (3,4) and gene regulation (5,6). We have known for decades that cancer is heritable and studies identifying specific germline genetic variants that predispose individuals to cancer span back beyond the 1990s—for example the discovery of the highly penetrant BRCA1 gene (7). The sequencing of the human genome, the development of affordable genotyping technologies and the subsequent use of genome-wide association studies (GWAS) further increased this list of known cancer risk variants. Larger sample sizes have allowed us to discover germline variants with an ever smaller effect on disease risk. For example the largest meta-analysis of GWAS data for breast cancer risk, which compared germline variation between 62,533 breast cancer patients and 60,976 controls, identified a total of 84 germline loci that explained about 16% of familial risk (8). Cheaper genome sequencing technologies are also allowing us to discover rare cancer risk variants. Given germline cancer risk variants have now been extensively catalogued, it is surprising that very little is known about the molecular function of most of these variants and in particular how they affect disease development. Indeed, studies that systematically evaluate the entire somatic and germline cancer genome and establish a link between the two are completely lacking.

In the current study, Carter *et al.* (9) take a first step in bridging this gap and propose studying the effect of the germline genome on cancer in a novel way, which aims at assigning function to these variants. Rather than comparing the germline genome between cohorts of cases (cancer patients) and controls, as in GWAS, they are comparing the germline genome *within* a large group of cancer patients. This allows them to ask two questions that have not previously been addressed: (a) How does germline genetic variation affect the propensity of cancer to occur in one tissue, rather than another? (b) Does germline genetic variation affect the somatic mutation profiles in cancers?

\*Corresponding Author: R. Stephanie Huang, 900 E 57th Street, KCBD room 7148, Chicago, IL 60637, USA. rhuang@medicine.bsd.uchicago.edu, Phone: 773-702-9363.

#### Conflict of interest

The authors disclose no potential conflicts of interest.

To address these questions, the authors use data from The Cancer Genome Atlas (TCGA), a very large study that has collected genomics information on over 10,000 cancer patients using multiple molecular profiling technologies. The germline genome in these patients has been measured using genotyping microarrays on samples collected from blood, which capture most common genetic variation (occurring at a minor allele frequency of >1%).

The first of the authors' questions "How does germline genetic variation affect the propensity of cancer to occur in one tissue, rather than another?" is addressed by using an approach similar to a conventional GWAS. However, instead of comparing cancer patients to healthy controls, the authors compare genotypes in each of the 22 cancer types included in TCGA to the patients from the other 21 types of cancer pooled. The authors' analysis identified relevant signal. Indeed of 916 markers that were prioritized in their discovery cohort, 395 were replicated in their validation cohort at an FDR of 0.25, a larger proportion than would have been expected by chance. The authors go on to demonstrate that the analysis identified several known risk variants and additionally identified specific examples that represent promising novel candidate genes for follow up. Given the encouraging results, in future, it may be possible to supplement such an analysis with genotype data collected from other cancer sequencing studies or even from cancer risk GWAS.

The second of the authors' questions ("How does germline genetic variation affect the somatic mutation profile in individuals who have cancer?") is also addressed using TCGA. TCGA carried out both germline genotyping and tumor exome sequencing. Thus, using these data it has been possible to compare the frequency of specific somatic mutations, given a specific germline genotype. The analysis focused on the association between germline genotype and the frequency of somatic mutations in 138 well established cancer genes. The sample size is small given the large number of statistical tests, but the authors argue that for some of the associations studied, the effect sizes recovered are much larger than for a conventional GWAS for complex traits or disease risk. For example, in TCGA data, the authors observed that a haplotype on 15q22.2 was associated with a 14-fold increased frequency of a CNV affecting *GNAQ*.

Leveraging this hypothesis generating computational approach, the authors discuss and experimentally validate two of their candidate associations. In any germline association study, establishing causal relationships is complicated by linkage disequilibrium and the fact that germline variants may be acting on a phenotype via a distal gene (10); here, establishing causality may be additionally complicated if the frequency of both germline and somatic variants differ between cancer types/subtypes. Thus, the authors wisely leverage additional biological knowledge and focused validation on instances where the germline variant and somatic variant pair likely affect genes that participate in the same pathway. They first highlight that an intronic SNP in *RBFOX1* is associated with somatic mutation of *SF3B1*, which in turn is associated with splicing of several genes. They also demonstrate that individuals who inherit specific germline variants on chromosome 19 are four times more likely to have somatic mutations in the oncogene *PTEN*, compared to cancer patients who do not inherit these variants. Two of the genes in the germline locus (*GNA11* and *STK11*) are known to act in the PIK3CA/mTOR pathway, in which PTEN plays a repressive role. In one of several experimental follow-ups, the authors demonstrated that an increase in *GNA11*

expression in HEK293T cells led to an increase in mTOR signaling, this effect was amplified in PTEN knockdowns. The findings are consistent with germline variants on chromosome 19 increasing the activity of GNA11, which in turn provides an additional selective advantage for PTEN inactivation. The validation work demonstrates the ability of creative computational analysis on large datasets to generate important biological hypotheses that can be investigated at the bench. One future direction may involve further integration of the results of this novel analysis with GWAS: while the authors suggest that many germline variants are associated with specific somatic mutation profiles among cancer patients, it may be interesting to assess whether the same variants are also strongly enriched for cancer risk variants. Although, it could be equally interesting to consider that germline variants that affect somatic mutation profiles of tumors may not necessarily affect cancer risk.

Finally, the authors propose a novel method of discovering new cancer genes, using the rationale that if certain germline loci are associated with somatic mutation profiles in known cancer genes, the same sets of loci may also be associated with somatic mutation profiles of as yet unknown cancer genes. This analysis yielded 20 additional genes, whose somatic mutation profiles differed between select germline variants. These genes represent a promising starting point for follow up work and if validated, could reveal new cancer genes, which may provide a selective advantage during tumorigenesis, but possibly only on a specific germline genetic background.

Overall, the authors have presented a systematic integrative analysis of germline genome variation and somatic mutation profiles in cancer patients. This new way of studying cancer has shed light on both disease risk and development. This paper opens the door for further integrative analysis of somatic and germline variation in cancer, which will help in furthering our understanding of the mechanisms by which germline variants are relevant in cancer. The findings also serve as an opportunity to highlight the benefit of investment in computational methodologies and projects focused on intelligent re-use of existing data. This continued investment in open science, data sharing and new computational approaches will certainly yield further benefit to the cancer research community, both in validating and broadening the analyses proposed by Carter *et al.* Furthermore, these investments will lead to novel methods to understand cancer in unforeseen ways, as our talented community of computational biologists continues to explore these invaluable public resources.

## Acknowledgments

### Grant support

RSH received support from the Avon Foundation research grant, NIH/NIGMS grant K08GM089941, NIH/NCI grant R21 CA139278, NIH/NIGMS grant UO1GM61393, Circle of Service Foundation Early Career Investigator award, University of Chicago Support Grant (#P30 CA14599), Breast Cancer SPORE Career Development Award (CA125183), the National Center for Advancing Translational Sciences of the NIH (UL1RR024999), University of Chicago CTSA core subsidy grant, and a Conquer Cancer Foundation of ASCO Translational Research Professorship award In Memory of Merrill J. Egorin, MD (awarded to Dr. MJ Ratain). PG received support from the Chicago Biomedical Consortium grant PDR-020.

## References

1. Chang, CQ., Yesupriya, A., Rowell, JL., Pimentel, CB., Clyne, M., Gwinn, M., et al. Eur J Hum Genet [Internet]. Vol. 22. Nature Publishing Group; 2014. A systematic review of cancer GWAS and candidate gene meta-analyses reveals limited overlap but similar effect sizes; p. 402-8.[cited 2017 Feb 13]Available from: <http://www.nature.com/doi/10.1038/ejhg.2013.161>
2. Kar SP, Beesley J, Amin AI, Olama A, Michailidou K, Tyrer J, Kote-Jarai Zs, et al. Genome-Wide Meta-Analyses of Breast, Ovarian, and Prostate Cancer Association Studies Identify Multiple New Susceptibility Loci Shared by at Least Two Cancer Types. Cancer Discov [Internet]. 2016; 6:1052–67. [cited 2017 Feb 13]. Available from: <http://cancerdiscovery.aacrjournals.org/cgi/doi/10.1158/2159-8290.CD-15-1227>.
3. Relling MV, Evans WE. Pharmacogenomics in the clinic. Nature [Internet] Nature Research. 2015; 526:343–50. [cited 2017 Feb 13]. Available from: <http://www.nature.com/doi/10.1038/nature15817>.
4. Morrison G, Lenkala D, LaCroix B, Ziliak D, Abramson V, Morrow PK, et al. Utility of patient-derived lymphoblastoid cell lines as an *ex vivo* capecitabine sensitivity prediction model for breast cancer patients. Oncotarget [Internet]. Impact Journals. 2014; 7:38359–66. [cited 2017 Feb 13]. Available from: <http://www.oncotarget.com/abstract/9521>.
5. Li Q, Seo J-H, Stranger B, McKenna A, Pe'er I, LaFramboise T, et al. Integrative eQTL-Based Analyses Reveal the Biology of Breast Cancer Risk Loci. Cell [Internet]. 2013; 152:633–41. [cited 2017 Feb 13]. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23374354>.
6. Ongen H, Andersen CL, Bramsen JB, Oster B, Rasmussen MH, Ferreira PG, et al. Putative cis-regulatory drivers in colorectal cancer. Nature [Internet] Nature Research. 2014; 512:87. [cited 2017 Feb 13]. Available from: <http://www.nature.com/doi/10.1038/nature13602>.
7. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science [Internet]. 1990; 250:1684–9. [cited 2017 Feb 13]. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2270482>.
8. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet [Internet]. Nature Research. 2015; 47:373–80. [cited 2017 Feb 15]. Available from: <http://www.nature.com/doi/10.1038/ng.3242>.
9. Carter H, Marty R, Hofree M, Gross A, Jensen J, Fisch KM, et al. Interaction Landscape of Inherited Polymorphisms with Somatic Events in Cancer. Cancer Discov. 2017
10. Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature [Internet]. Nature Research. 2014; 507:371–5. [cited 2017 Feb 15]. Available from: <http://www.nature.com/doi/10.1038/nature13138>.