JNCI: Journal of the National Cancer Institute, 2024, 116(12), 1882–1889
https://doi.org/10.1093/jnci/djae173
Advance Access Publication Date: July 25, 2024
Commentary

Variation to biology: optimizing functional analysis of cancer risk variants

Stefanie Nelson, PhD,[1,*] Danielle Carrick, PhD, MHS,[1] Danielle Daee, PhD,[1] Ian Fingerman, PhD,[2] Elizabeth Gillanders, PhD[1]

[1]Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD, USA
[2]Division of Cancer Biology, National Cancer Institute, Rockville, MD, USA

*Correspondence to: Stefanie Nelson, PhD, Division of Cancer Control and Population Sciences, National Cancer Institute, 9609 Medical Center Dr, Room 4E108, Rockville, MD 20850, USA (e-mail: stefanie.nelson@nih.gov).

Abstract

Research conducted over the past 15+ years has identified hundreds of common germline genetic variants associated with cancer risk, but understanding the biological impact of these primarily non-protein coding variants has been challenging. The National Cancer Institute sought to better understand and address those challenges by requesting input from the scientific community via a survey and a 2-day virtual meeting, which focused on discussions among participants. Here, we discuss challenges identified through the survey as important to advancing functional analysis of common cancer risk variants: 1) When is a variant truly characterized; 2) Developing and standardizing databases and computational tools; 3) Optimization and implementation of high-throughput assays; 4) Use of model organisms for understanding variant function; 5) Diversity in data and assays; and 6) Creating and improving large multidisciplinary collaborations. We define these 6 challenges, describe how success in addressing them may look, propose potential solutions, and note issues that span all the challenges. Implementation of these ideas could help develop a framework for methodically analyzing common cancer risk variants to understand their function and make effective and efficient use of the wealth of existing genomic association data.

On February 7-8, 2023, the National Cancer Institute (NCI) held a virtual meeting, Variation to Biology: Optimizing Functional Analysis of Cancer Risk Variants (1), to identify and discuss how to address scientific challenges and opportunities for understanding the path from common genetic variation to cancer phenotype. More than 200 participants attended, representing fields including cancer epidemiology, genetics, bioinformatics, and molecular biology.

Hundreds of common genetic variants that are associated with cancer risk have been identified through genomic association studies. Because most of these variants have small effect sizes and are located in non-protein coding regions of the genome, understanding how they impact molecular mechanisms and the underlying biology is challenging (2,3). The intent of this meeting was to focus on challenges for characterization of these primarily non-protein coding variants, conceptualize success in addressing these challenges, and brainstorm solutions.

To develop the meeting agenda, NCI solicited input from investigators working in this field, including genetic epidemiologists, molecular biologists, and cancer biologists, about significant challenges facing efforts to understand how common genetic variants impact cancer development. We received nearly 40 substantive responses to our survey and grouped these responses into 6 topics that were the focus of brief talks (Table 1): 1) When is a variant truly characterized; 2) Developing and standardizing databases and computational tools; 3) Optimization and implementation of high-throughput assays; 4) Use of model organisms for understanding variant function; 5) Diversity in data and assays; and 6) Creating and improving large multidisciplinary collaborations. These topics were used to generate a meeting agenda, featuring brief presentations on each topic, followed by small-group breakout sessions for in-depth discussions. The discussions were structured to define the problem, describe how success in addressing the problem would look, and brainstorm ideas for achieving success. In this report, we synthesize these discussions (summarized in Table 1). The ideas discussed during this workshop represent activities to facilitate the functional analysis of common cancer risk variants and maximize the value of existing genomic association data.

Topic 1: When is a variant truly characterized?

Defining the challenge

Understanding the impact of a genetic variant on cancer risk requires collaboration across multiple disciplines and consideration of numerous cancer types and cellular contexts. This complexity leads to the question of whether we consider a variant to be sufficiently characterized; in other words, when is there sufficient evidence to ensure that we accurately understand the impact, or lack thereof, of a variant on cancer risk?

What does success look like?

Deciding if a variant is sufficiently characterized will require different criteria for different downstream purposes. For example, biological and functional information may not necessarily be required for developing polygenic risk scores (PRS), where

Received: April 11, 2024. Revised: July 10, 2024. Accepted: July 16, 2024
Published by Oxford University Press 2024.

Table 1. Action items for advancing functional analysis of cancer risk variants

		Discussion topics					
Action item	**Details**	**Define sufficient characterization**	**Databases and computational tools**	**High-throughput assays**	**Model organisms**	**Diversity in data and assays**	**Collaborations**
Develop guidelines and/or standard for sufficient characterization of a genetic locus, based on intended purpose	Engage a small group of investigators to mine data and use benchmarks to adjudicate sources of data to identify a set of "truth" loci/variants. Identify consensus intermediate molecular phenotypes.	x					x
Develop and provide opportunities for cross-disciplinary training	Create opportunities to train across fields such as molecular biology, bioinformatics, epidemiology, genetics, etc. Promote opportunities for both early stage and mid-career investigators.	x				x	x
Encourage development of standards for in silico data curation and for high throughput assays	Develop standards for conducting research and reporting data/results. Work with existing databases to develop standards.		x	x	x		x
Build or connect data portals	All information/annotations for a locus should be easily integrated. Include published data from small studies that are not in larger databases. Connect data from various collaborative consortia.	x	x	x	x		x
Increase data, resources, and annotations relevant to underrepresented groups	Develop best practices for engaging under-represented groups. Ensure that new data generating efforts include all groups/ancestries.	x		x	x	x	x
Encourage creation of collaborative groups	Develop and encourage data jamborees or other activities focused on a specific project.	x	x				x

association with a clinical endpoint is the primary outcome of interest, although knowing the specific risk-promoting variant would add power. However, to understand the biological impact of risk variants and contribute to the development of potential interventions and treatments, more information is needed. An example of a "sufficiently characterized" variant may be exemplified by the rs11655237[A] allele within the locus *LINC00673* on chromosome 17q24.3, which was identified by genome-wide association studies (GWAS) to be associated with an increased risk for pancreatic ductal adenocarcinoma (PDAC) [(4,5) and reviewed in (6)]. Functional studies determined that the non-coding rs11655237[A] allele creates a miR-1231 binding site, reducing the ability of *LINC00673* to control PTPN11 degradation, thereby promoting proliferation and growth of PDAC cells (7).

Along with information on molecular mechanisms, data on interactions with environmental exposures, and how variants interact with each other, including how tumor mutations and germline variants interact, should be considered. Incorporating data from populations with a variety of genetic ancestries is highly important; although the biological mechanisms may be similar, the specific variants underlying these mechanisms may vary across racial and ethnic groups. Ideally, genetic (including family history of cancer), functional, and clinical information will be incorporated into variant analysis to create a clear picture of a variant's role in carcinogenesis.

Potential solutions

A consensus on sufficient variant characterization will require multiple independent lines of evidence that indicate the same conclusion (eg, in silico, epigenetic, functional assay data) and well-annotated, well-documented databases to capture and manipulate this information. Pipelines or standard processes to follow for characterization could be developed (8,9). Researchers could convene expert curation panels similar to those used by ClinGen (10) to develop and apply criteria to decide when a cancer risk variant is sufficiently characterized. Aspects of the variant curation guidelines adopted by the American College of Medical Genetics and Genomics could be incorporated, such as requiring data from multiple computational prediction programs, evaluating the strength of data from a range of sources

(eg, population data, computational and predictive data, functional assay data), and ensuring that potential risk variants have been evaluated in multiple ancestral groups (11). Common data and reporting standards could facilitate larger collaborations, and buy-in from journal editors could help develop evidence requirements and reporting standards. Although standard reporting guidelines exist for specific study types (eg, STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) for cohort and case-control studies, Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) for studies that use qPCR data) (12), research on non-coding variants encompasses multiple disciplines and data types, and reporting guidelines will need to consider this.

Topic 2: Standardizing computational tools and databases

Defining the challenge

The location of common cancer risk variants in non-protein coding regions means that extensive annotation of loci is needed to inform characterization approaches. A wealth of annotation data has been generated, but curating the data to make them accessible and usable to all investigators has been challenging. Multiple data types are needed to define the function of cancer risk variants, and although a great deal of data exist, they are often stored in separate databases; connecting and integrating these data to generate a unified description of a given locus is challenging. Analysis of germline cancer risk variants requires data from a broad set of potentially involved tissues, and the types of data available varies greatly across tissue types. In addition, few standards exist for assessing the accuracy and reliability of functional annotations. Another complicating issue is the lack of collaborative resources (eg, computational tools and databases) to allow investigators to easily integrate different types of data and/or annotations.

What does success look like?

Ideally, connected or federated databases, with data broadly available through controlled access, would allow investigators with different levels and areas of expertise to easily find and integrate all data related to a given locus or variant. For any given locus, a defined set of annotations should be available, such as allele frequencies in multiple populations; inter- or intragenic locations (eg, coding vs non-coding regions); filterable expression data for multiple contexts (including by ancestry, cell state, tissue of origin); chromatin state (eg, open/closed) for multiple contexts; and regulatory annotations (eg, transcription factor binding sites, promoters, enhancers). A well-characterized locus (eg, *LINC00673* as discussed above) could be used as a template for standardizing and optimizing the usability and accuracy of functional annotations, supported by laboratory assay data. Artificial intelligence-based and large-scale computational approaches could help analyze data (13), and these tools and their outputs should be broadly shared across the community.

Potential solutions

A primary focus for improving computational tools and databases for cancer research should be the connection or integration of the many different sources of data. Action items for achieving this goal include connecting or consolidating existing databases; although most atlases focus on coding variants, efforts such as the Atlas of Variant Effects, which strives to develop comprehensive variant effect maps that characterize the function of single nucleotide changes in a gene or functional element of interest (14,15), could serve as an example for developing resources for cancer-relevant variants. Agreed-upon standards for reporting meta-data and results, which could be modeled on those used by the Cancer Genome Atlas (TCGA) (16) or the Genotype-Tissue Expression (GTEx) project (17), for example, will help to ensure these disparate sources of data are interoperable and can be easily integrated. Although not strictly a matter of tools and databases, enhancing cross-training of and collaboration among bioinformaticists who can manipulate the data and molecular and cancer biologists who can interpret the data will help to ensure that all data are used widely and appropriately.

An initial step could include working with NCI's Cancer Research Data Commons (CRDC) (18), which has initially focused on tumor genomic data, to incorporate and annotate more data relevant to germline risk variant analysis. NIH's Common Fund Data Ecosystem is working to create a federated genomic data ecosystem (currently focused on Common Fund datasets) to ensure interoperability of a range of data (19). cBioPortal for Cancer Genomics (20) is another example of a data consolidation and visualization tool, although inclusion and annotation of germline variants will require additional data security considerations. Inclusion of cancer data in the Association to Function Knowledge Portal (21), which aggregates datasets and bioinformatic methods for several diseases, could also be a way to reach more scientists interested in exploring the germline basis of cancer risk.

Topic 3: High-throughput assays

Defining the challenge

Assays to understand the function of common variants must contend with the lack of knowledge concerning their precise molecular effects and the genes they may modulate. Because these variants have small effects and likely work in combination with others to promote carcinogenesis, diverse data types are needed to fully understand their impact on cancer risk. Multiplexed assays of variant effects (MAVEs) can generate large-scale functional data for all possible variants at a locus (22), but interpretation of these data requires a complex workflow, including choosing reliable, consistent assays; specialized knowledge to understand assay results; and considering the results in the context of other results (23,24). Curated, interoperable datasets are necessary; a near-complete understanding of a variant's biological impact will incorporate information ranging from DNA sequence changes to effects on processes including transcription, transcript stability, and protein properties, among others. Accurate integration of all these types of data underscores the need for "gold standards" that ensure that the data reported from an assay provide an accurate picture of variant action. Standards or agreements on how best to integrate the multiple levels of data are needed, including attention to context (eg, gene by gene interactions, cell type in which the assay is conducted, and appropriate use of organoids and other higher-level model systems), to accurately understand the molecular changes that lead to the cancer risk phenotype.

What does success look like?

Successfully addressing these challenges will result in confidence that the data from assays are reported accurately and can be built on by other groups to answer subsequent questions arising on the path to understanding the phenotype. The results of

assays should be easily findable and understandable—given a specific variant, it should be straightforward to find assay results describing its effects and the gene(s) it impacts. Investigators also should have access to sufficient information to combine data from different assays and/or laboratories and compare effects of a variant across different assays and cellular and organismal contexts. Similar to what was discussed for computational data, establishing an integrated repository of data from high-throughput multiplexed assays would help this area of research. Ideally, a single portal could be created, through which current knowledge can be easily accessed and gaps in knowledge identified. This portal would have, for a given variant, information on genes the variant may affect, intermediate phenotypes, data on positive and negative controls, and the ability to compare the data across different assays and cell types, as well as information on cell lines and other reagents. This portal should be usable by investigators who may not be experts in bioinformatics.

Potential solutions

A collaborative framework that standardizes the generation of data and reporting of results from high-throughput assays in a methodical way could help to ensure accurate assessment and broad use of the data. An important first step would be defining categories across the universe of cancer genetic data (eg, mechanisms, genetic changes, cascades). This process could follow the example of TCGA, which published reports defining, in a uniform way, genomic changes occurring in multiple cancer types [eg, (16,25)]. Although Clinical Interpretation of Variants in Cancer (ClinVar) and ClinGen largely focus on variants found in protein-coding regions, the standards these groups use to define variant function also could be considered. Development of functional standards and tools would need to be led by interested researchers and experts in standards development, perhaps reaching a point at which these standards could help guide editorial decisions for publication.

Use of high-throughput data could be facilitated by coordinating data from many research groups and sources in a user-friendly database. The database could be arranged around specific variants, or around specific genes or pathways, keeping in mind that for many common cancer risk variants, the target gene is unknown. Data on cell lines or organoids used in certain assays should be included, along with guidelines for conducting the assays. The Multiplexed Assays of Variant Effect database (MAVEdb) (26) provides a useful example of a database that consolidates the results of multiplex assays of variants effects, providing researchers with a way to store processed MAVE datasets and associated metadata in a standardized and searchable format. MAVEdb also has a web interface that allows researchers to easily access and analyze the data.

Topic 4: Model organisms

Defining the challenge

Model organisms have been instrumental in helping investigators move beyond a basic understanding of molecular mechanism to determine how rare, high-penetrance genetic variants impact carcinogenesis (27). Challenges for analyzing common genetic variants in models include their noncoding nature and small effect sizes that imply that these variants act in combination with other variants and over the lifespan. Although computational and in vitro assay approaches can generate a great deal of information about a variant, completing the trajectory from genetic variation to cancer will require testing in systems that can approximate tissue of origin, differentiation state, cell-cell interactions, and the impact of the immune system, among other issues. Diverse expertise will also be needed to design and interpret experiments using model organisms, because cancer arises in and affects multiple tissue types.

What does success look like?

Developing ways in which to model the impact of low-penetrance variants in systems that recapitulate real-life conditions will require consolidation and integration of data from computational and in vitro assay/MAVE approaches. Ideally, these models will allow for testing combinations of variants, perhaps based on contributions of specific variants to PRS, pleiotropy, or clinical relevance. This work will require large-scale collaborations, which will be aided and enhanced by developing standards and guidelines that are used across research groups to ensure that all data are comparable and can be integrated. For example, a standardized report could include descriptions of model development, assay details similar to those suggested for MAVEs, and standard reporting of results, for example, normalizing signals to an agreed-upon reference, among other details.

Potential solutions

Although each cancer type or variant will require slightly different approaches for characterization, researchers could consider creating guidelines for suggesting which models are most appropriate for the molecular mechanism or cancer type under study, and also pipelines and standards for reagents (eg, cell lines) and reporting results [such as that outlined in (8)]. Given that cancer develops slowly, researchers could define molecular events or intermediate phenotypes for model organisms that indicate the likelihood that a variant contributes to carcinogenesis.

Initial efforts to create this framework could focus on a locus with existing functional information (ie, *LINC00673*), variants with pleiotropic or stronger effects, or variants considered critical to characterize. This exercise could be used to develop standards for reagents (eg, cell lines), for reporting the results of experiments in model organisms (for example, normalizing signals to a common reference) and to develop guidelines for suggesting which models are most appropriate for the molecular mechanism or cancer type under study.

A comprehensive approach to analyzing function will include induced pluripotent stem cells (iPSCs) that can capture naturally occurring genetic variation and differentiate into multiple tissue types; organoids that include tumor microenvironment; and mouse models that recapitulate elements of cellular context (eg, tissue specificity), particularly the presence of an intact immune system. Although mostly used for genetic mapping or for the study of rare coding variants, researchers studying common cancer risk variants should also consider the utility of mouse models available through the Collaborative Cross and Diversity Outbred populations (28) as well as the Mouse Models of Human Cancers database (29) to study variant mechanisms on a complex yet defined genetic background. The advent of CRISPR-based approaches for engineering genetic variation into organisms will be helpful for exploring the effects of common cancer risk variants. CRISPR-based techniques can be used to create small alterations targeted to specific cell types (30), which are then introduced into mice and can lead to tumor development. Although this approach examines variant effects in a somatic rather than germline context, it could provide critical insight into variant function in vivo.

Topic 5: Diversity in data and assays

Defining the challenge

Although vast quantities of genomic data exist, these data are not representative of the world's populations; currently, more than 80% of existing GWAS data were generated from individuals of European ancestry (31,32). The recent publication of a new human pangenome reference (33) represents a significant advance in inclusion of other ancestry groups, but further efforts are needed to increase diversity of other types of data (eg, expression data) and resources. For example, data in TCGA and GTEx are derived mainly from people of European ancestry, as are 70% of the cancer cell lines in the Catalogue Of Somatic Mutations In Cancer (COSMIC) (34), and immortalized cell lines in general are of limited ancestral diversity (35).

What does success look like?

Successfully addressing lack of diversity in data will be achieved through investing in efforts to build trust, educate, and involve all participants (including researchers and other stakeholders, as discussed below) in biomedical research activities to ensure that all groups benefit. The All of Us program represents a useful example for expanding research participation (36), as does the Participant Engagement and Cancer Genome Sequencing Network (PE-CGS), which aims to promote direct engagement of diverse and underrepresented cancer patients and survivors as participants in cancer research (https://epi.grants.cancer.gov/events/pe-cgs/). Researchers working in the field of functional analysis should articulate the value of diversity, not only for variant discovery, but also to ensure that characterization considers a variety of contexts, endogenous and exogenous factors, environmental exposures, and lifestyle differences. If successful, sample and data distributions that reflect global populations will be generated. This will include but is not limited to variant allele frequencies from a diverse set of populations; expression data from diverse groups; reference controls for functional characterization that include population-specific or population-enriched variants; methods for characterizing admixture; and acknowledgment of the importance of diversity in functional characterization work. Researchers must determine whether standard models (ie, tissue culture systems and organoids; normal and tumor tissue samples) accurately reflect the ancestry of all population groups.

Potential solutions

A first step toward improving the diversity of data will be to assess the diversity of existing resources; for example, epidemiologic cohorts, research consortia, and patient-focused data collections could be surveyed to determine whether their data are representative of a range of ancestries. Development of diversity-focused genomic resources, such as "diversity GTEx" or "diversity TCGA" would help to provide the data needed to ensure all groups are represented. Importantly, when considering whether a variant is sufficiently characterized, investigators must be sure that functional analyses have included variants from all ancestries. Because results from this work are translated to screening tools and potential therapies, ensuring that these are built on data representing a variety of genetic ancestries will improve their relevance for all groups, potentially addressing current disparities (37,38). Researchers should consider working with PE-CGS and All of Us to learn how to engage with groups that have been traditionally under-represented in this research.

Topic 6: Large multidisciplinary collaborations

Defining the challenge

Multidisciplinary efforts will be crucial for optimal functional characterization of common cancer risk variants, especially because no 1 laboratory will have all the expertise needed for generating, analyzing, and integrating the many different data types needed to understand variant function. Assembling effective collaborations focused on cancer can be difficult because of its relative rarity and the heterogeneity of the disease, including differences in both genetic and environmental risk factors. During the discussions of each of the challenges, the need for multidisciplinary collaborations was evident.

What does success look like?

NCI's Genetic Associations and Mechanisms in Cancer initiative (39) provides an example of an effort that brought together genetic epidemiologists, biologists, and clinical scientists to identify cancer risk variants, determine their biological function, and understand their potential clinical impact. This and formal collaborations such as the International Common Disease Alliance (ICDA) (40) and the Impact of Genomic Variation on Function Consortium (IGVF) (41) could serve as models for collaborative efforts. ICDA and IGVF could provide partnership opportunities, as well as examples of ways to engage investigators and develop clear goals and structures for collaborative projects.

All collaborative efforts should include free and open sharing of data. Participants should develop and employ standards for generating and reporting data from computational analyses, high-throughput assays, and model organisms, as discussed previously. Additionally, agreeing on a definition of "sufficient" variant characterization will help to define and refine the goals of the group.

Inclusion must be a priority. From the start, collaborative efforts should work actively to include smaller institutions, junior faculty, trainees, and underrepresented researchers, to ensure diversity of both data and ideas. Opportunities for training in various disciplines or to learn new approaches can provide an incentive for participation, along with opportunities for publications, grants, and greater exposure and visibility (especially for more junior investigators). The National Cancer Plan specifically calls for engaging every person (https://nationalcancerplan.cancer.gov/goals/engage-every-person) and optimizing the workforce (https://nationalcancerplan.cancer.gov/goals/optimize-the-workforce) and includes activities and resources for achieving broad inclusion in collaborative efforts. As a specific example of fostering advancement of junior researchers, NCI's Human Tumor Atlas Network (https://humantumoratlas.org/) and PE-CGS encourage junior investigators to develop and run consortia meetings, providing additional experience and networking opportunities. Travel to other groups to learn new techniques could be sponsored or intensive workshops or jamborees and hackathons (42) could be held to help encourage collegiality and novel approaches to problems.

Potential solutions

Initial activities for building these collaborations should include investigating the feasibility of partnerships with existing efforts or building on lessons learned to establish new collaborations. A comprehensive survey of existing cancer-risk variants would help to focus and define the scope of a new collaboration, with special attention to ensuring that the survey includes data from

diverse populations. Examples from ICDA and IGVF could be used to develop experimental and data submission standards adapted to the cancer risk context. The crowdsourced ClinVar (43), cBioPortal for Cancer Genomics (20), ENCODE (44), and NCI's Cancer Research Data Commons (18) could provide models for data curation. Examples of collaborations that could be leveraged for standards development include organizations such as the Global Alliance for Genomics and Health (45), the Variant Interpretation for Cancer Consortium (46), and the Cancer Genomics Consortium (47). Although many of these groups focus on tumor mutations or rare variants with large effect sizes, the standards they describe could be adapted to common, germline variants.

Summary

This meeting convened a wide range of stakeholders to identify challenges in moving from identification of common, primarily non-coding, germline cancer risk variants to cancer phenotype. Participants were given the opportunity to brainstorm about what success in meeting those challenges would look like, and then were asked to discuss ways to achieve that success.

Based on the meeting discussions, key action items for advancing this field were identified: 1) create guidelines for sufficient variant characterization; 2) develop standards for in silico data curation and functional assays; 3) create interconnected data portals; 4) enhance data generation and annotation from under-represented groups; 5) create opportunities for cross-disciplinary training and research; and 6) encourage development of collaborative research groups. These items are summarized in Table 1, along with details for achieving these actions. The robust participation in this meeting and response to a post-meeting survey to assess interest in working to further develop the ideas discussed indicate significant interest from the research community in this area. We plan to convene smaller, focused working groups to develop detailed ways to realize action items.

Many of these action items pertain to more than 1 topic, underscoring the interdependence and interconnectedness of the 6 discussion topics. Taken together, these topics may define a framework for consistent, comprehensive characterization of risk variants (Figure 1). Understanding whether a variant is sufficiently characterized will require annotation of newly identified risk variants in silico, followed by high-throughput assays to begin to develop a functional understanding of and potential role for the variant in cancer risk, and testing variants in cells or model organisms to get closer to a "real-life" setting. This information can then be considered regarding whether our knowledge of a risk variant is "fit for purpose"; for example, confirming an association may be sufficient for use in PRS, but additional functional knowledge may be needed for possible therapeutic development. A common thread through all these activities is the need for diversity not only in source data, but also in the cells and model organisms used to test hypotheses and functions and in investigators working in this field. This interdependence also emphasizes the desirability of large, multidisciplinary collaborations to advance our understanding of the function of germline cancer risk variants.

Navigating the path from common genetic variant discovery to cancer phenotype will be challenging but, ultimately,

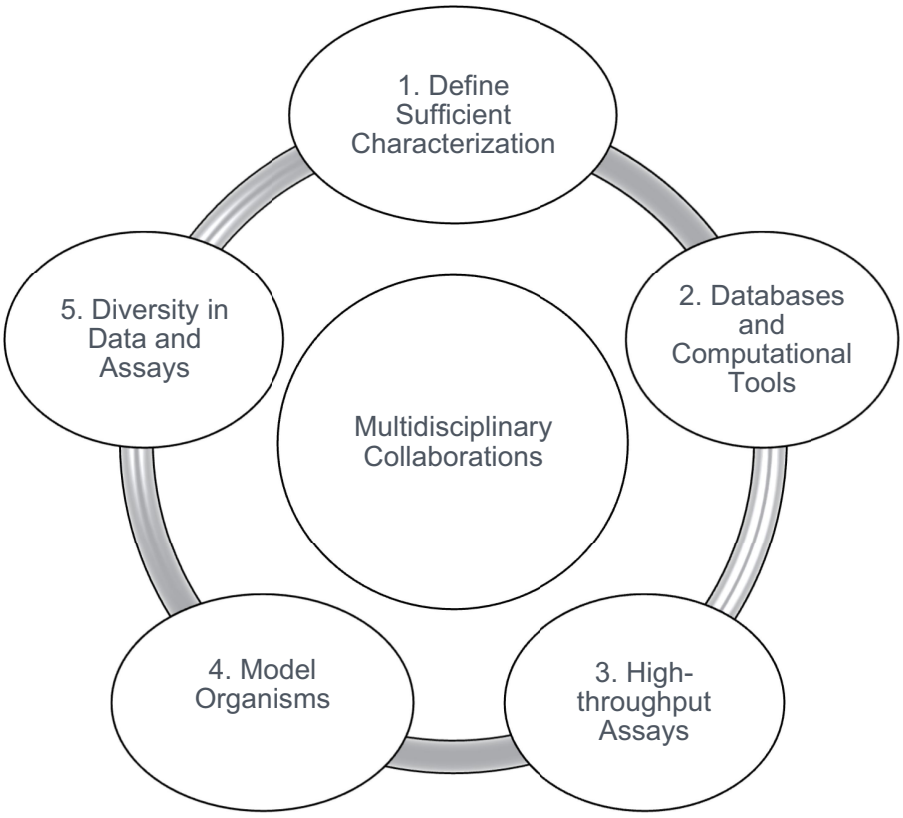

Figure 1. Framework for addressing challenges and opportunities to optimize functional analysis of germline cancer risk variants: The topics discussed during this meeting are connected and interdependent, forming a framework for variant characterization. The need for multidisciplinary collaboration is an overarching need that affects all topics. The goal of this framework is to promote efficient and consistent functional characterization of cancer risk variants.

rewarding by helping us develop a better understanding of how genetic variation affects risk, perhaps leading to the discovery of novel carcinogenesis mechanisms, insights into cancer cell vulnerabilities, and potentially new risk mitigation and treatment options.

Data availability

No new data were generated or analyzed for this Commentary.

Author contributions

Stefanie Nelson, PhD (Conceptualization; Project administration; Writing—original draft; Writing—review & editing), Danielle Carrick, PhD (Conceptualization; Writing—original draft; Writing—review & editing), Danielle Daee, PhD (Conceptualization; Writing—original draft; Writing—review & editing), Ian Fingerman, PhD (Conceptualization; Writing—original draft; Writing—review & editing), Elizabeth Gillanders, PhD (Conceptualization; Writing—original draft; Writing—review & editing).

Funding

This work was authored by employees of the National Cancer Institute.

Conflicts of interest

The authors declare no competing interests.

Acknowledgments

The authors would like to acknowledge Kevin Brown, Alexander Gusev, Alvaro Monteiro, Lea M. Starita, and Amanda E. Toland for their help in defining the topics covered in this meeting, preparing presentations, and moderating breakout sessions.

References

1. NCI. *Variation to Biology: Optimizing Functional Analysis of Cancer Risk Variants*. 2023. https://epi.grants.cancer.gov/events/functional-analysis/. Accessed March 2024.
2. Yang W, Zhang T, Song X, et al. SNP-target genes interaction perturbing the cancer risk in the post-GWAS. *Cancers (Basel)*. 2022;14(22):5636.
3. Abdellaoui A, Yengo L, Verweij KJH, et al. 15 years of GWAS discovery: realizing the promise. *Am J Hum Genet*. 2023;110(2):179-194.
4. Wu C, Miao X, Huang L, et al. Genome-wide association study identifies five loci associated with susceptibility to pancreatic cancer in Chinese populations. *Nat Genet*. 2011;44(1):62-66.
5. Childs EJ, Mocci E, Campa D, et al. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. *Nat Genet*. 2015;47(8):911-916.
6. Gentiluomo M, Canzian F, Nicolini A, et al. Germline genetic variability in pancreatic cancer risk and prognosis. *Semin Cancer Biol*. 2022;79:105-131.
7. Zheng J, Huang X, Tan W, et al. Pancreatic cancer risk variant in LINC00673 creates a miR-1231 binding site and interferes with PTPN11 degradation. *Nat Genet*. 2016;48(7):747-757.
8. Spisák S, Lawrenson K, Fu Y, et al.; GAME-ON/ELLIPSE Consortium. CAUSEL: an epigenome- and genome-editing pipeline for establishing function of noncoding GWAS variants. *Nat Med*. 2015;21(11):1357-1363.
9. Freedman ML, Monteiro ANA, Gayther SA, et al. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet*. 2011;43(6):513-518.
10. Clinical Genome Resource. *About ClinGen Expert Panels*. 2023. https://clinicalgenome.org/affiliation/. Accessed March 2024.
11. Richards S, Aziz N, Bale S, et al.; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, 2015;17(5):405-424.
12. JNCI. *Reporting Guidelines. General Instructions*. https://academic.oup.com/jnci/pages/general_instructions. Accessed March 2024.
13. Jiang P, Sinha S, Aldape K, et al. Big data in basic and translational cancer research. *Nat Rev Cancer*. 2022;22(11):625-639.
14. Fowler DM, Adams DJ, Gloyn AL, et al. An Atlas of Variant Effects to understand the genome at nucleotide resolution. *Genome Biol*. 2023;24(1):147.
15. Atlas of Variant Effects Alliance. *Atlas of Variant Effects*. 2023. https://www.varianteffect.org/about. Accessed March 2024.
16. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008;455(7216):1061-1068.
17. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318-1330.,
18. NCI. *NCI Cancer Research Data Commons*. 2023. https://datascience.cancer.gov/data-commons. Accessed March 2024.
19. NIH. *Common Fund Data Ecosystem (CFDE)*. 2023. https://commonfund.nih.gov/dataecosystem. Accessed March 2024.
20. cBioPortal. *cBioPortal for Cancer Genomics*. 2023. https://www.cbioportal.org/. Accessed March 2024.
21. A2F. *Association to Function Knowledge Portal*. 2023. https://a2f.hugeamp.org/. Accessed March 2024.
22. Starita LM, Ahituv N, Dunham MJ, et al. Variant interpretation: functional assays to the rescue. *Am J Hum Genet*. 2017;101(3):315-325.
23. Kircher M, Ludwig KU. Systematic assays and resources for the functional annotation of non-coding variants. *Med Genet*. 2022;34(4):275-286.
24. Gasperini M, Starita L, Shendure J. The power of multiplexed functional analysis of genetic variants. *Nat Protoc*. 2016;11(10):1782-1787.
25. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487(7407):330-337.
26. Esposito D, Weile J, Shendure J, et al. MaveDB: an open-source platform to distribute and interpret data from multiplexed assays of variant effect. *Genome Biol*. 2019;20(1):223.
27. Kersten K, de Visser KE, van Miltenburg MH, et al. Genetically engineered mouse models in oncology research and cancer medicine. *EMBO Mol Med*. 2017;9(2):137-153.
28. Saul MC, Philip VM, Reinholdt LG, et al.; Center for Systems Neurogenetics of Addiction. High-diversity mouse populations for complex traits. *Trends Genet*. 2019;35(7):501-514.
29. Mouse Models of Human Cancer Database. *Mouse Models of Human Cancer Database*. 2023. https://tumor.informatics.jax.org/mtbwi/index.do. Accessed March 2024.

30. Bu W, Creighton CJ, Heavener KS, et al. Efficient cancer modeling through CRISPR-Cas9/HDR-based somatic precision gene editing in mice. *Sci Adv*. 2023;9(19):eade0059.
31. Mills MC, Rahal C. The GWAS Diversity Monitor tracks diversity by disease in real time. *Nat Genet*. 2020;52(3):242-243.
32. Fatumo S, Chikowore T, Choudhury A, et al. A roadmap to increase diversity in genomic studies. *Nat Med*. 2022;28(2):243-250.
33. Liao W-W, Asri M, Ebler J, et al. A draft human pangenome reference. *Nature*. 2023;617(7960):312-324.
34. Kessler MD, Bateman NW, Conrads TP, et al. Ancestral characterization of 1018 cancer cell lines highlights disparities and reveals gene expression and mutational differences. *Cancer*. 2019;125(12):2076-2088.
35. Zaaijer S, Capes-Davis A. Ancestry matters: building inclusivity into preclinical study design. *Cell*. 2021;184(10):2525-2531.
36. Denny JC, Rutter JL, Goldstein DB, et al.; All of Us Research Program Investigators. The "All of Us" Research Program. *N Engl J Med*. 2019;381(7):668-676.
37. Lee KK, Rishishwar L, Ban D, et al. Association of genetic ancestry and molecular signatures with cancer survival disparities: a pan-cancer analysis. *Cancer Res*. 2022;82(7):1222-1233.
38. Jia G, Ping J, Guo X, et al. Genome-wide association analyses of breast cancer in women of African ancestry identify new susceptibility loci and improve risk prediction. *Nat Genet*. 2024;56(5):819-826.
39. NCI. *Genetic Associations and Mechanisms in Oncology (GAME-ON) Initiative*. 2024. https://epi.grants.cancer.gov/gameon/. Accessed April 1, 2024.
40. ICDA. *International Common Disease Alliance*. 2023. https://www.icda.bio/. Accessed March 2024.
41. IGVF Consortium. *Impact of Genomic Variation on Function Consortium*. 2023. https://igvf.org/. Accessed March 2024.
42. Peshkin L, Kirschner MW. A cell type annotation Jamboree—revival of a small acommunal science forum. *Genesis*. 2020;58(9):e23383.
43. National Library of Medicine. *ClinVar*. 2023. https://www.ncbi.nlm.nih.gov/clinvar/intro/. Accessed March 2024.
44. Luo Y, Hitz BC, Gabdank I, et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res*. 2020;48(D1):D882-D889.
45. Global Alliance for Genomics and Health. *GA4GH*. 2023. https://www.ga4gh.org/. Accessed March 2024.
46. Global Alliance for Genomics and Health. *Variant Interpretation for Cancer Consortium (VICC)*. 2023. https://cancervariants.org/. Accessed March 2024.
47. Cancer Genomics Consortium. *Cancer Genomics Consortium*. 2023. https://www.cancergenomics.org/. Accessed March 2024.

Published by Oxford University Press 2024.

JNCI: Journal of the National Cancer Institute, 2024, 116, 1882–1889
https://doi.org/10.1093/jnci/djae173
Commentary