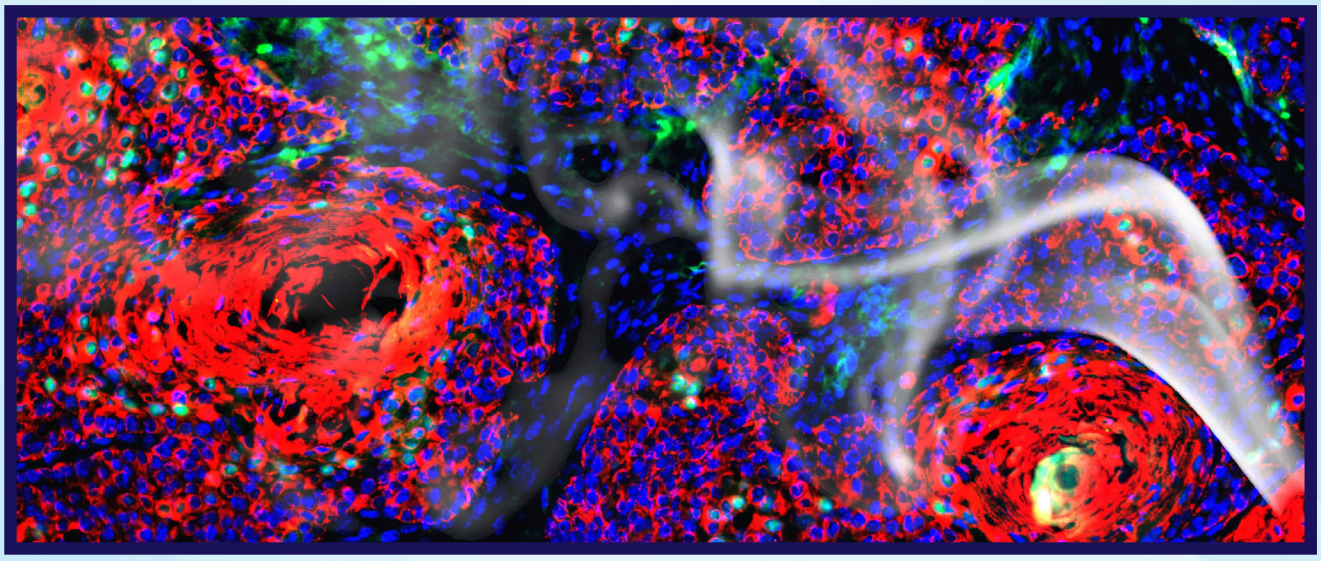
National Institute of Environmental Health Sciences

Molecular Signatures of Exposure in Cancer: A Joint NIEHS and NCI Workshop



Workshop Report

National Institutes of Health • U.S. Department of Health and Human Services

This workshop report was prepared by Megan Boland, MDB, Inc. This work has been funded with federal funds from the National Institute of Environmental Health Sciences, National Institutes of Health, under Contract No. HHSN273201800006I. The statements, conclusions, and recommendations contained in this document reflect opinions of the workshop participants and are not intended to represent the official position of the National Institute of Environmental Health Sciences, the National Cancer Institute, the National Institutes of Health, or the U.S. Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

The author thanks Ron Johnson, Dan Shaughnessy, the NIEHS-NCI Cancer and the Environment Working Group, and Julie Leibach for their review of earlier versions of this workshop summary.

The workshop planners gratefully acknowledge the contributions of all workshop participants, listed in Appendix 1, for their contributions to workshop deliberations and outcomes.

Suggested citation: Molecular Signatures of Exposure in Cancer. National Institute of Environmental Health Sciences & National Cancer Institute, June 29-30, 2023.
https://www.niehs.nih.gov/news/events/pastmtg/2023/molecular-signatures/index.cfm

Table of Contents

List of Abbreviations and Acronyms

AA	Aristolochic acid
APOBEC	Apolipoprotein B mRNA Editing Catalytic Polypeptide-like
BCERP	Breast Cancer and the Environment Research Program
BMI	Body mass index
CIMP	CpG island methylator phenotype
COSMIC	Catalogue of Somatic Mutations in Cancer
EBV	Epstein-Barr Virus
EWAS	Epigenome wide association study
GC-HRMS	Gas chromatography-high resolution mass spectrometry
HCC	Hepatocellular carcinoma
HPLC-ICPMS	High-performance liquid chromatography with inductively coupled plasma mass spectrometry
HTAN	Human Tumor Atlas Network
ICR	Imprint control region
ICPMS	Inductively couple plasma mass spectrometry
iPSC	Induced pluripotent stem cells
MC-ICPMS	Multicollector inductively coupled plasma mass spectrometry
MEC	Multiethnic Cohort
MEF	Mouse embryonic fibroblast
NCI	National Cancer Institute
NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
PFAS	Per- and polyfluoroalkyl substances
RCC	Renal cell carcinoma
SBS	Single base substitution
TCGA	The Cancer Genome Atlas

Glossary

Term	Definition
Epigenomics	Epigenomics describes the study of the epigenome. Derived from Greek, epigenome means “above” the genome. The epigenome consists of chemical compounds that modify DNA. Those modifications, also called marks, change how the body reads DNA to make proteins. Although epigenetic marks are not genetic material, they can be passed on through cell division, and from one generation to the next.[1]

[1] According to the Fact Sheet on Epigenomics published by the National Human Genome Research Institute: http://www.genome.gov/27532724

Term	Definition
Exposome	The exposome comprises all the exposures an individual experiences in a lifetime and how those exposures relate to health. Exposure begins in utero and includes effects from environmental and occupational sources. This report will discuss how exposures from an individual's environment, diet, and lifestyle interact with unique personal characteristics such as genetics, physiology, and epigenetics to impact health.[2]
Genetics	Genetics refers to the study of genes and their roles in inheritance and explores how specific traits or conditions are biologically passed down from one generation to another. Genes carry the instructions for making proteins, which direct the activities of cells and the functions of the body. Examples of genetic, or heritable, medical conditions include cystic fibrosis, Huntington's disease, and phenylketonuria (PKU).[3]
Genomics	Genomics, a more recent term than genetics, describes the study of all of a person's genes (the genome), including interactions among genes and the person's environment. Genomics includes the scientific study of complex diseases such as heart disease, asthma, diabetes, and cancer. Such diseases are typically caused by a combination of genetic and environmental factors rather than by individual genes. Genomics offers new possibilities for more targeted therapies and treatments for complex diseases, as well as new diagnostic methods.[4]
Metabolomics	Metabolomics is the study of the biological metabolic profile of a cellular specimen in a specific environment at an isolated timepoint. This discipline depicts the physiological states of cells and organisms by focusing on carbohydrates, lipids, and other metabolites. Several analytical techniques, such as mass spectrometry and electrophoretic applications, are utilized to quantify the metabolic content of specimens.[5]
Mutational signature	Different mutational processes generate unique combinations of mutations, called mutational signatures that are commonly represented as single base substitution (SBS), doublet base substitution (DBS), small insertions-deletions (ID) or copy number variation (CN) signatures. Mutations accumulate throughout life in the human body. Errors in DNA replication, exposures to mutation-causing agents produced inside or outside the body, enzymatic modifications to DNA, and defective DNA repair can all cause mutations.[6]

[2] National Institute for Occupational Safety and Health (NIOSH). Exposome and Exposomics. Available at https://www.cdc.gov/niosh/topics/exposome/
[3] Frequently Asked Questions About Genetic and Genomic Science. NHGRI. Available at http://www.genome.gov/19016904
[4] See footnote 3 above.
[5] Derived from the NCI Thesaurus found at https://ncit.nci.nih.gov/ncitbrowser/pages/multiple_search.jsf?nav_type=terminologies
[6] https://cancer.sanger.ac.uk/signatures/

Term	Definition
Omics	The informal term "omics" describes the study of related sets of biological molecules.[7] Examples of omics disciplines include genomics, transcriptomics, proteomics, metabolomics, and epigenomics. This report, broadens the definition to encompass exposomics and phenotypes resulting from environmental exposures.
Organoid	An organoid is a tiny, three-dimensional mass of tissue derived from stem cells, or cells that can develop into specialized cells. Organoids can simulate human tissues and organs or specific types of tumors. Scientists use organoids to study how normal tissues or cancers form and to test new drugs and other types of treatment before they are given to people.[8]
Phenotype	A phenotype is the set of observable characteristics resulting from interactions between an individual's genes and the environment. Phenotype can refer to common traits, such as height or hair color, or to the presence or absence of a disease, among other qualities.[9]
Polygenic risk score (PRS)	A polygenic risk score provides a measure of an individual's disease risk based on their genetic makeup. Combining polygenic risk scores with other factors that affect disease risk, such as environmental exposures, can improve insight into how likely an individual is to develop a specific disease.[10]
Precision Environmental Health	Precision environmental health focuses on individualized risk assessment and interventions to prevent disease. Practitioners aim to reduce adverse health effects from exposures through air, water, and food by identifying individuals who are specifically susceptible and enabling precise, targeted, and effective prevention.[11]
Proteomics	Proteomics is the study of the structure and function of proteins, including the way they work and interact with each other inside cells.
Transcriptomics	Transcriptomics is the study of the transcriptome, or the complete set of RNA transcripts — instructions for making proteins — that are produced by the genome, under specific circumstances or in a specific cell.[12]

[7] IOM (Institute of Medicine). 2012. Evolution of Translational Omics: Lessons Learned and the Path Forward. Washington, DC: The National Academies Press.
[8] NCI Dictionary of Cancer Terms. Available at https://www.cancer.gov/publications/dictionaries/cancer-terms/def/organoid
[9] Phenotype. NHGRI Talking Glossary of Genomic and Genetic Terms. Available at https://www.genome.gov/genetics-glossary/Phenotype
[10] Polygenic Risk Scores. CDC, Genomics & Precision Health. Available at https://www.cdc.gov/genomics/disease/polygenic.htm
[11] Walker CL, Dolinoy D, Baccarelli A. Perspectives on Precision Environmental Health. National Advisory Health Sciences Council, February 2021.
[12] https://www.nature.com/subjects/transcriptomics

Executive Summary

The National Institute of Environmental Health Sciences (NIEHS) and the National Cancer Institute (NCI) convened a virtual workshop, Molecular Signatures of Exposure in Cancer, on June 29 and 30, 2023. The meeting aimed to i) determine how research in mutational, epigenetic, and other omics signatures can further inform cancer etiology, ii) assess whether signatures of exposure can be identified from other omic data types, and iii) determine how these approaches can be applied to cancer prevention and uncover cancer etiologies linked to environmental exposures.

The workshop goals were to:

- Define the current state of the science of using molecular signatures to link environmental exposures to cancer.
- Prioritize pressing needs and opportunities that, if pursued, would aid progress in identifying molecular signatures of exposure in cancer.
- Explore ways to use molecular signatures of exposure to improve cancer prevention.

The workshop was co-chaired by Cheryl Walker, Ph.D., of Baylor College of Medicine, and Hannah Carter, Ph.D., of the University of California, San Diego, and planned by members of the NIEHS-NCI Cancer and the Environment Working Group. Each of the five topic sessions offered a state-of-the-science overview followed by three brief, targeted presentations and an extensive moderated discussion panel. Combined, the sessions featured talks from 26 leading investigators with expertise in computational biology, epidemiology, exposure science, and cancer biology.

A sixth and final session, comprising all the invited experts, summarized the workshop findings and discussed future directions. Over 1,200 people attended the meeting and participated by sending questions and comments during presentations and discussion sessions. Below are key points identified in each of the topic sessions as well as a summary of key needs within the field.

Session 1: Mutational Signatures of Exposure in Cancer

- The effect of exposures on mutational signatures can be complex, multi-factorial, and dependent on context; mutational signature profiles can be altered by factors such as tissue and cell type, carcinogen metabolism, genetic background, exposure dose, and competing endogenous mutational processes.
- Many exposures are not linked to detectable mutational signatures, and many mutational signatures have an unknown cause or may be reflective of multiple exposures.
- Integration of mutational signatures with other data modalities may identify new etiological linkages or refine existing ones.
- Detection of mutational signatures in normal tissues of healthy individuals may complement studies in cancerous tissues to better characterize the evolution from healthy tissue to cancer. Studies across the lifetime are needed to understand the effects of exposure-induced mutations versus mutations in late-stage tumors.
- In-depth studies of the effects of promoters on histologically normal tissue may help to link initiation mutations induced by environmental agents to changes in the cellular microenvironments. Such studies can help develop interventions that prevent advancement of initiated tumors.

Session 2: Other Data Types as Signatures of Exposure in Cancer

- Omic data types including the epigenome, microbiome, proteome, and metabolome, as well as immune and inflammatory profiles, may be developed as molecular signatures of exposure in cancer.
- Biological responses to exposures can be transient and may limit the identification of molecular signatures.
- The role of epigenetic changes in driving cancer formation is not well understood. Many epigenetic changes are highly dynamic and specific epigenetic changes may depend on the developmental windows of the exposure.
- Systems biology and computational modeling approaches are needed for understanding the non-mutational processes of cancer development.
- Development of cell-free (liquid biopsy) profiles is needed for omic data types to identify non-mutational signatures of cancer risk across tissues.
- Development of longitudinal cohorts or serial sample resources with robust meta-data is needed for greater temporal resolution of environmental carcinogenic processes.
- Other omic data types, including epigenetics, may be applied to aging signatures that can be associated with cancer risk.

Session 3: Computational Challenges and Integrating Multi-Omics to Identify Signatures

- Better analysis methods are needed to understand mutational signatures and incorporate other data types with mutational signatures. Network approaches can incorporate multi-scale data, including genetic diversity, with respect to carcinogenesis.
- Network and nonlinear machine approaches to molecular signature analysis require further development. Deep-learning approaches have great potential but require significant amounts of data that are not readily available. Machine learning methods that are nonlinear or constrained by biological understanding may be more effective.
- Better data and analysis linking exposure data to molecular signatures are needed. Multi-omic profiles of exposures could utilize experimental models, population data, or single cell profiles of tissue. Time course data and normal tissue comparators are critical for identifying and interpreting molecular signatures.
- Development of multi-omic profiles of cells or tissues across cancer development (normal to malignancy) are necessary to understand how exposures affect cancer initiation and progression.
- Benchmarking of methods, validation of signatures, and standardized data would greatly help computational analysis and calibrate the performance of new computational methods.

Session 4: Challenges in Tracking Signatures of Exposures

- Continued identification of the cell types and model systems that best recapitulate human population effects and that are generalizable is needed. Experimental systems include mouse embryo fibroblasts, human induced pluripotent stem cells, human tissue organoids, mouse models and data derived from human populations.
- Better understanding of biomarkers of exposure is needed, particularly to determine that biomarkers are stable, dependent on life stage, and reflective of dose and timing.

- Continued study of biomarkers and their specificity, interactions, and related physical measures is needed. These biomarkers include circular RNAs, DNA methylation, DNA adducts, DNA mutation, metal isotopes, and protein adducts.
- Methods are needed to address chemical mixtures and the exposome, including laboratory methods that systematically test chemicals of interest and advanced statistical methods for mixtures analysis.

Session 5: Population-Based Cancer Studies

- More frequent and closer interactions between population scientists and experimental scientists would stimulate collaboration and accelerate research.
- Connect human studies to mechanisms using experimental models, such as organoids, to validate mutational signatures from human populations.
- Initiate new forms of cohorts, like mother-child cohorts, to look at early-life exposures and increase population diversity.
- Develop enhanced intermediate outcomes, like clonal expansion, and more accurate, scalable exposure assessments.

Overall Key Messages

- The associations between mutational signatures, exposures, and cancer have many complexities such as specificity, stability, and context dependency with no generalizable rules.
- Including multi-omic data as signatures of exposure in cancer is preferable to relying on single modality measurements.
- More frequent and closer interactions between population scientists and experimental scientists would stimulate collaboration and accelerate research.
- Implementing or improving data standards, datasets for calibration, and benchmarking of computational methods would aid computational studies that link molecular signatures of exposure to cancer.
- Focusing on a timely and tractable problem, such as an exposure window, mixtures, or longitudinal sampling, might be a useful model for applying multi-omic measures to identify molecular signatures.
- Multi-omic and exposure datasets are needed that are accessible, standardized, well-annotated, and represent longitudinal samples. Data sets from experimental models or human studies would be useful.

Introduction and Overview

On June 29 and 30, 2023, the National Institute of Environmental Health Sciences (NIEHS) and National Cancer Institute (NCI) convened a workshop on Molecular Signatures of Exposure in Cancer. The workshop was co-chaired by Cheryl Walker, Ph.D., of Baylor College of Medicine and Hannah Carter, Ph.D., of University of California, San Diego. The workshop was planned by members of the NCI/NIEHS Cancer and the Environment Working Group. Because the mining of genomic data for mutational and epigenetic signatures has been successful in identifying links between environmental exposures with cancer, the workshop aimed to inform how research in mutational and epigenetic signatures can be further advanced to inform cancer etiology, prevention, and population studies. The workshop brought together a multidisciplinary group, including environmental scientists, epidemiologists, toxicologists, physicians, and bioinformatics researchers (See Appendix 1 for the full list of participants and biographies). Presentations and moderated discussions highlighted challenges and opportunities related to defining and using molecular signatures of exposure in cancer. The full workshop agenda is included in Appendix 2. Key publications are available in Appendix 3.

The workshop was organized into six main sessions over two days. Each session included a *State of the Science* speaker to give a broad overview of the session theme, three *Synopsis* speakers to apply the session theme to a particular research project or idea, and an extended panel discussion to address key questions on the session topic. A session moderator led a brief question-and-answer period following the *State of the Science* presentation and guided the panel discussion, which included all session speakers.

The session titles for each day are indicated below.

Day 1

- Session I: Mutational signatures of exposure in cancer
- Session II: Other data types as signatures of exposure in cancer
- Session III: Computational challenges and integrating multi-omics to identify signatures

Day 2

- Session IV: Challenges in tracking signatures of exposures
- Session V: Population-based cancer studies
- Session VI: Workshop summary and future directions

Trevor Archer, Ph.D., Deputy Director of NIEHS, and Dan Gallahan, Ph.D., Director of the Division of Cancer Biology at NCI, gave opening remarks to begin the workshop.

Archer explained that the workshop would build upon the key takeaways from a previous workshop held in February 2023, titled Integrating Environmental Data with Other Omics for Cancer Epidemiology. That workshop discussed how to incorporate environmental exposures into genetic and other omics data, with emphasis on study design, data harmonization, and reproducibility. This workshop was an opportunity to dig deeper into omics studies that are characterizing signatures of exposure in cancer. Participants described current efforts to identify and apply signatures for cancer prevention studies and identified gaps and needs. The importance of considering existing and potentially new data streams was emphasized.

Gallahan discussed the cancer research community's role in finding mutations and molecular profiles related to cancer through programs like The Cancer Genome Atlas (TCGA) and the Human Tumor Atlas Network (HTAN), including both germline and somatic level targets and mutations that occur in tumors.

The speakers selected for this workshop included computational biologists, epidemiologists, exposure scientists, and other experts in cancer research who could field a range of questions.

Following opening remarks, Carter and Walker expressed their thanks for the speakers and excitement for the workshop. The purpose and intended outcomes of the workshop were presented by members of the planning committee, Ron Johnson, Ph.D., of NCI, and Dan Shaughnessy, Ph.D., of NIEHS. The workshop goals were to define the current state of the science of using molecular signatures to link environmental exposures to cancer; prioritize pressing needs and opportunities that, if addressed and pursued, will aid progress in identifying molecular signatures of exposure in cancer; and explore ways to use molecular signatures of exposure to improve cancer prevention.

Throughout the sessions, the workshop sought to answer these key questions:

- What are the major gaps in, and opportunities for, identifying molecular signatures of exposure in cancer?
- What resources and approaches are needed to accelerate this research area in the next five years?
- How can molecular signatures of exposure be applied to improve cancer prevention?

The remainder of this report summarizes speaker presentations and accompanying discussion, as well as key takeaways and suggested future directions for NCI and NIEHS.

Session 1: Mutational Signatures of Exposure in Cancer

Session 1 was moderated by Ludmil Alexandrov, Ph.D., of the University of California, San Diego. The session discussed opportunities, caveats, and challenges related to using mutational signatures of exposure in cancer, mutational signatures to identify novel causes of cancer, and alternative models of cancer initiation and progression. The session aimed to critically discuss the state of the science on the specificity of signatures, the types of exposures associated with known signatures, whether nongenotoxic exposures elicit mutational signatures, and association of early signatures with stages of cancer.

State of the science: Emerging opportunities and caveats of using mutational signatures of environmental exposures

Serena Nik-Zainal, Ph.D., University of Cambridge

Mutational signatures are readouts of mutagenic processes that have occurred through a tissue's or tumor's history and are a product of a combination of DNA damage and DNA repair. Some mutational signatures result from endogenous, or intrinsic processes, such as deamination of methyl-cytosine or dysregulation of APOBEC (Apolipoprotein B mRNA Editing Catalytic Polypeptide-like) enzymes. Other signatures stem from exogenous, or environmental, mutagens, such as UV light, tobacco smoke, or

aristolochic acid (AA). Each one could leave a characteristic imprint on the genome, regardless of whether it is caused by an endogenous or exogenous source. Sequencing the genome gives a composite of these different processes. Mutational signatures are also a function of the magnitude (dose, duration) of the exposure. Many algorithms have been developed to identify signatures indicative of these exposures. Ultimately, mutational signatures can help researchers identify cancer risk and develop early interventions.

Learning how mutations arise in a controlled environment is a critical part of identifying mutational signatures arising from environmental mutagens. When studying cancer in cellular model systems, such as human, yeast, mouse embryo fibroblasts (MEFs), and in vivo models like *Caenorhabditis elegans* and mice, it is important to consider:

- A single environmental mutagen can cause multiple mutational patterns.
- Many environmental mutagens are a combination of compounds, not single agents.
- Some unrelated compounds can produce similar signatures.
- Environmental agents can cause mutagenesis through various mechanisms.
- Environmental agents can induce mutations as well as trigger other epigenetic alterations.
- Exposure does not necessarily indicate causality for cancer.
- Environmental agents can promote tumorigenesis without leaving a mutational signature.

Modern genomic sequencing technologies can track mutational signatures of environmental carcinogens in human cells. Epidemiological observations coupled with systematic studies on mutational signatures of environmental carcinogens may uncover previously unknown mechanisms of mutagenesis and cancer.

Can we use mutational signatures to identify novel causes of cancer?

Paul Brennan, Ph.D., International Agency for Research on Cancer

Because epidemiology only explains about 40% of the burden of cancer, various initiatives have explored international differences in cancer incidences to uncover unknown causes of cancer. One such example is the Mutographs[13] study, which investigated 5,000 cancers representing five cancer types across five continents with high- and low-risk areas for each cancer type. Interestingly, esophageal squamous cell carcinoma and renal cell carcinoma (RCC) revealed different results.

A case study[14] of 552 esophageal squamous cell carcinomas sequenced across eight high- and low risk countries revealed no differences in mutational signatures. For example, cancer incidence was 20-fold

[13] https://www.mutographs.org/

[14] Moody S, Senkin S, Islam SMA, Wang J, Nasrollahzadeh D, Cortez Cardoso Penha R, Fitzgerald S, Bergstrom EN, Atkins J, He Y, Khandekar A, Smith-Byrne K, Carreira C, Gaborieau V, Latimer C, Thomas E, Abnizova I, Bucciarelli PE, Jones D, Teague JW, Abedi-Ardekani B, Serra S, Scoazec JY, Saffar H, Azmoudeh-Ardalan F, Sotoudeh M, Nikmanesh A, Poustchi H, Niavarani A, Gharavi S, Eden M, Richman P, Campos LS, Fitzgerald RC, Ribeiro LF, Soares-Lima SC, Dzamalala C, Mmbaga BT, Shibata T, Menya D, Goldstein AM, Hu N, Malekzadeh R, Fazel A, McCormack V, McKay J, Perdomo S, Scelo G, Chanudet E, Humphreys L, Alexandrov LB, Brennan P, Stratton MR. Mutational signatures in esophageal squamous cell carcinoma from eight countries with varying incidence. Nat Genet. 2021 Nov;53(11):1553-1563. doi: 10.1038/s41588-021-00928-6.

higher in China than in the United Kingdom despite no difference in mutational burden. The researchers also found no mutational signatures associated with known or suspected causes of esophageal squamous cell carcinomas, such as hot drinks, indoor air pollution, and poor diet. The team observed some associations between environmental or lifestyle exposures and mutational signatures in U.K., Japan, and Brazil between alcohol consumption and SBS16 signature, and in Iran between opium usage and SBS288J signature.

By contrast, a study investigating 954 cases of renal cell carcinoma from 11 high- and low-risk countries showed differences in mutational signatures. For example, signature 12 — previously identified in some liver cancers — appeared in about 75% of the RCC cases from Japan but was largely absent elsewhere, which could be indicative of an exposure only present in Japan. As another example, Aristolochic acid (AA) signature SBS22 was present in almost all cases in Romania, about half of cases in Serbia, a few in Thailand, and almost none elsewhere. A novel signature, SBS40b, was found to be strongly associated with the incidence of RCCs across all countries.

These two examples highlight how mutational signatures may explain international differences in incidence for some cancers but not others. They also illustrate that many known or suspected causes of cancer may affect non-mutagenic pathways. Emerging evidence suggests that certain mutagens are nearly ubiquitous in some populations but largely absent elsewhere, and traditional epidemiology has been poor at detecting these exposures.

Mutational signatures of exposures: examples from lung cancer studies

Maria Teresa Landi, M.D., Ph.D., National Cancer Institute

The mere presence of a mutational signature related to an exposure within a tumor tissue does not imply or confirm causation. For example, SBS22 has been identified as a signature for AA exposure in multiple cancer types, including lung cancer. In a study of never-smokers in Taiwan, researchers found that mutations associated with SBS22 were mostly clonal, suggesting the exposure occurred early in tumor evolution. However, the team found no evidence of these types of mutations associated with AA in the major cancer driver genes, and SBS22 is also found in normal tissue. These observations raise the question of whether SBS22 is a signature of exposure or if it is associated with causality.

In another study with smokers and non-smokers, there was no association between the dose of tobacco smoking and tobacco smoking signatures in lung cancer. This observation could suggest a saturation effect beyond which the signature cannot capture dose, potentially requiring large sample sizes to capture minimal differences in exposures. Further, when investigating subjects with chronic, episodic, old, and recent exposures, the researchers found no major differences in the signatures. They also found no signature of second-hand tobacco smoking in never-smokers. Simulation studies estimated that at least 10% of the mutations associated with tobacco exposure per sample would be needed to be detected by the algorithms at that given sample size. The idea that current algorithms may miss subtle mutations associated with cancer risk is an important consideration for cancer prevention.

A third study involved two groups of heavy smokers, which were mostly identical in terms of sex, age, and cancer stage type, but differed in their presence of the APOBEC signature. In those with no or low APOBEC signatures, tobacco mutational signatures SBS4, DBS2 and ID3 were clearly identified, including epigenetic changes. However, in those with a strong APOBEC signature, no signatures of tobacco

smoking were seen, including epigenetically. This finding is likely due to altered cell state compositions in the tumors due to the mutations. Determining the effect of the co-occurrence of multiple mutational processes and genetic backgrounds is another challenge.

Based on these three studies, actionable steps to consider for future studies include:

- Use cohort studies that include pre-diagnostic samples.
- Time the occurrence of mutational signatures in pre-cancer/tumor evolution.
- Integrate other omics, cell of origin, cell state, and subjects' genetic background.
- Improve algorithm sensitivity.
- Use biospecimens linked to detailed exposure information/exposomes.
- Use large sample size/single repository of signatures.
- Test different doses and mixtures of agents in experimental models.

Alternative models of cancer initiation and progression

Allan Balmain, Ph.D., FRS, University of California, San Francisco

A classic model of cancer initiation and progression is based on the sequential accumulation of mutations, which lead to an increased mutational burden over time that in turn leads to sequential genetic activation of cancer hallmarks. An alternative model includes two major stages: initiation of cancer and promotion.[15] When researchers compared these two models in mice, the classic model showed that most environmental factors tested resulted in few mutations and did not cause any obvious signatures. In the initiation-promotion model, a single mutagen exposure resulted in many mutations. This finding emphasizes that the number of mutagens in the models did not matter and that a single mutagen, when combined with a promoting factor, was enough to lead to the development of cancer.

In a separate study that used deep sequencing to detect signatures, mice received one single exposure of DMBA. Two weeks later, researchers measured mutational signatures and found a DMBA signature. They also measured mutational signatures one year later, and the DMBA signature persisted. However, the mice did not develop cancer because they were not exposed to a promoter. The results show that mutational signatures can persist over time but not result in a tumor.

In another assay, researchers used error-corrected duplex sequencing to find specific driver mutations. Mice treated with DMBA showed only four mutations in the canonical oncogenes, Nras and Hras. However, when mice were treated with DMBA and exposed to a tumor promoting agent for four weeks, a large clonal expansion of cells carrying mutations occurred, with tumors appearing weeks and months later. This finding again indicates that the rate-limiting step is exposure to the promoting factor, not to the initiating factor.

Future directions for the field to begin documenting tumor promoting factors and how they operate to change the clonal architecture include:

- Detection: development of short-term in vitro/in vivo assay for detection of promoter activity and clonal expansion in different tissue types.

[15] Balmain A. Peto's paradox revisited: black box vs mechanistic approaches to understanding the roles of mutations and promoting factors in cancer. Eur J Epidemiol. 2022. doi: 10.1007/s10654-022-00933-x.

- Understanding: large scale single cell RNAseq and epigenetic analysis of the dynamic impact of promoters on normal tissues. This work could investigate anti-promoters and non-toxic dietary or other factors that could be used to first understand, and then implement, cancer prevention strategies targeting the promotion stage.
- Prevention: development of approaches to identification and elimination of potentially dangerous clones in normal tissues, using small molecules or activators of innate of adaptive immune systems

Discussion

Key discussion points from Session 1 included:

Genetic background effect on signatures. Research in mice shows that genetic background clearly has an influence, not necessarily on mutational signatures, but on driver mutations found in tumors. The prevalence of a driver mutation can change dramatically between mouse strains, and there is a segregation of driver mutations across different populations. In lung cancer, an association has been found between germline retrotransposon insertions and tumor signatures, pointing to complex interactions between different factors. In some cases, genetic background and germline variants have a secondary impact, so determining what is happening on a molecular basis is difficult. For example, some people have a combination of cytochrome P450 enzymes that correlate with either fast or slow metabolism. That combination can affect metabolic activation or breakdown of chemical exposures into mutagenic compounds.

Mutational signature complemented with epigenetic factors. Not all agents that cause cancer leave distinct mutational signatures. Therefore, other types of signatures may be useful, such as promoting signatures or adductomic signatures. For example, smoking causes seven different cancers, but when researchers examined signature SBS4 in human tumors among smokers and nonsmokers, they found it only in the respiratory tract, not in other tissues. Other epigenetic factors may contribute to tumor promotion without causing DNA adducts or direct DNA damage characteristic of mutational signatures. Environmental agents cause tumors via chronic inflammation, tissue damage, and tissue regeneration.

Analyzing normal tissue. Growing evidence suggests that the cancer process starts at birth. If cancer is driven by an early-life environmental exposure, the associated mutational signature may not be found by looking at late-stage tumors that may be associated with a late-stage endogenous signature. For example, in esophageal tumors, the APOBEC signature — a late-stage phenomenon — dominates, perhaps obscuring important mutational signatures that occurred earlier in life. This concept suggests an opportunity for investigating normal tissues or placenta.

Embracing complexity of biological systems. Cancer is not caused by a single mutation or a single event, but multiple mutations and a combination of events. In other words, elements work together in complex networks that drive disease. These elements also work in combination with genetic background and exposure history, among other factors. Therefore, it is important to integrate multi-omic data to understand how these elements interact with each other.

Signatures for prevention. Because limited mutational signatures exist for predicting cancer, other modalities, like transcriptomic or epigenetic approaches, could potentially identify signatures that relate to other environmental factors that create tumor promotion conditions. For example, clonal architecture

can be used to distinguish between cancer risk among mice that share the same spectrum of mutations. Deep sequencing of normal tissue could reveal if certain clones expanded after the mice were exposed to a promoter. New research is investigating normal tissue clonal architecture in human samples to understand clonal evolution and develop potentially preventative approaches. A major question will be the degree to which agents influence specific clonal architecture.

Future directions. Speakers shared their hopes and predictions for the field in the next 10 years, which included integrating more data on the microenvironment and other model features; applying signatures for clinical value; determining molecular alterations in normal tissue before tumor development; studying mutational signatures in different populations and early life to reduce health disparities; reconstructing the process that occurs after a cell first interacts with a carcinogen to develop a more complete understanding of the carcinogenic process; using tools like deep sequencing, single cell sequencing, and artificial intelligence (AI); gathering more data and more tissue resources for a wider perspective.

Session 2: Other Data Types as Signatures of Exposure in Cancer

Session 2 was moderated by Scott Auerbach, Ph.D., of NIEHS. The session discussed non-mutational data types, such as epigenetic, metabolomic, and microbial changes, as signatures of exposure in cancer. The session aimed to answer questions related to the potential of non-mutational omic data types to identify signatures of tumor progression, whether certain omic data types are more applicable to certain exposures, and other emerging technologies could be used for molecular signatures of exposure.

State of the science: Other data types as signatures of exposure in cancer

Ting Wang, Ph.D., Washington University

Genetic and epigenetic mechanisms work together during tumorigenesis. The epigenome is at the boundary between the genome and the environment and mediates most of the environment's effects. Therefore, the epigenome and other cellular molecular profiles may provide an important opportunity to capture signatures of exposure during cancer evolution. These profiles include mutational signatures and other data related to genomics, metagenomics, epigenomics, transcriptomics, proteomics, metabolomics, and immune profiling.

Omics is a collection of methods to measure and functionally characterize different types of biomolecules in cells of tissues. The past 20 years have seen rapid evolution of sequencing, mass spectrometry, methods for collecting specimens, and single-cell technologies. These new approaches produce large volumes of data that require different skills and paradigms to integrate, analyze, and interpret. As a result, advances have occurred in bioinformatics, computational algorithms, tool development, and AI used to analyze multi-omic data sets. These developments critically affect how molecular signatures in cancer are defined, identified, and applied.

The impact of technology development is evident when looking at three important consortium projects that have created large data sets and that may provide a key resource for developing epigenetic measures of exposure: the NIH Roadmap Epigenome Project, Toxicant Exposures and Responses by Genomic and Epigenomic Regulators of Transcription (TaRGET), and the Human Pangenome Project.

The goal of the NIH Roadmap Epigenome Project[16] was to create the first reference epigenome. Although the project ended in 2017, epigenomic data continues to accumulate: Over a hundred thousand complete human epigenome datasets now exist. This reference epigenome has enabled comparisons with the epigenomes of diseased tissue.

The goal of the TaRGET consortium[17] is to define a standard approach to investigate, in a systematic manner, how environmental exposures affect the genome. First, given a very specific exposure, researchers investigate how the epigenome responds in a target tissue. Second, they determine whether epigenetic changes in the target tissue can be discerned by studying a surrogate tissue. In one study, consortium members used a mouse model to test several toxicants, including arsenic, BPA, and PM2.5, and generated several thousand epigenomic datasets, which will be made public. Although the intent was not to investigate connections with cancer, the team identified a cancer outcome and epigenetic signatures at early stages of exposure.

The Human Pangenome Project,[18] which launched in 2019, will build upon the success of the Human Genome Project by incorporating more diversity. The project will sequence 350 individuals' haplotype genomes, finish them telomere to telomere, and incorporate them into a new data structure that will represent diversity. This project presents a new opportunity to look at signatures of exposure and cancer in the context of different genetic backgrounds.

Can developmental window-specific molecular fingerprint/surrogates clarify links of environmental exposure and cancer risk?

Cathrine Hoyo, Ph.D., North Carolina State University

Over 80,000 chemicals have been identified as environmental contaminants, and approximately 2,000 more are added each year. Toxicity data is lacking for many of these chemicals, including 40% of "high production volume" substances. Beyond chemical toxicity, where people live affects risk. For example, the liver cancer risk in some contiguous North Carolina counties is about half the rate in other counties, but the reason is unclear. This discrepancy highlights the importance of using intermediate surrogates to link environmental exposures to cancer.

Social, chemical, and lifestyle stressors all shape a cumulative epigenetic "fingerprint," resulting in changes to the transcriptome, proteome, and metabolome. However, how epigenetic changes drive cancer is not well understood.

---

[16] Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS, Thomson JA. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010;28(10):1045-8. doi: 10.1038/nbt1010-1045.

[17] TaRGET: https://targetepigenomics.org/

[18] Wang T, Antonacci-Fulton L, Howe K, Lawson HA, Lucas JK, Phillippy AM, Popejoy AB, Asri M, Carson C, Chaisson MJP, Chang X, Cook-Deegan R, Felsenfeld AL, Fulton RS, Garrison EP, Garrison NA, Graves-Lindsay TA, Ji H, Kenny EE, Koenig BA, Li D, Marschall T, McMichael JF, Novak AM, Purushotham D, Schneider VA, Schultz BI, Smith MW, Sofia HJ, Weissman T, Flicek P, Li H, Miga KH, Paten B, Jarvis ED, Hall IM, Eichler EE, Haussler D; Human Pangenome Reference Consortium. The Human Pangenome Project: a global resource to map genomic diversity. Nature. 2022;604(7906):437-446. doi: 10.1038/s41586-022-04601-8.

The penetrance of epigenetic changes is variable due to host genetics, exposure dose, window of susceptibility, and other environmental factors. These factors should be measured in the context of features such as chromatin and proteome changes that are associated with epigenetic changes. Additionally, more empirical data from exposed populations could inform efforts to model the effects of epigenetic changes on cancer risk and establish a better understanding of epistatic events that occur between the genome and environment. Certain attributes of DNA methylation, including its responsiveness to environmental stressors (archives of exposure), can be leveraged for exposure assessment, using high throughput sequencing and analytic methods (arrays). This approach could be applied to DNA demethylation and activation of genomic transposable elements, which are sensitive to environmental perturbations and have been shown to affect carcinogenic cell pathways. However, some shortcomings of the approach include reliance on published arrays, which are limited in genome coverage, and case-control comparison feasibility, which depend on accessible tissues like blood and saliva.

There is a need to leverage what is known about the genome and to develop a semblance of whole genome responsive elements. As an example, one study investigated DNA methylation throughout the genome in search of chromosomal regions that reflect genomic imprinting called imprint control regions (ICRs). The team identified 1,488 ICRs, representing more than 22,000 CpG sites — areas in DNA where a cytosine nucleotide occurs next to a guanine nucleotide. Methylation marks at ICRs were established before tissue specification in response to environmental exposures and were similar across tissues and stable over the life course. Using GeneHub, the team found 28 ICUs related to all cancers investigated.[19] Future studies using ICR array data can investigate whether the contributions of early exposures to cancer can be quantified and whether other developmental windows of susceptibility exist.

Integration of cell-free DNA epigenetic and antibody signals as signatures of exposure in Epstein-Barr virus-associated cancers

Ben Gewurz, M.D., Ph.D., Harvard University

Epstein-Barr virus (EBV) was the first human virus discovered to cause cancer. A particular challenge with EBV is that 95% of adults carry the virus, but only 200,000 cases of associated cancer occur per year, typically in specific populations.

EBV was discovered about 60 years ago in sub-Saharan Africa, where a high incidence of Burkitt lymphoma occurred. Shortly after, researchers investigated a possible association between the virus and cancer. To determine if a relationship existed between antibody response to the virus and the onset of cancer, the team administered a survey of 45,000 children enrolled from birth, which showed elevated antibodies to a viral antigen that held true at the time of diagnosis.[20] A few year after the conclusion of

[19] Jima DD, Skaar DA, Planchart A, Motsinger-Reif A, Cevik SE, Park SS, Cowley M, Wright F, House J, Liu A, Jirtle RL, Hoyo C. Genomic map of candidate human imprint control regions: the imprintome. Epigenetics. 2022;17(13):1920-1943. doi: 10.1080/15592294.2022.2091815.
[20] Burkitt D. A children's cancer dependent on climatic factors. Nature. 1962 Apr 21;194:232-4. doi: 10.1038/194232a0.

the survey, an association between EBV and nasopharyngeal carcinoma was discovered. Antibody surveys again showed a higher antibody response in people with nasopharyngeal carcinoma.[21]

Most recently, The Cancer Genome Atlas Program (TCGA) found that one out of four types of gastric cancer is associated with an EBV infection. In these cancers, the virus does not make new viruses, but rather exists as a latent genome inside the nucleus of cells. A salient feature of EBV-positive gastric cancer is the CIMP hypermethylation phenotype, the highest level of DNA methylation of any human cancer. Whether the virus is responsible for that phenotype or whether it stems from a whole cell response to shut down the virus is unclear. Regardless, the feature could potentially be used diagnostically to determine if somebody will likely develop cancer.

One diagnostic approach is to use liquid biopsy. As the infected cells are killed by surveilling immune cells or dying as the tumor is evolving, they release DNA into the circulation that can retain methylation marks from the original cells. Then, a technology called cell-free methylated DNA IP sequencing (cfMeDIP-seq), can immunoprecipitate the methylated DNA fragments and sequence them using next generation sequencing.[22] Another recently developed technology, VirScan, in which phages display a peptide from viruses such as EBV, can analyze as little as one microliter of serum and return a readout of antibody production. The results provide a high-resolution view of immune system response. Finally, next generation sequencing of T-cell receptors and machine learning can predict what T-cell receptors are responding to, such as viral or cancer antigens related to EBV.

Research on viruses and cancer should develop approaches that integrate risk factors, viral load, serology, cell-free DNA methylation, and T/B-cell receptor signals. For example, researchers could use a risk score to search for associations between human leukocyte antigen and EBV-positive cancer in different populations. Next, they could obtain small blood samples from volunteers to determine viral load, as well as apply high-resolution serology screens such as VirScan to understand a person's response at different points in time. If a high risk presents, researchers could apply methylation DNA precipitation sequencing (cfMeDIP-seq) and TCR/BCR-seq.

Chemical exposures from the microbiome

Michael Fischbach, Ph.D., Stanford University

Fischbach and colleagues developed a complex gut microbiome model system that was used to manipulate the levels of bile acids in colonized mice.[23] When they removed bile acid metabolite-

[21] Ji MF, Wang DK, Yu YL, Guo YQ, Liang JS, Cheng WM, Zong YS, Chan KH, Ng SP, Wei WI, Chua DT, Sham JS, Ng MH. Sustained elevation of Epstein-Barr virus antibody levels preceding clinical onset of nasopharyngeal carcinoma. Br J Cancer. 2007 Feb 26;96(4):623-30. doi: 10.1038/sj.bjc.6603609.

[22] Shen SY, Singhania R, Fehringer G, Chakravarthy A, Roehrl MHA, Chadwick D, Zuzarte PC, Borgida A, Wang TT, Li T, Kis O, Zhao Z, Spreafico A, Medina TDS, Wang Y, Roulois D, Ettayebi I, Chen Z, Chow S, Murphy T, Arruda A, O'Kane GM, Liu J, Mansour M, McPherson JD, O'Brien C, Leighl N, Bedard PL, Fleshner N, Liu G, Minden MD, Gallinger S, Goldenberg A, Pugh TJ, Hoffman MM, Bratman SV, Hung RJ, De Carvalho DD. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. Nature. 2018 Nov;563(7732):579-583. doi: 10.1038/s41586-018-0703-0.

[23] Cheng AG, Ho PY, Aranda-Díaz A, Jain S, Yu FB, Meng X, Wang M, Iakiviak M, Nagashima K, Zhao A, Murugkar P, Patil A, Atabakhsh K, Weakley A, Yan J, Brumbaugh AR, Higginbottom S, Dimas A, Shiver AL, Deutschbauer A, Neff N, Sonnenburg JL, Huang KC, Fischbach MA. Design, construction, and in vivo augmentation of a complex gut microbiome. Cell. 2022 Sep 15;185(19):3617-3636.e19. doi: 10.1016/j.cell.2022.08.003.

producing bacteria, *Clostridium hylemonae* DSM 15053 or *Clostridium scindens* ATCC 35704, from the mice, secondary bile acid production did not occur. The results suggested an opportunity to study how the bacteria individually affect secondary bile acids production.

To investigate whether either of the two bacteria alone would be sufficient to metabolize bile acids in the gut, two additional communities were built: one with *C. hylemonae* but not *C. scindens* (ΔCh), and one with *C. scindens* but not *C. hylemonae* (ΔCs). Both communities had identical levels of other bacteria. Mice colonized by either community displayed normal bile acids pools. However, mice containing the *C. scindens* community showed high levels of aromatic amino acid metabolites compared to other communities. The profile of aromatic amino acid metabolites also shifted dramatically.

Whether *C. hylemonae* or *C. scindens* was removed had little bearing on other bacterial levels. However, some organisms disappeared when either *C. hylemonae* or *C. scindens* were removed, and some organisms appeared from below the limit of detection. The findings indicate that the presence of certain bacteria is sufficient to dramatically change aromatic amino acid metabolite pool. The findings emphasized a need to determine microbiome exposures empirically, rather than by computational prediction. They also highlighted the potential influence of microbial ecology on drug/infection/chemical exposures.

Discussion

Key discussion points from Session 2 included:

Continuous monitoring. Leveraging a technology like continuous glucose monitoring could be a powerful way of measuring metabolites and uncovering physiologic changes in real time. Using technology like wristbands or other easy-to-wear accessories could provide continuous information about molecules. Future research could investigate whether this approach could be used to examine tissue changes over time.

Some data types are plastic. Mutational signatures are historical records of permanent genetic changes whereas some data types mentioned in this session are plastic, making it more challenging to capture their effect. Epigenetic marks are largely transient, except for some that are more stable depending on the developmental window and type of exposure. For example, viruses can cause changes to the methylome that persist. Regarding the metabolome, some metabolite levels change quickly, like those that respond to diet, but others are stable over time. Studying ephemeral and permanent data types such as these can provide insight into the mechanistic bases of disease.

Metabolomics. The type of model microbiome described by Fischbach,[23] could suggest causality between a specific molecule and the onset of cancer. For example, levels of the amino methionine, the methyl donor for methylation reactions, can alter the epigenome. For microbiome-derived molecules that are known to be difficult to control, the model could provide an opportunity to carry out mechanistically sound experiments.

Aging (non-chemical exposures). Studies examining the interaction between exposures and the epigenome have shown a clear association with biological aging, although the mechanism is unknown. As people age, their immune response wanes, which can alter the balance between tumor viruses and host. Certain epigenetic enzymes like DNA methylases have been implicated in heart diseases. Epigenetic

clocks are well measured using DNA methylation but not for histone marks. Another future direction for aging studies could examine the tissue of tumor origin, rather than tumors, to determine whether early-life environmental exposures affect the aging trajectory. The aging trajectory could also be factored into risk analysis by determining whether a person is on or off the aging trajectory.

Session 3: Computational Challenges and Integrating Multi-Omics to Identify Signatures

Session 3 was moderated by Mona Singh, Ph.D., of Princeton University. The session discussed the use of gene regulatory networks to embrace biological complexity; mutational signatures as a composite of DNA damage and repair mechanisms; the use of molecular signatures and single cell RNA sequencing to understand premalignant tumors; and computational challenges related to multi-omic integration. The session aimed to answer questions related to using different omics data types to identify signatures of exposure; what data types can be used to link signatures to genotoxic or nongenotoxic exposures; and using multi-omic signatures to improve cancer risk prediction models.

State of the Science: Why networks matter: embracing biological complexity

John Quackenbush, Ph.D., Harvard University

A large part of multi-omic and life-course data analysis is organizing the data for interpretation. The concept of regulatory networks embraces biological complexity, which is important when considering large-scale data. Although every cell has the same genome, different genes are expressed in different tissues and organ systems. With analysis of gene expression data, it is important to recognize that expression of genes is driven by regulatory networks. Different regulatory networks represent different biological states, and networks capture differences that occurred based on a variety of factors, including mutations that alter transcription factor binding or changes in epigenetic regulation that affect gene expression. These are important to understanding the effectors of genetic programs. Differences in the topology and structure of networks can be used to identify connections that are distinct to individual phenotypes. This concept can be combined with gene expression data to understand what drives the manifestation of biological states. Taken further, networks can be thought of as representing not just individual physical states, but individual biomarkers for exploring health and disease. Overall, the structure of a network informs the understanding of the biology of the system being studied and networks in each tissue, in each biological state, in each individual are unique.

There are different ways to represent networks, all of which provide insight into a disease state. Some, such as DRAGON,[24] PANDA,[25] EGRET,[26] LIONESS,[27] are more strongly associated with gene regulatory networks.

PANDA uses a network fusion approach to capture information about the way in which gene regulation occurs. For example, a transcription factor might regulate three genes, but if two of those genes are co-expressed, the transcription factor likely co-regulates those genes. A network fusion approach is based on a technique called message passing, focused on protein-protein interactions, correlation data, and expression data, to establish an initial transcriptional network. By examining three starting networks — cooperative networks, regulatory networks, and co-regulatory networks — the system captures transcription factor interactions in the initial network, in an iterative process that arrives at the final network.

An extension of this method that can be used with almost any gene regulatory network is a method called LIONESS, which allows for individual networks to be extracted from population data. It works by establishing a network representing contributions from many samples, i.e., network $e^{(a)}$. Then one sample, q, is removed from the network to give network $e^{(a-q)}$. Network $e^{(a)}$- network $e^{(a-q)}$ is scaled by number of samples, and network $e^{(a-q)}$ is added back to give sample q's network: $N_s(e^{(a)} - e^{(a-q)}) + e^{(a-q)} = e^{(q)}$. This process can then be done for a population to obtain individual networks for each sample and can be treated as a way of linking structural differences in networks with diseases and disease states. For example, when investigating sex differences in colon cancer, it was shown that male and female networks have fundamental differences in the structure of the regulatory networks that were predictive of clinical endpoints, including response to chemotherapy. Among females, those females who had networks that more closely resembled male networks responded more like males to chemotherapy treatment.[28] Another study investigated what different types of networks reveal compared to analyzing genomic data alone. Using data from The Cancer Genome Atlas and other large-scale pancreatic studies, Quackenbush and colleagues focused on two subtypes of pancreatic cancer, basal and classical, with respect to differentially expressed, co-expressed, or differentially targeting genes to understand what the different data types revealed about the disease. Gene expression data showed differences between subtypes, but none that were informative about the nature of the disease. The co-expression data showed some processes that linked back to disease development and progression, and some other processes that could not be seen otherwise, such as autocrine signaling. Gene regulatory networks were then built that linked co-expressed genes and their regulators. When looking at differential patterns of

[24] Shutta KH, Weighill D, Burkholz R, Guebila MB, DeMeo DL, Zacharias HU, Quackenbush J, Altenbuchinger M. DRAGON: Determining Regulatory Associations using Graphical models on multi-Omic Networks. Nucleic Acids Res. 2023 Feb 22;51(3):e15. doi: 10.1093/nar/gkac1157..
[25] Glass K, Huttenhower C, Quackenbush J, Yuan GC. Passing messages between biological networks to refine predicted interactions. PLoS One. 2013 May 31;8(5):e64832. doi: 10.1371/journal.pone.0064832..
[26] Weighill D, Ben Guebila M, Glass K, Quackenbush J, Platig J. Predicting genotype-specific gene regulatory networks. Genome Res. 2022 Mar;32(3):524-533. doi: 10.1101/gr.275107.120.
[27] Kuijjer ML, Hsieh PH, Quackenbush J, Glass K. lionessR: single sample network inference in R. BMC Cancer. 2019 Oct 25;19(1):1003. doi: 10.1186/s12885-019-6235-7.
[28] Lopes-Ramos CM, Kuijjer ML, Ogino S, Fuchs CS, DeMeo DL, Glass K, Quackenbush J. Gene Regulatory Network Analysis Identifies Sex-Linked Differences in Colon Cancer Drug Metabolism. Cancer Res. 2018 Oct 1;78(19):5538-5547. doi: 10.1158/0008-5472.CAN-18-0454.

regulation, there was an additional subset of immune-related, epigenetic, and cell cycle processes that were only found using the regulatory network approach to synthesize different data types.[29]

Another method that can be used to integrate other types of data is EGRET,[26] which accounts for how individual genetic backgrounds influence the patterns of expression and regulation that occur in individual networks. It starts with the same basic PANDA model, but for each individual in the data set, the network prior is modified based on data from expression quantitative trait loci (eQTL) studies about how genetic variance might perturb the binding of specific transcription factors. This method was used to show that genetic variants long associated with cancer risk were those that were associated with the regulation in transcription factor binding of oncogenes and tumor suppressor genes. This result showed the risk of developing disease is linked back to the genetic background individuals carry and indicates how those potentially mutated genes are regulated. This method can also be used to understand response to drugs and potential therapies. A database called GRAND,[30] containing more than 200,000 regulatory networks — including data from studies that have investigated drug treatments — has been used to identify new drugs and drug candidates by altering network structures.

Overall, network methods have the potential to allow a principled approach to multi-omic analysis. By investigating changes in network structure, drivers of disease and potential therapeutic targets can be identified. In numerous applications, networks have been able to provide insight into disease that is not found using expression or co-expression data alone.

Mutational signatures as a composite effect of DNA damage, repair, and other cellular processes

Teresa Przytycka, Ph.D. National Library of Medicine (NLM)

Mutational signatures are the result of a complex process that starts with DNA damage followed by DNA repair. What is observed as a mutational signature is a superposition of the two processes. However, in cancer, the DNA repair machinery is often compromised, so there may be a superposition of the same mutagenic processes with a somewhat altered DNA repair machinery, which will result in different mutational signatures. Researchers have looked at whether there is a different way to model mutational signatures to capture this non-linearity.

A recently proposed model, RepairSig,[31] approaches mutagenic processes as one of two types: primary, or those which are caused by the environment or chemical reactions and are reasonably approximated by an additive model, and secondary, or those related to DNA repair or some other DNA repair deficiency. The proposed total model is a composite model where primary processes are modified by the secondary processes, $M_{total}=M_{primary}*DQ$ (D= exposures, Q= repair signatures).

---

[29] Weighill D, Ben Guebila M, Glass K, Platig J, Yeh JJ, Quackenbush J. Gene Targeting in Disease Networks. Front Genet. 2021 Apr 23;12:649942. doi: 10.3389/fgene.2021.649942.
[30] Ben Guebila M, Lopes-Ramos CM, Weighill D, Sonawane AR, Burkholz R, Shamsaei B, Platig J, Glass K, Kuijjer ML, Quackenbush J. GRAND: a database of gene regulatory network models across human conditions. Nucleic Acids Res. 2022 Jan 7;50(D1):D610-D621. doi: 10.1093/nar/gkab778.
[31] Wojtowicz D, Hoinka J, Amgalan B, Kim YA, Przytycka TM. RepairSig: Deconvolution of DNA damage and repair contributions to the mutational landscape of cancer. Cell Syst. 2021 Oct 20;12(10):994-1003.e4. doi: 10.1016/j.cels.2021.07.004.

To better characterize mutational signatures, methods are needed to infer repair processes active in individual signatures. For example, cancer treatments are designed to induce cell death by DNA damage. However, tumor cells can initiate or adapt DNA repair pathways to resist these anticancer agents during chemotherapy. This presents an opportunity to uncover potential epistasis with chemotherapy drugs. There is also a need for new methods to examine relationships between signatures. This is currently determined by showing a signature has a high cosine similarity to another signature. Alternatively, a new approach, RePrint, can attempt to capture similarities between common repair mechanisms.[32]

An important step to understanding mechanisms of mutational signatures is to combine the mutational signatures with other types of data, like gene expression data. Learning how a mutational signature is related to changes in gene expression may help uncover the processes behind signatures. To that end, two complementary methods were developed: EcoSig[33] and NetSig.[34] EcoSig is based on clustering; understanding mutational signatures present in a cancer; and correlation profiles of all genes, from clusters to the signatures, including clock-like signatures and cell cycle signatures. This approach allows for identification of the groups of genes that are related to the signatures. The NetSig model is a network that contains signatures and gene expression, and statistical methods are used to understand the information flow between genes and signatures.

Understanding molecular signatures of response to exposure and development of premalignant lesions

Joshua Campbell, Ph.D., Boston University

Understanding early development of lung cancer, such as identifying key molecular changes that happen at the premalignant stage, can be useful for devising ways to intercept cancers before it progresses to an invasive lesion. This approach could be applied to understanding the effect of cigarette smoke exposure on molecular changes in the airway epithelium, for example.

Lung squamous cell carcinoma is strongly associated with smoke exposure, and the cell of origin is thought to come from the central airway. The airway is a composite of many different cell types — basal, club, goblet, ciliary — that all work together for mucociliary clearance. Researchers performed single cell RNAseq on six never-smokers and six current smokers and clustered data to identify the major cell types.[35] The data revealed three key findings. First, a clear transition occurred from club cells, predominant in the non-smokers, to a goblet-like phenotype in the smokers. Second, some detoxification genes were more enriched in particular cell types. For instance, an aldo-keto reductase (AKR1B10) was

[32] Wojtowicz D, Leiserson MDM, Sharan R, Przytycka TM. DNA Repair Footprint Uncovers Contribution of DNA Repair Mechanism to Mutational Signatures. Pac Symp Biocomput. 2020;25:262-273.
[33] Kim YA, Wojtowicz D, Sarto Basso R, Sason I, Robinson W, Hochbaum DS, Leiserson MDM, Sharan R, Vadin F, Przytycka TM. Network-based approaches elucidate differences within APOBEC and clock-like signatures in breast cancer. Genome Med. 2020 May 29;12(1):52. doi: 10.1186/s13073-020-00745-2.
[34] Kim YA, Hodzic E, Amgalan B, Saslafsky A, Wojtowicz D, Przytycka TM. Mutational Signatures as Sensors of Environmental Exposures: Analysis of Smoking-Induced Lung Tissue Remodeling. Biomolecules. 2022 Sep 27;12(10):1384. doi: 10.3390/biom12101384.
[35] Duclos GE, Teixeira VH, Autissier P, Gesthalter YB, Reinders-Luinge MA, Terrano R, Dumas YM, Liu G, Mazzilli SA, Brandsma CA, van den Berge M, Janes SM, Timens W, Lenburg ME, Spira A, Campbell JD, Beane J. Characterizing smoking-induced transcriptional heterogeneity in the human bronchial epithelium at single-cell resolution. Sci Adv. 2019 Dec 11;5(12):eaaw3413. doi: 10.1126/sciadv.aaw3413.

upregulated and enriched in ciliary cells. Third, a novel peri-goblet intermediate cell state was found, which was halfway between a basal and goblet cell. This intermediate did not express any distinct cell markers but was a hybrid between the basal and goblet populations and represented a transitional state that was enriched in the smokers. This study is an example of how single cell RNAseq can be used to understand the effects of cigarette smoke or other exposures on complex tissues.

In another study, biopsies of the abnormal premalignant lesions in the airway were profiled across different timepoints to understand the major drivers of the premalignant lesions. When performing DNA sequencing, the top driver was *NOTCH1*, which was enriched in dysplastic lesions compared to normal-appearing lesions. Bulk RNAseq was performed on the same set of samples to correlate genes with *NOTCH1* status, and clear signatures associated with *NOTCH1* mutation status were found. To understand what the signatures meant, the researchers applied a standard functional enrichment for particular cell types. Among the genes upregulated in the *NOTCH1* mutants, normal cell type signatures were not upregulated. However, when the team used the peri-goblet intermediate state signature from smokers, they found a strong association with the *NOTCH1* mutant cell signatures. This finding suggests that *NOTCH1* mutants caused cells to be in intermediate state and produce a dysplastic-like phenotype.

Future opportunities in this area include improving algorithms and technologies to allow for somatic mutation calling in single-cell data. A Human Cell Exposure Atlas (similar to the Human Tumor Atlas Network[36]) would also be useful for defining cell states in the context of exposures.

Thoughts on computational challenges & integrating multi-omics to identify signatures

Mark Gerstein, Ph.D., Yale University

Signature decomposition is challenging, and researchers have attempted to deconvolute observed mutations into a linear combination of known multinomial mutation probability distributions. One such method, SigLASSO,[37] aimed to jointly optimize the sampling likelihood and use regularization, and to allow priors for soft-thresholding. Using a renal cancer dataset and investigating how it populated different signatures using both SigLASSO and another tool called deconstructSigs, researchers showed the tools produced very different results. The finding points to the issue of the lack of standards for these types of analyses.

Therefore, to move the field forward, a number of challenges must be addressed. First, there is a need to develop reliable datasets to use as benchmarks. Further, gold-standard benchmarks must be defined for evaluating methods related to signatures and to evaluate performance. Second, regarding omics data, another challenge lies in determining how to calibrate, scale, and normalize different types of data relative to each other and in relation to signatures so that they can be integrated and their relative performance can be analyzed. Finally, many methods infer signatures based on linear approaches. Deep learning architectures may have relevant applications, though gathering sufficient training data also poses a challenge.

[36] https://humantumoratlas.org/

[37] Li S, Crawford FW, Gerstein MB. Using sigLASSO to optimize cancer mutation signatures jointly with sampling likelihood. Nat Commun. 2020 Jul 17;11(1):3575. doi: 10.1038/s41467-020-17388-x.

Discussion

Key discussion points from Session 3 included:

Deep learning. When using multi-omic data and large datasets, there is a potential opportunity to take advantage of deep learning. However, challenges include the large datasets required to train the model and the difficulty interpreting models as they get bigger and more complex. Deep learning could be useful for predicting signatures, but if the goal is to learn about a specific signature, linear deconvolution methods may be more relevant. It will also be important to constrain AI models with knowledge about biological systems, so they reflect evolutionary constraints.

Data needs. One challenge when attempting to link signatures to exposures across studies is the integration of different data types from different samples. Having the same type of data for different samples across studies will be critical to answering questions on cancer progression and risk. There is also a need for normal data (i.e., from non-disease tissue) and data on large cohorts over time to understand the influence of epigenomics and genetics over a lifetime and their progression along the risk trajectory. Better human datasets are needed for network analysis to be able to encompass age span.

Challenges for transcriptomic and epigenomic signatures. For single cell RNAseq, challenges lie in having controls (unexposed people) and detailed exposure information. Also, limited clinical variables or quantitative measures of exposure exist for assessing correlation with transcriptomics and other omics signatures. When integrating different types of omics data, often different datasets reflect the same underlying transcriptional process, and integrating multiple data types can address underlying noise to uncover the stronger underlying biological signal. However, epigenetic reprogramming from early-life exposures can precede altered transcriptional profiles later in life, often through exposure to a second type of stressor, such as a high-fat diet. Exploiting known variables, such as sex differences, can also significantly contribute to developing network models for cancer risk and variable response to cancer therapies.

Time course data. This type of data could remove signal to noise issues. However, time series data from humans is close to impossible to gather. Using pseudo-time methods to order individuals along a progression is an alternative approach.

Standardization and benchmarking. To assess the performance of new computational and network approaches, datasets must be standardized, and benchmarks applied. Different data types have different error profiles, thus interpreting them in relation to each other when combining large datasets is challenging. The field should consider the best steps for validation and benchmarking.

Session 4: Challenges in Tracking Signatures of Exposures

Session 4 was moderated by Rebecca Fry, Ph.D., of the University of North Carolina at Chapel Hill. The session discussed inducing mutational signatures in various in vitro models, key questions and challenges related to signatures of exposure, and the use of metallomics for cancer diagnosis and prevention. The session aimed to answer questions related to tissues and specimens to target in searching for signatures; how long molecular signatures of exposure persist in carcinogenesis; if cell-free

markers can be used; whether signatures can quantify the exposure and reflect duration; and how acute and chronic exposures manifest as signatures.

State of the Science: Inducing mutational signatures by environmental carcinogens and chemotherapeutic agents in vitro: progress, prospects, and limitations

David Phillipps, Ph.D., King's College

There are a number of key questions related to tracking signatures of exposures in laboratory-based studies, such as the choice of cell type, if exposures are acute or chronic, and whether to use selection or cloning prior to examining signatures. Other questions should consider if signatures are unique or species/tissue-specific, are qualitative or quantitative, and if this approach can inform risk assessment. Three experimental systems that have been used to investigate some of these questions include mouse embryo fibroblasts (MEF), human-induced pluripotent stems cells (iPSC), and human tissue organoids.

For example, MEFs were used to study mutations on a single gene and the whole genome using humanized mice.[38] To investigate TP53 mutations, researchers treated MEFs in culture using various agents (benzopyrene, aristolochic acid, UV, etc.) and passaged them for several months. The MEFs eventually went through senescence crisis and emerged as immortalized cells, which were clones that contained TP53 mutations. The TP53 mutations roughly mirrored what was seen in human tumors exposed to the same agents. Whole genome sequencing of the clones revealed thousands of mutations specific to exposures, even though cells were only briefly exposed. While this system was effective given the significant number of mutations, the non-human cells and time intensity were drawbacks. Pluripotent stem cells (iPSCs) are an alternative option because they are cloneable, quasi-normal, have fairly unlimited growth potential, and have no selection process (i.e., they are not transformed).

In another study, researchers investigated the effect of 79 environmental agents on iPSCs.[39] Whole genome sequencing of subclones from cells treated at a concentration that resulted in 40-60% viability for 3-24 hours. Whole genome sequencing of the subclones revealed 41 mutagen-associated substitution signatures, six mutagen-associated double substitution signatures, eight mutation-associated Indel signatures, and mutation asymmetries along the genome topography. There was a variable mutational burden — some agents did not give more mutations than the controls, and therefore detecting any signatures from those was not possible. Some had high mutation counts, like polycyclic aromatic hydrocarbons (PAHs) and Nitro-PAHs, alkylating agents, and chemotherapeutic drugs. Overall, some of the results were expected as observed with MEFs, but some unexpected results occurred. Also, some dissimilar agents gave similar signatures. When the experimental signatures were compared against COSMIC signatures, which represent those in human tumors, there were similarities. For example, aristolochic acid (AA) had strong similarity with COSMIC signature 22.

[38] Nik-Zainal S, Kucab JE, Morganella S, Glodzik D, Alexandrov LB, Arlt VM, Weninger A, Hollstein M, Stratton MR, Phillips DH. The genome as a record of environmental exposure. Mutagenesis. 2015 Nov;30(6):763-70. doi: 10.1093/mutage/gev073.

[39] Kucab JE, Zou X, Morganella S, Joel M, Nanda AS, Nagy E, Gomez C, Degasperi A, Harris R, Jackson SP, Arlt VM, Phillips DH, Nik-Zainal S. A Compendium of Mutational Signatures of Environmental Agents. Cell. 2019 May 2;177(4):821-836.e16. doi: 10.1016/j.cell.2019.03.001.

A third system using human-derived organoids was assessed as part of the Mutographs[2] project. The initial protocol involved cloning organoids, treating them, and cloning again before performing whole genome sequencing, which was time intensive. However, with the emergence of duplex sequencing, which sequences both DNA strands and therefore removes artifacts that occur during PCR, the protocol could be modified to remove cloning and instead simply treat the organoid with a mutagen, expand for seven to 10 days, isolate the DNA, and perform duplex sequencing. This approach was successful, and many of the compounds were metabolically activated efficiently by the organoids, which has not been achievable with iPSCs. Compared to other methods, there were some similarities, such as benzo(a)pyrene signature, which was similar to that from the iPSCs and COSMIC signature 4. A possible way to distinguish similar signatures from different compounds is by examining double-substitutions and Indels. The researchers also found that some signatures did not change from one tissue to another. For example, AA displayed the same signature in gastric, colon, kidney, pancreas, and liver tissues. Temozolomide had a similar signature in gastric organoids and COSMIC signature 11, but was very different in iPSCs. This was also tested in an MGMT knockout mouse which showed the same signature, suggesting the differences between the systems could be related to DNA repair.

Investigating environmental carcinogens and chemotherapeutic agents in these three in vivo systems resulted in some key takeaways, including:

- Specific mutational signatures of carcinogens can be detected in in vitro experimental models.
- NanoSeq (duplex sequencing) identifies genome-wide mutations without the need for subcloning.
- Organoids are a useful, but expensive, in vitro model that enable long-term culture and testing of cells from primary tissues; they are more metabolically competent than iPSCs.
- Some agents produced signatures after chronic treatment.
- Initial findings indicate that signatures are not tissue-specific, but mutational signatures can be influenced by DNA repair.
- Mutational signatures from exogenous exposures may not always be apparent.

Key questions and challenges

John Essigmann, Ph.D., Massachusetts Institute of Technology

A typical mutational spectrum can pose the question: Why do these patterns look the way they do? Key questions and challenges related to tracking signatures of exposure include:

- The complexity of mutational spectra points to a need to provide chemical/biological explanations for the types of mutations induced and address why the mutational patterns are punctuated by hot and cold spots. The respective roles of adduct formation, adduct repair, and adduct mutagenic bypass in molding end-stage mutational spectra must be defined.
- Animals and cell-based models with defined DNA repair and replication defects are needed to recapitulate mutational patterns that occur in human cancers.
- Chemoprevention (e.g., agents such as sulforaphane) can be effective, but high-throughput systems to evaluate agents that will erase or mitigate mutational spectra are needed. However, there can be a downside to such agents, as some make tumor tissues resistant to chemotherapeutic agents.

- Dose-response is important, but working with the right doses given likely human exposures and the fact that dose-responses are often non-linear is a challenge. Further, high doses saturate DNA repair systems and people might never experience those doses in real life.
- Life-course matters. Young animals are more sensitive to DNA damage, and there is a need to develop methods to look at spectra as early as possible, even in utero.
- Simple "founder" mutational spectra by carcinogens mature into much more complex tumor spectra, so following mutational evolution longitudinally in animals is necessary. This approach will help assess whether normal-looking tissue surrounding the tumor retains the simple "founder" spectrum and if that tissue can be used to understand the history of all mutagenic exposures experienced by the animal.
- Actionable advice to move the field forward includes:
    - Emphasize dose-response studies to use doses that are environmentally realistic, which requires more accurate and sensitive sequencing tools.
    - Distinguish spectra from different mutagen exposures in mixtures.
    - Integrate mutational signature/spectra studies with other omics and share tissues across platforms.
    - Carry out proof of concept for liquid biopsy starting with something relatively simple (e.g., temozolomide-treated patients/animals).
    - Investigate the role of a damaged nucleotide pool as a source of mutational spectra/signatures (e.g., 8-oxoG).

Challenges in tracking signatures of exposure

Yvonne Fondufe-Mittendorf, Ph.D., Van Andel Institute

When examining how a toxicant drives disease, it is important to determine how to mimic human exposure, which is often to a mixture and not a single compound. The challenge with mixture models is determining how many toxicants to use, the chemical composition of each toxicant in the mixture, the dose of each toxicant, and the number of exposures.

Another challenge is differentiating the initiator, driver, and passenger mutations and the epigenetic mechanisms that might drive the disease process. For example, researchers built a model for inorganic arsenic carcinogenesis and examined cells exposed to arsenic over time. Cells became migratory in transwell migration assays (and reverted with time) and wound healing, more colonies were formed, and the cells became more invasive with the appearance of protrusions. When these were injected into mice, they formed tumors. This model could help to understand what biomarkers reflect each process.

A simple assay to test what biomarker to use would be helpful. Early-expressed circular RNAs remain circulating in the blood throughout the transformation process and are stable. Researchers also found a circular RNA that co-expresses with its linear mRNA in both acute and chronic conditions. Because circulating blood can be measured easily and circular RNAs appear early as cells are exposed to arsenic, they could be used as a potential biomarker.

Metallomics and cancer: opportunities for diagnosis, prevention and intervention

Ana Navas-Acien, M.D., Ph.D., Columbia University

Metals are a class of carcinogens that can provide opportunities for prevention and intervention and diagnostics. Metallomics is the collective characterization, quantification, and speciation of metal and metalloid molecules that translate into the structure, dynamics, and function of an organism or system. Advances in technology allow for the characterization of elements (Inductively coupled plasma mass spectrometry (ICPMS)), species (High performance liquid chromatography (HPLC)-ICPMS)), and isotopes (Multi-Collector (MC)-ICPMS)). The precise quantification of metal ions is a recent development and may complement information on metabolites, pharmacokinetics, pharmacodynamics, and toxicity to elucidate what is occurring in the presence of cancer. A proof-of-concept study for cancer diagnosis found that zinc isotopes in urine are related to pancreatic cancer, especially lighter zinc.[40]

An important consideration for using this metallomics approach in humans is how to access the target tissue. For example, multi-omics technology can be combined with measures of metals and DNA methylation to develop biomarkers of lead levels in bone. Using this approach had similar results to established measures of lead in the patella and tibia.[41] This approach could be used in the future to measure different carcinogens at the target tissue level and connect the new biomarkers with health endpoints.

Metallomics also factors into chemoprevention when considering whether chelation could prevent cancer through the elimination of divalent toxic metals. For example, cadmium, which has an extremely long half-life, could be chelated and eliminated with EDTA. Chelation therapy has already been shown to be beneficial for cardiovascular disease, and a replication trial is ongoing.[42] Applying this preventative approach to cancer, while provocative, could be promising.

Discussion

Key discussion points from Session 4 included:

Prioritizing measurements for molecular signature analysis. In considering which measurements are most critical for molecular signatures analysis, one approach is to focus on using already-available technologies and data, such as DNA methylation arrays, which are widely accessible through cohort studies. For example, when determining lead biomarkers, most studies do not have whole blood samples to measure lead, so researchers can consider using DNA methylation array data to develop biomarkers for lead levels. Distinguishing biomarkers (e.g., those associated with cancer initiation versus promotion

[40] Schilling K, Larner F, Saad A, Roberts R, Kocher HM, Blyuss O, Halliday AN, Crnogorac-Jurcevic T. Urine metallomics signature as an indicator of pancreatic cancer. Metallomics. 2020 May 27;12(5):752-757. doi: 10.1039/d0mt00061b.

[41] Colicino E, Just A, Kioumourtzoglou MA, Vokonas P, Cardenas A, Sparrow D, Weisskopf M, Nie LH, Hu H, Schwartz JD, Wright RO, Baccarelli AA. Blood DNA methylation biomarkers of cumulative lead exposure in adults. J Expo Sci Environ Epidemiol. 2021 Feb;31(1):108-116. doi: 10.1038/s41370-019-0183-9.

[42] Lamas GA, Anstrom KJ, Navas-Acien A, Boineau R, Kim H, Rosenberg Y, Stylianou M, Jones TLZ, Joubert BR, Santella RM, Escolar E, Aude YW, Fonseca V, Elliott T, Lewis EF, Farkouh ME, Nathan DM, Mon AC, Gosnell L, Newman JD, Mark DB; TACT2 Investigators. The trial to assess chelation therapy 2 (TACT2): Rationale and design. Am Heart J. 2022 Oct;252:1-11. doi: 10.1016/j.ahj.2022.05.013.

and maintenance) can be identified using model systems applied to human studies, which often reflect maintenance biomarkers.

Applying signatures to improve cancer prevention. Understanding the mechanisms for cancer prevention is important. For example, understanding what causes cancer as revealed through readouts like COSMIC signatures could lead to preventive measures. It is also important to consider biochemical pathways induced by preventive agents to ensure a growth advantage is not being provided to cancers that may be present. Linking exposures to signature(s) is needed to establish causality.

Signatures to quantify duration of exposure. DNA methylation-based biomarkers could reflect past exposures. Circular RNA biomarkers could also reflect duration of exposure because research has shown that increasing the dose of exposure increases circRNA expression. Testing in human samples to determine if the biomarker is present in blood samples and whether it is correlated to a specific exposure is a next step. DNA adducts could provide another way to monitor exposure.

Mixtures. Using a multi-omic approach in a standardized system, like organoids, to examine DNA and protein adducts and lesions, along with related gene expression, could be an approach to studying mixtures. A key issue with mixtures is that they are specific to a certain population, and signatures for a mixture in one population are difficult to generalize to another population. However, advancements in statistical methods may help address this challenge. There is interest among statisticians to develop better methods for how to transport findings from one population to another that may have different mixture compositions. Dose-response is not linear, which poses another challenge for replication across different populations with different doses.

Proteins. Investigating whether mutational patterns in a COSMIC array reveal immunogenic information about proteins after translation would be helpful. Proteins may be valuable because biobanks are full of plasma, and protein adducts are persistent and abundant.

Session 5: Population-Based Cancer Studies

Session 5 was moderated by Paul Brennan, Ph.D., of the International Agency Research on Cancer. The session discussed using signatures of exposure in epidemiologic studies, precision environmental health for cancer prevention, the environmental exposome in cohort cancer studies, and race and ethnicity as modifiers of exposure. The session aimed to answer questions related to how molecular signatures can be used to identify, quantify, and predict cancer-related environmental exposures in human populations; the potential for using signatures of exposure to screen populations for cancer risk assessment; the effect of genetic background variation on signatures; appropriate biological samples for assessment of signatures; and how timing affects signatures.

State of the Science: Using molecular signatures of exposures in environmental epidemiologic studies of cancer

Wei Zheng, M.D., Ph.D., Vanderbilt University

Throughout lifetime, humans are exposed to a large set of chemicals and their mixtures, but most environmental exposures have not been studied in relation to cancer risk. It is also unclear how low-level

exposures in the general environment may be related to cancer and if there is any synergistic effect of exposure to mixtures.

Low-income and racial/ethnic minority populations tend to live in heavily polluted and economically deprived communities, but they are underrepresented in epidemiologic studies.

Biomarkers to assess environmental exposures are limited. For many of these biomarkers, the internal dose, biologically effective dose, and early biologic effects in humans remain to be validated. Molecular signatures may measure the early biologic effect and altered structure/function. It is well established that environmental exposures can alter cellular functions and cause changes in DNA, lipids, and proteins, and these signatures may be detectable.

Similar approaches to the iPSC model could be used to discover molecular signatures using epigenomics, transcriptomics, proteomics, and metabolomics tools. The challenge is translating these findings to humans since these signatures may appear differently in cell culture models. In addition to in vitro models, molecular signatures can be studied directly in humans. Many CpG sites have been identified in relation to environmental exposures such as tobacco, alcohol, maternal diet, air pollutants, chemicals, drug use, BMI, and age. For example, an epigenome-wide association study (EWAS) of about 15,900 adults using DNA methylation identified 2,623 more CpG sites in current smokers compared to never-smokers and 185 more CpG sites in former smokers compared to never-smokers. Additionally, dose-response associations were identified for about 60% of the CpGs with pack-years smoked. Thirty-six CpGs did not return to never-smoker levels, even after 30 years of smoking cessation.[43] These types of association studies are isolated, and with the exception of studies focused on aging and smoking, few have put these signatures together to build a model to assess exposure.

EWASs can also be used for the identification of endogenous metabolites associated with environmental exposures, but the key is the need to assess external exposures as well. For example, the Deep Exposome Project, part of the Southern Environmental Health Study, assesses the external exposome using polydimethylsiloxane (PDMS) wristbands — which can measure about 4,000 signals by gas chromatography-high resolution mass spectrometry (GC-HRMS) — as well as surveys and geographic information system (GIS) data. Using metabolomics of blood and urine, researchers assess external exposome components that have infiltrated the blood, as well as endogenous metabolites formed in response to the external exposures. The internal exposome is also assessed via the epigenome with about 850,000 CpGs by BeadChip and via inflammation biomarkers. The goal is to assess both the external and internal exposome, identify biomarkers/molecular signatures for external exposures and

[43] Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, Guan W, Xu T, Elks CE, Aslibekyan S, Moreno-Macias H, Smith JA, Brody JA, Dhingra R, Yousefi P, Pankow JS, Kunze S, Shah SH, McRae AF, Lohman K, Sha J, Absher DM, Ferrucci L, Zhao W, Demerath EW, Bressler J, Grove ML, Huan T, Liu C, Mendelson MM, Yao C, Kiel DP, Peters A, Wang-Sattler R, Visscher PM, Wray NR, Starr JM, Ding J, Rodriguez CJ, Wareham NJ, Irvin MR, Zhi D, Barrdahl M, Vineis P, Ambatipudi S, Uitterlinden AG, Hofman A, Schwartz J, Colicino E, Hou L, Vokonas PS, Hernandez DG, Singleton AB, Bandinelli S, Turner ST, Ware EB, Smith AK, Klengel T, Binder EB, Psaty BM, Taylor KD, Gharib SA, Swenson BR, Liang L, DeMeo DL, O'Connor GT, Herceg Z, Ressler KJ, Conneely KN, Sotoodehnia N, Kardia SL, Melzer D, Baccarelli AA, van Meurs JB, Romieu I, Arnett DK, Ong KK, Liu Y, Waldenberger M, Deary IJ, Fornage M, Levy D, London SJ. Epigenetic Signatures of Cigarette Smoking. Circ Cardiovasc Genet. 2016 Oct;9(5):436-447. doi: 10.1161/CIRCGENETICS.116.001506.

biological responses, evaluate within-person variations, conduct EWAS to identify potential environmental carcinogens, and develop exposome risk scores.

Many molecular signatures have been identified in relation to environmental exposures. However, few have been validated for exposure assessment in epidemiologic studies. Proper use of biomarkers will greatly facilitate epidemiologic studies to assess exposures and determine causality.

Precision environmental health: an emerging approach for cancer prevention

Jesse Goodrich, Ph.D., University of Southern California

Precision environmental health is the intersection of precision medicine and environmental health. Overarching goals include understanding how the totality of individual environmental exposures impact biological factors and cause disease; identifying molecular signatures; and developing personalized prevention and intervention strategies. Key to reaching these goals is integrating the exposome with omics data to enhance understanding of cancer risk and help develop novel prevention strategies.

For example, using data from the Multiethnic Cohort Study,[44] researchers examined PFAS and risk for cellular carcinoma. Through untargeted metabolomics, they found that PFAS exposure was associated with higher risk of hepatocellular carcinoma (HCC) and identified metabolites that may link perfluorooctanoic acid (PFOS) exposure to outcome. For instance, glucose and bile acids associated with PFOS were also associated with higher risk for HCC. Approaches that integrate multiple omics are essential to improving understanding of the impact of environmental exposures on cancer.

A second case study used data from the Human Early Life Exposome (HELIX) Study, including in utero arsenic exposures and omics data (DNA methylation, gene expression, and proteins) to assess risk of childhood leukemia and non-alcoholic fatty liver disease (NAFLD). Researchers attempted to integrate this data to identify children at highest risk of liver injury, which could lead to increased risk of cancer later in life. They extended the known method of Latent Unknown Clustering Integrating Multi-View Data[45] and identified groups of children defined by their levels of arsenic exposure and disease risk. Two groups had the same level of exposure, but differing disease risk because their transcription profiles varied. This study shows that groups of individuals can have different disease risks despite the same environmental exposure, and they can be identified using omics profiles.

Actionable steps towards cancer prevention using precision environmental health include:

- Involving transdisciplinary exposure scientists to ensure population-based studies incorporate measures of the exposome.
- Investigating windows of exposure, windows of susceptibility, and potentially windows of omics profiles to determine the best point to identify altered omics profiles.
- Measuring multiple omics layers.

[44] Goodrich JA, Walker D, Lin X, Wang H, Lim T, McConnell R, Conti DV, Chatzi L, Setiawan VW. Exposure to perfluoroalkyl substances and risk of hepatocellular carcinoma in a multiethnic cohort. JHEP Rep. 2022 Aug 8;4(10):100550. doi: 10.1016/j.jhepr.2022.100550.

[45] Peng C, Wang J, Asante I, Louie S, Jin R, Chatzi L, Casey G, Thomas DC, Conti DV. A latent unknown clustering integrating multi-omics data (LUCID) with phenotypic traits. Bioinformatics. 2020 Feb 1;36(3):842-850. doi: 10.1093/bioinformatics/btz667.

- Recruiting larger sample sizes, especially when looking at specialized markers of exposure.

The environmental exposome in cohort studies of cancer

Karin Michels, Sc.D., Ph.D., University of California, Los Angeles

When considering population-based cohorts, studies investigate environmental exposures and their molecular signatures, or molecular signatures and their association with cancer, but incorporating all three into one study is challenging given the latency of cancer. However, including the entire sequence — environmental exposure, molecular signature, and cancer — in one study is important for assessing causality.

Challenges in studying the environmental exposome in cohort studies of cancer include the fact that many environmental exposures do not have suitable biomarkers or biomarkers in an accessible tissue. Short half-lives require multiple samples (e.g., BPA), which may not be possible in a prospective study. There are also varying windows of susceptibility for different chemicals.

Opportunities to address these challenges include focusing on known windows of susceptibility and chemicals with known long-lasting molecular signatures. Developmental Origins of Health and Disease (DOHaD) should also be considered, given that the early-life exposures leave long-lasting molecular signatures that, in effect, serve as "memories" of the early exposome. Epigenetics is suspected to be the most important in DOHaD phenomena, but the microbiome likely also plays a role.

Future directions for this field should include creating an inventory of existing prospective cohort studies that can be exploited for their existing biorepositories to link environmental exposures to stable molecular signatures. Multiple biospecimens could be used, including blood, urine, stool, hair, nails (E.g. Breast Cancer and Environment Research Program (BCERP) cohorts[46] and Avon Longitudinal Study of Parents and Children (ALSPAC)[47]). In addition, appropriate intermediate endpoints should be defined for cancer, like mammographic density or colon polyps, to fill gaps in the sequence from exposure to cancer. Considering how future cohort studies should be designed, and determining which samples, biospecimens, and information are missing from existing cohorts would also be helpful.

Race/ethnicity as a modifier of exposure and molecular signatures

Loic Le Marchand, M.D., Ph.D., University of Hawai'i

Of similarly exposed individuals, only few will develop cancer; considering race and ethnicity could help explain why. Comparing racial or ethnic groups living in the same environment magnifies inter-individual differences (in exposure and biology) that may contribute to disease development. In most cases, these subtle differences are also present in racially and culturally homogenous populations but at lower frequency and thus are not as apparent. This concept can be demonstrated with data from the Multiethnic Cohort Study (MEC).

[46] https://bcerp.org/
[47] https://www.bristol.ac.uk/alspac/

In one study, researchers investigated ethnic differences in risk associated with body mass index (BMI) and type 2 diabetes[48] or breast cancer[49] and found several important differences: Japanese Americans and Native Hawaiians with higher BMIs had higher risk for both type 2 diabetes and breast cancer. They further investigated the association by looking at body fat composition. Specific fat tissues are more metabolically active, like visceral fat, and are more strongly associated with diabetes, hypertension, and heart disease. They found significant differences in visceral fat, adjusted for dual-energy X-ray absorptiometry total adiposity. At the same level of adiposity, Japanese American and Hawaiians had higher amounts of visceral fat, while African Americans had lower amounts.[50] The team developed a prediction score for visceral fat using nine biomarkers and applied it to the breast cancer cases. After adjusting for BMI and other risk factors, they found an association of 1.5 for upper level of visceral fat score. A two-fold increase was found for Japanese Americans. Overall, these findings show that BMI is not optimal for developing an omic signature of obesity.

Another study investigating smoking also found major risk differences in the association of smoking and lung cancer, with about 50% higher risk for African American and native Hawaiians compared to Caucasians, and a 50% lower risk for Japanese and Latinos compared to Caucasians.[51] When looking at internal smoking dose (total nicotine equivalents adjusted for cigarettes per day) as a measure of smoking, Japanese had lower internal smoking dose compared to Caucasians and African Americans.[52] Researchers then explored reasons for differences in the uptake of nicotine per cigarette and found the most likely reason for lower uptake in Japanese participants was a higher frequency of polymorphisms in the CYP2A6. People of Japanese heritage often have more non-functional or deleted alleles, meaning they take in less nicotine per cigarette, which could explain their lower risk of lung cancer.[53] To explain the higher uptake of nicotine per cigarette for African Americans, researchers are looking at whether contextual factors, such as social stress, might be a reason. Overall, smoking history is not optimal for developing an omics signature of smoking dose.

To identify signatures of exposure, there is a need to assess multiple omics in a standardized way in well-characterized samples. However, exposure measurements must first be optimized by accounting for

[48] Maskarinec G, Grandinetti A, Matsuura G, Sharma S, Mau M, Henderson BE, Kolonel LN. Diabetes prevalence and body mass index differ by ethnicity: the Multiethnic Cohort. Ethn Dis. 2009;19(1):49-55.
[49] White KK, Park SY, Kolonel LN, Henderson BE, Wilkens LR. Body size and breast cancer risk: the Multiethnic Cohort. Int J Cancer. 2012 Sep 1;131(5):E705-16. doi: 10.1002/ijc.27373.
[50] Lim U, Monroe KR, Buchthal S, Fan B, Cheng I, Kristal BS, Lampe JW, Hullar MA, Franke AA, Stram DO, Wilkens LR, Shepherd J, Ernst T, Le Marchand L. Propensity for Intra-abdominal and Hepatic Adiposity Varies Among Ethnic Groups. Gastroenterology. 2019 Mar;156(4):966-975.e10. doi: 10.1053/j.gastro.2018.11.021.
[51] Haiman CA, Stram DO, Wilkens LR, Pike MC, Kolonel LN, Henderson BE, Le Marchand L. Ethnic and racial differences in the smoking-related risk of lung cancer. N Engl J Med. 2006 Jan 26;354(4):333-42. doi: 10.1056/NEJMoa033250.
[52] Park SL, Carmella SG, Ming X, Vielguth E, Stram DO, Le Marchand L, Hecht SS. Variation in levels of the lung carcinogen NNAL and its glucuronides in the urine of cigarette smokers from five ethnic groups with differing risks for lung cancer. Cancer Epidemiol Biomarkers Prev. 2015 Mar;24(3):561-9. doi: 10.1158/1055-9965.EPI-14-1054.
[53] Park SL, Tiirikainen MI, Patel YM, Wilkens LR, Stram DO, Le Marchand L, Murphy SE. Genetic determinants of CYP2A6 activity across racial/ethnic groups with different risks of lung cancer and effect on their smoking intensity. Carcinogenesis. 2016 Mar;37(3):269-279. doi: 10.1093/carcin/bgw012.

important modifiers, such as race and ethnicity, genetics, and, possibly, contextual variables like social determinants of health. Differences can be expected in omics signatures for racial and ethnic minorities.

Discussion

Key discussion points from Session 5 included:

Windows of susceptibility. Available population studies do not always overlap with windows of susceptibility because most recruit middle-aged individuals. The BCERP puberty cohort is an example of a study that may address windows of susceptibility by enrolling girls 6-8 years of age to study the effect of puberty as an indicator of breast cancer risk. A study that contains cohorts focused on different windows of susceptibility with the same outcome and environmental chemical exposure would be useful. A national birth cohort of lifetime study should also be considered. Another future direction could identify markers that reflect early-life exposure, like methylation signatures that reflect prenatal exposure to tobacco smoking.

Population-based studies for ubiquitous exposures. The first challenge in studying ubiquitous exposures is determining how to measure and quantify them, because large differences in dose and interactions with other chemicals could complicate analysis. For example, exposures to PFAS are ubiquitous, but variability exists even within highly exposed populations. Disease biomarkers like methylation, age, or changes in metabolomic profiles could inform an EWAS to identify potential environmental carcinogens. Considering diversity — different ethnic groups in different geographic locations — could provide insights to this challenge as well.

Need for transdisciplinary research teams. Researchers who work with epidemiological cohorts and researchers who work with experimental model systems can come together to collaboratively tackle more comprehensive cancer research questions. Language barriers may exist, but making those connections, especially with organoid models, will be important.

Other forms of tissue for biomarkers. Mutational signature studies can be incorporated into population-based studies using circulating free DNA, urine, oral cells, and buccal cells, where mutational signatures can be measured using Nanoseq or duplex sequencing. These types of cells could represent “normal” tissue and are easily available. Hair and nail samples would also be useful as time-integrated markers for assessing long-term exposures.

Precision prevention. A challenge with precision prevention is that implies that a large proportion of the burden of cancer comes from a small part of the population. Integrating more omics data could potentially improve risk stratification, and different groups could be identified based on different omics profiles even with similar exposures.

Session 6: Workshop Summary and Future Directions

The final session was moderated by co-chairs Hannah Carter, Ph.D., and Cheryl Walker, Ph.D. Moderators from each session recapped key questions, challenges, and opportunities. Afterward, all speakers and moderators convened to discuss future directions for the field. Discussions aimed at identifying the most pressing scientific needs to move the field forward and how this information can be used to improve cancer prevention.

Session Summaries

Session 1: Mutational Signatures of Exposure in Cancer, Ludmil Alexandrov, Ph.D.

Key gaps and opportunities

- A complex and intricate connection exists between environmental carcinogens, mutational signatures, and cancer risk:
  - Some carcinogens have clear mutational signatures and contribute to cancer risk through mutagenesis.
  - Carcinogens can also cause large numbers of mutations without affecting cancer risk.
  - Carcinogens may act through mutational mechanisms, e.g., by activating endogenous processes.
  - There are also non-mutagenic carcinogens that drive cancer through other mechanisms (e.g., non-mutagenic promoting agents).
- Key major gaps in mutational signatures research:
  - For most mutational signatures, there is a limited direct link with cancer causality.
  - Many mutational signatures are orphaned, and their underlying mutational processes and molecular mechanisms are unknown.
  - For some cancers, mutational signatures do not explain international differences in cancer risk (e.g., ESCC) or known environmental exposures (e.g., asbestos).
- Opportunities in mutational signatures research:
  - Better understanding of mutagenic and non-mutagenic carcinogenesis.
  - Integrating mutational signatures with other data modalities for providing a better personalized understanding of cancer risk.
  - Detection of mutational signatures in normal tissues of healthy individuals.
  - Utilizing mutational signatures (in combination with other measurements) for predictive understanding of cancer risk.

Needs for the next five years

- Generation of additional data modalities (e.g., multi-omics, microbiomics, adductomics, etc.) from multiple tissues in clinically and epidemiologically well-annotated cohorts, such as cancer patients, matched healthy controls, and children (for tracking early-life exposures and their role in cancer later in life).
- Better mathematical approaches and additional computational methods for holistic integration and examination of multi-modal datasets.
- Reconstruction of mutagenesis (and other molecular processes) from the first cell to develop more complete predictive understanding of its role in cancer (and other diseases).

Applications to cancer prevention

- Moving from research settings to the field (example SBS22) to improve:
  - Ability to detect signatures in healthy individuals.
  - Cost-effective biomarkers that provide quantitative predictions.
  - Intervention strategies for cancer prevention (e.g., public engagement, policy, chemoprevention, etc.).
- There is a limited general commitment to cancer prevention (when compared to treatment and diagnostics). Students and researchers, research institutions, and funding agencies are critical for advancing this field.

Session 2: Other Data Types as Signatures of Exposure in Cancer, Scott Auerbach, Ph.D.

Key gaps and opportunities

- Further resolution of environmentally induced epigenetic changes and how they play a role in the formation of pre-cancer and cancer — they can be detected and associated, but it is often unclear what influence they actually have. This effort will require functional omic (proteomics, metabolomics, phenomics, etc.) linkage to epigenetic changes.
- Understanding the complex, systems-level dark energy/matter (non-mutational) that drive the causal aspects of the carcinogenic process and to what degree modifiable environmental exposures influence these processes. This will require more advanced data- and systems biology-modeling approaches.
- Development of liquid biopsy/cell-free/exosome omics (genome, epigenome, transcriptome, proteome, and metabolome) technologies to identify non-mutational signatures of cancer risk across all tissues and organs.
- Understanding the complex ecological landscape (both temporal and systemic) related to the development of pre-cancerous states.
- Development of longitudinal cohorts and sample resources with robust metadata that allow for greater temporal resolution of the environmental carcinogenic processes and application of causal models that can deconvolute drive versus passenger processes.

Needs for the next five years

- Initiation of longitudinal studies and samples starting early in life, serial sampling, and technologies for continuous monitoring of major dietary components (e.g., glucose).
- Development of technologies that can use non- and minimally invasive samples to allow for detection of tissue/organ age/cancer risk state (e.g., epigenetic clocks).
- Development of signatures of disease state based on more transient or dynamic technologies (e.g., metabolomics, transcriptomics) that can serve as intermediate phenotypes and flags for clinical intervention.
- Data to better understand the complex genetic, temporal, and ecological influences that drive promotional processes and modify the risk status in different tissues and organs. More empirical data is needed to model what effects measured changes have on cancer risk.
- More effective integrations of multi-omic technologies, with the goal of understanding the relationship between mutagenesis, promotion, and the ecological landscape (e.g., microbiome, metabolome) of interactions that impact carcinogenic processes.
- Technologies that can provide more accurate records of impactful exposures (e.g., adductomics, epigenetic changes).
- A better understanding of the epistasis between the genome (background genetics) and cancer promotional process influences in the environment.

Overall thoughts

- The development of cancer is a highly complex process with profound epistasis at many levels. If substantive progress in understanding how this happens in a genetically and culturally diverse human population is to be made, there are some critical needs:
    - Well annotated longitudinal data (control people to themselves over time), and massive amounts of it. This points to a need for:
        - Technologies that use minimally to non-invasive methods to query deeply and specifically into complex biological patterns across all organs and tissues.
        - Technologies that can capture longitudinal changes in the exposome.

  - Better solutions to the "small n, big p problem." Causal inference, data reduction, and Bayesian methods could address this challenge.
  - Complex test systems to validate causal inferences.

Session 3: Computational Challenges and Integrating Multi-Omics to Identify Signatures, Mona Singh, Ph.D.

Key gaps and opportunities

- There is a need to move beyond mutational signatures.
- Within mutational signatures:
  - Need new methods to capture contributions of DNA damage versus DNA repair, which interact together to yield mutational profiles.
  - Need better methods to compare signatures to unveil common DNA repair mechanisms and common exposures.
- Network approaches are a powerful paradigm to integrate mutations, epigenetics, expression, and pathways.
- With respect to machine learning,
  - The most current mutation signature detection methods are linear.
  - More powerful deep-learning methods need significantly more data.
  - Other nonlinear machine learning methods may yield good performance while being more interpretable.
- Methods for linking signatures to exposures need access to data about the molecular impacts of exposures.
- There is a need for better validation of methods to detect signatures and for gold-standard benchmarks.

Needs for the next five years

- Large omic datasets derived from model systems and cell lines exposed to known combinations and timings of different exposures. Ideally, omics would be measured for all data types that algorithms would use. Such datasets are critical for testing signature detection and linking to exposures.
- Diverse omics data across time (pre-malignant to cancer and beyond) is necessary to understand how exposures affect cancer initiation and progression.
- Development of a "Human Cell Exposure Atlas" of multi-omic single cell data to assess cellular response to exposure in complex tissues.
- Standardized data in repositories.

Session 4: Challenges in Tracking Signatures of Exposures, Rebecca Fry, Ph.D.

Key gaps and opportunities

- New model types may be needed to identify molecular signatures of exposure, which should consider which cell types and animal models best recapitulate human population effects and whether they are generalizable.
- Determining which critical measurements (epigenetics, adducts, metabolomics, mutations, circular RNAs) should be included in molecular signatures analyses. This should consider the stability of biomarkers, life stage, dose and timing, and how these relate to individual susceptibility factors (genetics, DNA repair).

- Methods for measuring exposure of the complex environment (chemical mixtures, non-chemical mixtures, and the exposome).
- Vigilance in finding analytical tools for mixtures, pathway analysis, databases, and other tools that can be used to query exposure-biomarker-outcome relationships.

Needs for the next five years

- Continue to enhance the use of various model systems to investigate molecular signatures that indicate dose, timing of exposure, and initiation process. These include the appropriate use of mouse embryo fibroblasts, human-induced pluripotent stem cells (iPSCs), human-derived organoids (3D), and mouse models with specific mutations.
- Continue to study biomarkers, their specificity (chemical, dose, timing, tissue), and their interactions and related physiologic measurements (cancer phenotypes, cancer progression. These biomarkers include circular RNAs, metabolomics, DNA methylation, DNA adducts, DNA mutation, metal isotopes, and protein adducts.
- Continue to develop methods to address chemical mixtures and the exposome. These include the laboratory methods that systematically test chemicals of interest and advanced statistical methods for mixtures analysis.

Application to cancer prevention

- Understanding the mechanism of chemical-induced carcinogenesis is the first step toward prevention. This builds from data model systems with controlled chemical exposures, to comparisons of human physiologic endpoints (COSMIC signature).
- Approaches that focus on biological pathways may help identify cancer-prevention methods related to the environment. Aflatoxin-broccoli sprout example serves as a model.
- Methods currently used for non-cancer treatment, such as metal chelation, could be considered.

Session 5: Population-Based Cancer Studies, Paul Brennan, Ph.D.

Key gaps and challenges

- It is unclear how low-level exposures relate to cancer risk.
- There is a need to consider exposure mixtures, timing of exposure, susceptibility windows, and total exposome.
- Suitable biomarkers for many environmental exposures are lacking, those with short half-life (BPA) require multiple samples, and accounting for developmental windows is challenging.
- There is an absence of cross talk between population scientists and experimental scientists.
- Most population-based studies involve populations of European ethnicity; race can be an important modifier of risk.
- Addressing ubiquitous exposures.
- Some mutational signatures developed on case series lack accurate measures of exposure (mainly TCGA/ICGC).
- Most population cohorts are involve participants recruited in middle aged, relatively limited exposure assessment, or one-off collection of samples.
- Sharing of data and samples is extremely difficult. The field must reduce the administrative burden associated with data and biosample sharing and emphasize the financial and scientific cost associated with putting up barriers.

Key opportunities

- Technology development for more accurate exposure assessment on many fronts, such as duplex sequencing (for normal tissue), metabolomics, proteomics, and metallomics. This effort will require appropriate biological samples and standardization of methods and labs. The cost for large-scale exposure assessment is another important issue.
- Enhanced measures of intermediate outcomes like clonal hematopoiesis or normal tissue clonal development.
- Sampling blood and urine for inflammation biomarkers and methylation for multi-omics of the exposome.
- New forms of cohorts, such as mother-child cohorts, to look at early-life exposures, and more diverse cohorts representing different populations.
- Investigating alternative forms of tissue, including oral cells, urine, benign breast tissue, colonic polyps, and nasal swabs.
- A greater focus on microbiome studies.
- Mendelian randomization studies that can build on large GWA studies and large population cohorts with genetic data.

Needs for the next five years

- Develop new cohorts and studies that cut across the life range.
- Consider multiple collections of biological samples. This approach is expensive, but important for early detection studies.
- Identify molecular signatures that reflect early-life exposures, which may include child cohorts.
- Connect epigenetic studies to mechanisms using experimental models like organoids to validate signatures from human populations.
- Develop mutational signature studies based on large case series across multiple cancers with accurate exposure measurements, or within specific high risk populations.
- Large scale mutational signature studies based on normal cells from easily accessible sources of tissue like oral cells.
- Enhance international collaborations by bringing together large-scale population-based studies and researchers from diverse settings.

Future Directions

Detecting low-level exposures. At the population level, there are challenges in investigating low-level exposures that are difficult to quantify. Identifying what is needed to get resolution to detect these exposures is important to consider when designing future studies and cohorts. A good foundation may be to start with dissecting mechanisms of individual exposures. To understand exposures, a large cohort study followed over a long period of time, possibly a lifetime, would be useful. To understand the mechanisms, more studies using organoids would be useful. Multi-omics and other data types would inform understanding of tumors after development, as well development of therapeutic interventions. Developing in-depth benchmarks will be important for assessing exposure detection, how well an algorithm works, and what it misses.

Tackling one problem at a time. Future research could address the categories of risk assessment, hazard identification, risk prediction, and early detection one issue at a time. Research could also focus on the

biggest emerging problem first (e.g., early-onset colorectal cancer or lung cancer in never-smokers), then use that work as a successful example for other cancer types. Specific tissues or biomarkers also merit focus. The field could look at previous success stories in cancer, like aflatoxin and liver cancer, to use as a model.

Microenvironment. Research should investigate whether an exposure works by affecting the microenvironment of a tissue. A major research gap exists in understanding how exposures affect tissues at a single-cell level.

Access to normal data. Studies where "normal" means morphologically normal tissue adjacent to a tumor can be problematic because of field effects where the tumor may influence the surrounding tissue. Access to normal tissue from disease-free individuals has been lacking across many studies. Identifying an accurate representation of "normal" is necessary as a basis for comparison. Obtaining blood, normal adjacent tissue, and tumor tissue could give perspective, for example, on clonal expansion for episodic exposures.

Data sharing. Data are often deposited in ways that are not useful. Data are also processed in different ways, encumbering efforts to combine them. The community should consider standards for processing data for sharing. It is also critical to share metadata.

Appendix 1: Participant Biographies

Ludmil Alexandrov, Ph.D.

University of California San Diego

Ludmil Alexandrov is an associate professor in the Departments of Bioengineering and Cellular & Molecular Medicine at University of California San Diego (UC San Diego). Alexandrov's research focuses on understanding large-scale molecular sciences datasets and leveraging this knowledge for developing better cancer-prevention strategies and improving cancer treatment. He has co-authored 135 peer-reviewed manuscripts, including 30 publications in Nature, Science, or Cell. Alexandrov is best known for creating the concept of mutational signatures, developing bioinformatics tools for analyses of cancer genomics data, and pioneering novel AI approaches for improving cancer treatment and cancer prevention. He received his doctoral degree in cancer genetics from the Sanger Institute and University of Cambridge.

Trevor Archer, Ph.D.

National Institute of Environmental Health Sciences

Trevor Archer, Ph.D., received a doctoral degree in biochemistry in 1987 at Queen's University, Kingston, Ontario, Canada, after which he completed postdoctoral training on chromatin gene transcription and steroid receptors at the National Cancer Institute in Bethesda, Maryland. In 1992, Archer joined the University of Western Ontario in Canada, as a National Cancer Institute of Canada scientist. Archer was recruited to the National Institute of Environmental Health Sciences (NIEHS) in 1999 as head of the Chromatin & Gene Expression Group and was later appointed as chief of the Laboratory of Molecular Carcinogenesis in February 2003. In 2014, Archer became the founding chief of the new Epigenetics & Stem Cell Biology Laboratory at NIEHS. Archer has made numerous original and important contributions to the study of chromatin structure and function, epigenetics, and gene transcriptional regulation in breast cancer cells while publishing roughly 120 peer-reviewed manuscripts.

Scott Auerbach, Ph.D.

National Institute of Environmental Health Sciences, Division of Translational Research

Scott Auerbach, Ph.D., received a dual bachelor's degree from The Pennsylvania State University in physiology and biochemistry/molecular biology, and his doctoral degree in pharmacology from the University of Washington. The focus of his doctorate research was characterizing the impact of structural variation of nuclear receptors caused by alternative mRNA splicing. He was then a postdoctoral researcher at Duke University, where he studied the genetics of human pulmonary fibrosis. Subsequently, he accepted a fellowship at the Division of the National Toxicology Program (DNTP), now the Division of Translational Research (DTT). In 2009, he became a staff scientist at the DNTP at NIEHS and a diplomat of the American Board of Toxicology. He is currently the toxicoinformatics group leader in the Predictive Toxicology Branch of the DTT. His research focuses on applying molecular and high-dimensional data to toxicology, with the goal of increasing efficiency and bridging the knowledge of toxicology to novel means of quantifying biological change. Since joining the DTT he has led efforts to apply machine learning to carcinogenicity and toxicity prediction, the DNTP's rapid response to the Elk River chemical spill, and the in vivo genomic dose response analysis and reporting group. Further, he has been guiding the development of the genomic dose-response software, BMDExpress, in addition to other software to enhance the interpretation omic and high-throughput toxicology data.

Allan Balmain, Ph.D.

University of California San Francisco, Helen Diller Family Comprehensive Cancer Center

Allan Balmain, Ph.D., F.R.S.E., F.R.S., F.A.A.C.R., is the Barbara Bass Bakar Distinguished Professor of Cancer Genetics at the University of California San Francisco (UCSF) Helen Diller Family Comprehensive Cancer Center in San Francisco. He has a long-standing interest in the interactions between genetic background and environmental factors, including mutagens and tumor promoters, that lead to initiation and promotion of cancer. He addresses these questions mainly using mouse models to understand the genetic and biological changes during multistage carcinogenesis.

Paul Brennan, Ph.D.

International Agency for Research on Cancer

Paul Brennan, Ph.D., leads the Genomic Epidemiology at the International Agency for Research on Cancer (IARC). IARC is the cancer research Agency of the WHO. The main focus of his work my is to use genomics and epidemiology to better understand causes of cancer. He has a particular focus on trying to understand causes of esophageal cancer (with a focus in North East Iran), as well as renal and pancreas cancer (in central Europe), and head and neck cancers (in Latin America).

Joshua Campbell, Ph.D.

Boston University Department of Medicine, Division of Computational Biomedicine

Joshua Campbell, Ph.D., received a bachelor's degree in biology, computer science, and mathematics from Anderson University in Anderson, Indiana. He then received his doctoral degree in bioinformatics from Boston University. He performed a postdoctoral fellowship in cancer genomics at the Dana-Farber Cancer Institute and the Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard. Campbell is currently an assistant professor at the Boston University Department of Medicine within the Division of Computational Biomedicine. Campbell has many research interests including computational biology and bioinformatics, identifying early drivers of lung cancer, and the therapeutic development and pathogenesis of chronic obstructive pulmonary disease (COPD).

Hannah Carter, Ph.D.

University of California San Diego

Hannah Carter, Ph.D., is an associate professor in the UC San Diego Department of Medicine, Division of Medical Genetics. Carter develops and applies computational approaches to aid the interpretation of genetic variation and advance precision cancer medicine. A major focus of the lab is on joint analysis of tumor and inherited genomes to uncover the role of genetic background in shaping cancer risk, somatic tumor evolution, and therapeutic response. Carter is a member of the UC San Diego Institute for Genomic Medicine, the Bioinformatics and Systems Biology Program, and the Moores Cancer Center. She received her doctorate degree in biomedical engineering from Johns Hopkins University and her master's degree in electrical and computer engineering from the University of Louisville. She is a CIFAR Azrieli Global Scholar, a Siebel Scholar, a Mark Foundation Emerging Leader, a Jaime Wyatt Miller Fellow, and a recipient of the 2013 NIH Director's Early Independence Award.

Phillip Daschner, M.Sc.

National Cancer Institute

Phillip Daschner, M.Sc., is a program director in the Cancer Immunology, Hematology, and Etiology Branch in the Division of Cancer Biology, National Cancer Institute (NCI). He currently manages a

portfolio of basic research grants that investigate mechanisms of biological agents (most viral and bacterial) and host predisposing states that are etiological factors or cofactors in carcinogenesis. He is currently in leadership roles on both the Trans-NIH Microbiome Working Group (TMWG) and the NCI Microbiome Working Group, which coordinate ongoing microbiome-related initiatives and activities at the NIH and Institute levels. His research interests include the role of the microbiome in carcinogenesis, tumor immunity and therapy efficacy, infection derived cancers, cellular defense mechanisms, dietary phytochemicals and metabolites in chemoprevention, inflammation and cell stress response pathways, and cancer health disparities.

John Essigmann, Ph.D.

Massachusetts Institute of Technology

John Essigmann, Ph.D., is the William and Betsy Leitch Professor of chemistry, biological engineering, and toxicology at MIT. His laboratory studies the responses of cells to DNA-damaging agents, with a specific emphasis on mechanisms of mutagenesis and genotoxicity. One major line of work involves the synthesis of oligonucleotides containing known carcinogen- or drug-DNA adducts, insertion of the modified oligonucleotide into the genomes of viruses, and replication of the modified viral genome in living cells; this work defines the type, amount and genetic requirements for mutagenesis and toxicity. In other work, his laboratory uses high-resolution mutational spectrometry to determine the mutational spectra of DNA damaging agents, such as aflatoxin B_1 and *N*-nitrosodimethylamine.

Michael Fischbach, Ph.D.

Stanford University

Michael Fischbach, Ph.D., is an associate professor in the Departments of Bioengineering and Microbiology & Immunology at Stanford University, an institute scholar of Stanford Chemistry, Engineering, and Medicine for Human Health (ChEM-H), and the director of the Stanford Microbiome Therapies Initiative. Fischbach is a recipient of the National Institutes of Health (NIH) Director's Pioneer and New Innovator Awards, a Howard Hughes Medical Institute (HHMI) HHMI-Simons Faculty Scholars Award, a Fellowship for Science and Engineering from the David and Lucille Packard Foundation, a Medical Research Award from the W.M. Keck Foundation, a Burroughs Wellcome Fund Investigators in the Pathogenesis of Infectious Disease award, and a Glenn Award for Research in Biological Mechanisms of Aging. His laboratory uses a combination of genomics and chemistry to identify and characterize small molecules from microbes, with an emphasis on the human microbiome. Fischbach received his doctoral degree as a John and Fannie Hertz Foundation Fellow in chemistry from Harvard University in 2007, where he studied the role of iron acquisition in bacterial pathogenesis and the biosynthesis of antibiotics. After two years as an independent fellow at Massachusetts General Hospital, Fischbach joined the faculty at the University of California San Francisco, where he founded his lab, before moving to Stanford in 2017. Fischbach is a co-founder and director of Federation Bio and Kelonia, a co-founder of Revolution Medicines, a member of the scientific advisory boards of NGM Biopharmaceuticals and Chan Zuckerberg Science, and an innovation partner at The Column Group.

Yvonne Fondufe-Mittendorf, Ph.D.

The Van Andel Institute

Yvonne Fondufe-Mittendorf, Ph.D., is a professor of epigenetics at The Van Andel Institute. Fondufe-Mittendorf obtained her doctoral degree from the University of Goettingen, Germany, and did a postdoctoral fellowship at Northwestern University in the lab of the late Jonathan Widom. She then

went on to her first faculty job at the University of Kentucky, where she became a professor in the Department of Molecular and Cellular Biochemistry. Fondufe-Mittendorf moved to The Van Andel Institute in January of 2022 as a professor in the Department of Epigenetics. Her lab studies how the epigenome is reprogrammed in response to an environmental toxicant to drive diseases such as cancer.

Rebecca Fry, Ph.D.

University of North Carolina at Chapel Hill, Gillings School of Global Public Health

Rebecca Fry, Ph.D. is the Carol Remmer Angle Distinguished Professor in Children's Environmental Health and associate chair in the Department of Environmental Sciences and Engineering at the Gillings School of Global Public Health at the University of North Carolina at Chapel Hill (UNC-Chapel Hill). Fry is the founding director of the newly launched Institute for Environmental Health Solutions (IEHS) at UNC-Chapel Hill. Fry received her doctoral degree in biology from Tulane University with postdoctoral training in toxicogenomics and environmental health sciences at MIT. A primary goal of Fry's research is to increase awareness of the deleterious impacts of toxic exposures during the prenatal period with a focus on the epigenome and developmental origins of health and disease.

Dan Gallahan, Ph.D.

National Cancer Institute

Dan Gallahan, Ph.D., is the NCI director for the Division of Cancer Biology. Gallahan started as NCI an intramural researcher focusing on the utilization of model systems to help understand the role of genetic alterations in breast cancer and the role of human papillomaviruses in cancer. He also spent time in private industry exploring the commercial and applied side of research, helping to establish a molecular diagnostic test. In the NCI extramural community, he has been responsible for the establishment of many important programs and scientific innovations having direct impact on new knowledge and cancer advances, such as the Stamp Out Breast Cancer Act, Trans-NCI Innovative Molecular Analysis Technologies (IMAT) program, and the Integrative Cancer Biology Program (ICBP). In his relentless pursuit of innovation and desire for a better understanding of cancer, he has advanced quickly to become the deputy director of the Division of Cancer Biology and subsequently the director of the division in 2019.

Mark Gerstein, Ph.D.

Yale University, Yale Computational Biology & Bioinformatics Program

After graduating from Harvard with a bachelors in physics in 1989, Mark Gerstein, Ph.D., earned a doctorate in theoretical chemistry and biophysics from Cambridge in 1993. He did postdoctoral research at Stanford, then came to Yale in 1997 as an assistant professor. In 2003, he became co-director of the Yale Computational Biology & Bioinformatics program. Gerstein has published over 600 publications, including several in prominent journals, such as Science and Nature. His research is focused on biomedical data science and he is particularly interested in machine learning, macromolecular simulation, human-genome annotation, disease genomics, and biomedical privacy.

Ben Gewurz, M.D., Ph.D.

Harvard Medical School, Brigham & Women's Hospital

Ben Gewurz, M.D., Ph.D. is an associate professor at the Brigham & Women's Hospital and Harvard Medical School. He is the associate chair of the Harvard Graduate Program in Virology and a founding member of the Broad Institute Center for Integrative Solutions in Infectious Diseases. His laboratory studies Epstein-Barr virus (EBV) driven B-cell lymphomagenesis and gastric carcinogenesis, including EBV-

driven metabolism remodeling, epigenetic control of viral oncogene expression, the EBV lytic switch, oncogene pathways, and host/virus interactions. Gewurz is the president of the International EBV Association, a PLoS Pathogens associate editor, and a member of the Virology, Journal of Virology and Tumor Virus Research editorial boards.

Jesse Goodrich, Ph.D.

University of Southern California

Jesse Goodrich, Ph.D., is an assistant professor in the Department of Population and Public Health Sciences at the University of Southern California. His research combines data on mixtures of environmental exposures with information from omics datasets, to better characterize the effects of environmental pollutants on cancer risks. In particular, his recent work has focused on how per- and poly-fluoroalkyl substances (PFAS), a persistent and ubiquitous group of chemicals detected in blood of over 99% of people in the U.S., increase the risk of liver cancer via alterations in key metabolic pathways linked to glucose and amino acid metabolism.

Michelle Heacock, Ph.D.

National Institute of Environmental Health Sciences, Hazardous Substances Research Branch

Michelle Heacock, Ph.D., received her doctorate from Texas A&M University in College Station, Texas, for her work on the interplay between DNA repair proteins and telomeres. Her postdoctoral work was conducted at the NIEHS where she studied the DNA repair pathway, base excision repair. Her research focused on understanding the causes of cellular toxicity caused by DNA-damaging agents. Heacock is currently serving as the acting branch chief of the Hazardous Substances Research Branch and is a health science administrator overseeing Superfund Research Program (SRP) grants that span basic molecular mechanisms of biological responses from exposures to hazardous substances, movement of hazardous substances through environmental media, detection technologies, and remediation approaches. She has been with NIEHS since 2007.

Cathrine Hoyo, Ph.D.

North Carolina State University

Cathrine Hoyo, Ph.D., is the Goodnight Distinguished Innovation Chair and professor in biological sciences and directs the Epidemiology and Environmental Epigenomics Laboratory at North Carolina State University (NC State). Her group's research program aims to improve our understanding of how early development influences the risk of common chronic diseases, especially those that exhibit racial/ethnic differences in outcomes, including liver cancer and metabolic diseases. Her group has used a two-pronged approach to accomplish this. The first tactic is to identify environmentally responsive epigenetic elements that can be evaluated as a link between environmental stressors and common chronic diseases in children and adults. She also has assembled and is following multiple cohorts of newborns and otherwise healthy adults to identify environmentally responsive epigenetic targets that mediate environmental exposures and chronic-disease susceptibility in children and in adults.

Ron Johnson, Ph.D.

National Cancer Institute

Ron Johnson, Ph.D., is a program director in the DNA and Chromosome Aberrations Branch in the Division of Cancer Biology, NCI. Johnson oversees a portfolio of cancer biology research awards related to chemical and physical carcinogens, DNA damage, and gene expression with a focus on lung, bladder,

and liver cancers. Johnson received a doctorate in biochemistry from the Johns Hopkins School of Medicine and completed postdoctoral studies in developmental biology at the Stanford School of Medicine.

Maria Teresa Landi, M.D., Ph.D.

National Institutes of Health, National Cancer Institute

Maria Teresa Landi, M.D., Ph.D., has training in clinical oncology and molecular epidemiology. She is a senior advisor for the Genomic Epidemiology, Trans-Divisional Research Program, and a senior investigator for the Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH. She focuses her research on the genetic and environmental determinants of lung cancer and melanoma, and on the genomic characterization of these tumors. She is the principal investigator of both EAGLE and Sherlock-Lung, two landmark studies of lung cancer in smokers and never-smokers, respectively, which identified subtypes with distinct genomic features, mutational signatures, and evolutionary trajectories. She is also the leader of the MelaNostrum consortium, with the largest family study of melanoma worldwide.

Somdat Mahabir, Ph.D., M.P.H.

National Cancer Institute

Somdat Mahabir, Ph.D., is a program director in the Environmental Epidemiology Branch of the Epidemiology and Genomics Research Program (EGRP) in the NCI Division of Cancer Control and Population Sciences (DCCPS). His responsibilities include managing research that focuses on cancer epidemiology of modifiable risk factors such as environmental exposures and lifestyle factors, and the development of scientific research initiatives. Mahabir leads the Cohorts for Environmental Exposures and Cancer Risk (CEECR) program and was involved with the development of the NIH Environmental Influences on Child Health Outcomes (ECHO) program. Mahabir served as co-chair for research on the 2019-2023 Trans-NIH Strategic Plan for Women's Health Research, and currently serves as co-chair of the NIEHS-NCI Cancer and the Environment Working Group. Prior to joining EGRP in 2009, Mahabir was an assistant professor in the Department of Epidemiology at The University of Texas MD Anderson Cancer Center. Mahabir is the recipient of an NCI Cancer Prevention Research Training Merit Award, NCI Director's Award, CDC/Agency for Toxic Substances and Disease Registry (ATSDR) Honor Award, NIH Director's Award and academic awards from New York Medical College and New York Institute of Technology.

Loïc Le Marchand, M.D., Ph.D.

University of Hawai'i Cancer Center

Loïc Le Marchand, M.D., Ph.D., is a professor of cancer epidemiology in the Population Sciences in the Pacific Program and currently serves as associate director for population sciences and community outreach and engagement at the University of Hawai'i Cancer Center. His research focuses on the role of biological and environmental factors in the etiology of colorectal, lung, and breast cancers, especially regarding ethnic/racial differences in cancer risk.

Karin Michels, Sc.D., Ph.D.

University of California, Los Angeles, Fielding School of Public Health

Karin Michels, Sc.D., Ph.D., is professor of epidemiology at the University of California, Los Angeles (UCLA) Fielding School of Public Health in Los Angeles. From 2016-2021, she served as chair of the

department at UCLA after a 25-year stint at Harvard. Michels received her doctoral training in epidemiology at Harvard School of Public Health and earned an additional doctorate in biostatistics from Cambridge University in the United Kingdom. Her research focuses on the developmental origin of cancer, particularly breast cancer. She served as the principal investigator of one of the U01 grants in The Breast Cancer and the Environment Research Program (BCERP) consortium that explored the role of various environmental chemicals on pubertal maturation and development of the mammary gland. Michels also studies the influence of nutrition on health and heads several ongoing intervention studies of diet and the microbiome. She is also the cofounder of the area of epigenetic epidemiology and published the leading textbook in this field.

Ana Navas-Acien, M.D., Ph.D.

Columbia University, Mailman School of Public Health

Ana Navas-Acien, M.D., Ph.D., is a professor of environmental health sciences at Columbia University's Mailman School of Public Health. Her research investigates the health effects of environmental exposures (metals, tobacco smoke, e-cigarettes, air pollution), molecular pathways and gene-environment interactions, and effective interventions for reducing involuntary exposures and their health effects, with the goal of improving people's health and advance environmental justice. She obtained her medical degree from the University of Granada, Spain, and completed her residency training in preventive medicine and public health at the Hospital La Paz, Madrid, and her doctoral degree in epidemiology at Johns Hopkins University, Baltimore. She is recognized for bridging medical and environmental health sciences using a participatory approach. She directs the Columbia University Northern Plains SRP, a center that integrates science, technology, and traditional knowledge to protect the Northern Plains water resources and Indigenous communities from hazardous metal exposures.

Serena Nik-Zainal, Ph.D.

University of Cambridge

Serena Nik-Zainal, Ph.D., is a Professor of Genomic Medicine and Bioinformatics and an NIHR Research Professor at the University of Cambridge. She studied medicine at the University of Cambridge in 2000 and completed a doctorate degree at the Wellcome Sanger Institute (WSI) in 2009 exploring breast cancer using whole genome sequencing (WGS). She demonstrated how detailed downstream analyses of all mutations present in WGS breast cancers could reveal mutation signatures, which are imprints left by mutagenic processes that have occurred through cancer development. She identified a novel phenomenon of localized hypermutation termed "kataegis". Nik-Zainal was awarded a Wellcome Trust Intermediate Clinical Fellowship in 2013. She joined the Sanger Institute faculty team in 2014 and continued to develop expertise in the analysis and interpretation of WGS tumors. Apart from using computational approaches, she also studies mutational signatures experimentally using cell-based model systems. Nik-Zainal ran a clinical project, Insignia, recruiting patients with DNA repair/replication defects, aging syndromes and neurodegeneration, and people who have been exposed to environmental/occupational mutagens to gain biological insights into mutational phenomena in these patients. Nik-Zainal moved to the Department of Medical Genetics in 2017 to accelerate the translation of her genomics expertise towards clinical applications and further her work into the physiological mechanisms underpinning mutagenesis.

Arun Pandiri, Ph.D.

National Institute of Environmental Health Sciences, Molecular Pathology Group

Arun Pandiri, Ph.D., leads the Molecular Pathology Group at NIH. He is a diplomate of the American College of Veterinary Pathologists and the American Board of Toxicology. He was previously an NIEHS Intramural Research Training Award (IRTA) fellow, then a contract pathologist from Experimental Pathology Laboratories, Inc. He earned his veterinary degree from Acharya N. G. Ranga Agricultural University (ANGRAU), Hyderabad, India; master's degree from the University of Arkansas, Fayetteville; doctoral degree from Michigan State University and the United States Department of Agriculture (USDA) ARS Avian Disease and Oncology Laboratory, East Lansing; and pathology residency training at NC State, Raleigh. His group is currently working on multiple -omics projects related to chemical carcinogenesis using rodent and human tumor samples as well as several projects to understand the environmental contributions of early onset colorectal cancers.

David Phillips, Ph.D.

King's College London

David H. Phillips, Ph.D., D.Sc., FRCPath, is a Professor of Environmental Carcinogenesis at King's College London. His research interests are in the mechanisms of metabolic activation of environmental carcinogens, the detections and identification of DNA adducts, and the biological consequences of such DNA damage — what cells do to carcinogens and what carcinogens do to cells. In recent years, his attention has focused on generating whole genome mutational signatures in *in vitro* systems with the aim of determining the environmental origins of mutations found in human tumors.

Teresa Przytycka, Ph.D.

National Library of Medicine, National Center for Biotechnology Information

Teresa Przytycka, Ph.D., is a senior investigator at the National Center for Biotechnology Information in the National Library of Medicine (NLM) and NIH. The research in her group focuses on computational methods advancing the understanding of biomolecular systems, including gene regulation, biological networks, and the emergence of complex phenotypes, including cancer. In 2021, she was elected an International Society for Computational Biology (ISCB) fellow.

John Quackenbush, Ph.D.

Harvard University

John Quackenbush, Ph.D., is professor of computational biology and bioinformatics and chair of the Department of Biostatistics at the Harvard T.H. Chan School of Public Health, professor in the Channing Division of Network Medicine, and professor at the Dana-Farber Cancer Institute. Quackenbush completed his doctorate degree in theoretical physics, but in 1992 he received a fellowship to work on the Human Genome Project. This led him through the Salk Institute, Stanford, The Institute for Genomic Research (TIGR), and to Harvard in 2005. Quackenbush uses massive data to probe how many small effects combine to influence human health and disease. He has more than 300 scientific papers and over 73,000 citations. Among his honors is recognition in 2013 as a White House Open Science Champion of Change.

Daniel Shaughnessy, Ph.D.

National Institute of Environmental Health Sciences, Exposure, Response, and Technology Branch

Daniel Shaughnessy, Ph.D., joined the Division of Extramural Research and Training in 2006. As a postdoctoral fellow in the Laboratory of Molecular Carcinogenesis at NIEHS, he conducted research on the risks and protective effects of dietary factors on DNA damage in humans. Shaughnessy manages a portfolio of grants related to DNA repair and mutagenesis. He also manages grants on the development and validation of biomarkers of response to environmental stress, with a current focus on early biomarkers of mitochondrial dysfunction and altered signaling in response to environmental stress. He is the program contact for the small business programs (SBIR/STTR) at NIEHS. He received a doctoral degree from UNC-Chapel Hill in 2002 and a master's degree from UNC-Chapel Hill in 2000, studying the molecular mechanisms of dietary antimutagens.

Mona Singh, Ph.D.

Princeton University

Mona Singh, Ph.D., is a professor of computer science in the Lewis Sigler Institute for Integrative Genomics. She has been on the faculty at Princeton University since 1999. She received her bachelor's and master's degrees from Harvard University, and her doctorate from MIT, all in computer science. She works broadly in computational molecular biology and its interface with machine learning and algorithms. Much of her work is on developing algorithms to decode genomes at the protein level and she is especially interested in developing data-driven methods for predicting and characterizing protein sequences, functions, interactions, and networks, both in healthy and disease contexts. Among her awards are the Presidential Early Career Award for Scientists and Engineers (PECASE) in 2001, and the Rheinstein Junior Faculty Award from Princeton's School of Engineering and Applied Science in 2003. She was named a fellow of the Association for Computing Machinery (ACM) in 2019 and of the Informational Society for Clinical Biostatistics (ISCB) in 2018.

Cheryl Walker, Ph.D.

Baylor College of Medicine

Cheryl Walker, Ph.D., is the director of the Center for Precision Environmental Health and a professor in the Departments of Molecular & Cell Biology and Medicine at Baylor College of Medicine. She currently directs the NIEHS Center for Translational Environmental Health Research and serves on the board of scientific advisors for the National Cancer Institute. Walker's studies on the role of the epigenome in gene-environment interactions have yielded significant insights into mechanisms by which early-life exposures influence health and disease across the life course. Her work has also led to the discovery of new tumor suppressor functions in the cell. Walker earned her bachelor's degree in 1977 from the University of Colorado-Boulder in molecular, cellular, and developmental biology, and a doctorate in 1984 in cell biology from The University of Texas Southwestern Medical School, with additional post-doctoral training as a staff fellow at National Institute of Environmental Health Sciences. She has been recognized with the 2016 Leading Edge in Basic Research Award from the Society of Toxicology, is a fellow of the Academy of Toxicological Sciences and the American Association for the Advancement of Sciences (AAAS), and in 2016 was elected to National Academy of Medicine.

Ting Wang, Ph.D.
Washington University School of Medicine in St. Louis

Ting Wang, Ph.D., is the inaugural Sanford C. and Karen P. Loewentheil Distinguished Professor of Medicine at Washington University School of Medicine. His group is known for defining the widespread contribution of transposable elements (TEs) to the evolution of gene regulatory networks as well as to the 3D genome architecture, and for revealing that epigenetic dysregulation of TEs is a major mechanism driving oncogenesis. His lab is home to the WashU Epigenome Browser, utilized by investigators around the world to access hundreds of thousands of genomic datasets generated by large Consortia including the NIH Roadmap Epigenome Project, Encyclopedia of DNA Elements (ENCODE), 4D Nucleome, TaRGET, Impact of Genomic Variation on Function (IGVF), and the Human Pangenome Project. Wang currently directs the NIEHS Environmental Epigenomics Data Center, the Human Pangenome Reference Consortium, the IGVF Data Administrative and Coordination Center, the SMaHT Network Organization Center and Genome Characterization Center.

Wei Zheng, M.D., Ph.D., M.P.H.
Vanderbilt University School of Medicine, Division of Epidemiology

Wei Zheng, M.D., Ph.D., is professor and director of the Division of Epidemiology at Vanderbilt University School of Medicine. He also serves as the associate director for Population Science Research at the Vanderbilt-Ingram Cancer Center. Zheng has published more than 1,200 research papers and served as the principal investigator for more than 35 NIH-funded large epidemiologic and genetic studies, including three large prospective cohort studies including over 200,000 study participants. His research focuses on nutrition and the molecular and genetic epidemiology of cancer.

Appendix 2: Workshop Agenda

All times are Eastern Daylight Time (EDT)

Day 1: Thursday, June 29, 2023

11:00 a.m. **Welcome and Introductory Remarks**

- Daniel Shaughnessy, Ph.D., National Institute of Environmental Health Sciences (NIEHS)
- Trevor Archer, Ph.D., Deputy Director, Distinguished Investigator, NIEHS
- Ron Johnson, Ph.D., National Cancer Institute (NCI)
- Dan Gallahan, Ph.D., Director, Division of Cancer Biology, NCI

11:15 a.m. **Workshop Structure and Goals**

- Cheryl Walker, Ph.D., Baylor College of Medicine
- Hannah Carter, Ph.D., University of California, San Diego
- Daniel Shaughnessy, Ph.D., NIEHS
- Ron Johnson, Ph.D., NCI

11:30 a.m. **Session 1: Mutational Signatures of Exposure in Cancer**
SESSION HOST: Arun Pandiri, Ph.D., NIEHS
MODERATOR: Ludmil Alexandrov, Ph.D., University of California, San Diego
State of the Science: Emerging Opportunities and Caveats of Using Mutational Signatures of Environmental Exposures

- Serena Nik-Zainal, Ph.D., University of Cambridge

Synopsis Talks

- Paul Brennan, Ph.D., International Agency for Research on Cancer
- Maria Teresa Landi, M.D., Ph.D., NCI
- Allan Balmain, Ph.D., FRS, University of California, San Francisco

Panel Discussion

12:40 p.m. **Break**

1:20 p.m. **Session 2: Other Data Types as Signatures of Exposure in Cancer**
SESSION HOST: Phil Daschner, M.Sc., NCI
MODERATOR: Scott Auerbach, Ph.D., NIEHS
State of the Science: Other Data Types as Signatures of Exposure in Cancer

- Ting Wang, Ph.D., Washington University

Synopsis Talks

- Cathrine Hoyo, Ph.D., North Carolina State University
- Ben Gewurz, M.D., Ph.D., Harvard University

- Michael Fischbach, Ph.D., Stanford University

Panel Discussion

2:30 p.m. **Break**

2:40 p.m. **Session 3: Computational Challenges and Integrating Multi-Omics to Identify Signatures**

SESSION HOST: Daniel Shaughnessy, Ph.D., NIEHS

MODERATOR: Mona Singh, Ph.D., Princeton University

State of the Science: Why Networks Matter: Embracing Biological Complexity

- John Quackenbush, Ph.D., Harvard University

Synopsis Talks

- Teresa Przytycka, Ph.D., National Library of Medicine
- Joshua Campbell, Ph.D., Boston University
- Mark Gerstein, Ph.D., Yale University

Panel Discussion

3:50 p.m. **Close of Day 1**

Day 2: Friday June 30, 2023

11:00 a.m. **Welcome**

- Michelle Heacock, Ph.D., NIEHS

11:05 a.m. **Session 4: Challenges in Tracking Signatures of Exposures**

SESSION HOST: Michelle Heacock, Ph.D., NIEHS

MODERATOR: Rebecca Fry, Ph.D., University of North Carolina, Chapel Hill

State of the Science: Inducing Mutational Signatures by Environmental Carcinogens and Chemotherapeutic Agents In Vitro: Progress, Prospects, and Limitations

- David Phillips, Ph.D., King's College, London

Synopsis Talks

- John Essigmann, Ph.D., Massachusetts Institute of Technology
- Yvonne Fondufe-Mittendorf, Ph.D., Van Andel Institute
- Ana Navas-Acien, M.D., Ph.D., Columbia University

Panel Discussion

12:15 p.m. **Break**

12:30 p.m. **Session 5: Population-Based Cancer Studies**

SESSION HOST: Somdat Mahabir, Ph.D., MPH, NCI

MODERATOR: Paul Brennan, Ph.D., International Agency for Research on Cancer

State of the Science: Use of Molecular Signatures of Exposures in Epidemiologic Studies of Cancer

- Wei Zheng, M.D., Ph.D., Vanderbilt University

Synopsis Talks

- Jesse Goodrich, Ph.D., University of Southern California
- Karin Michels, Sc.D., Ph.D., University of California, Los Angeles
- Loïc Le Marchand, M.D., Ph.D., University of Hawai'i

Panel Discussion

1:40 p.m. **Break**

2:20 p.m. **Workshop Summary and Future Directions**

SESSION HOST: Ron Johnson, Ph.D., NCI

MODERATORS: Cheryl Walker, Ph.D., Baylor College of Medicine and Hannah Carter, Ph.D., University of California, San Diego

Session Summaries

- Ludmil Alexandrov, Ph.D., University of California, San Diego
- Scott Auerbach, Ph.D., NIEHS
- Mona Singh, Ph.D., Princeton University
- Rebecca Fry, Ph.D., University of North Carolina, Chapel Hill
- Stephen Chanock, M.D., NCI

Future Directions

Panel discussion among all invited participants moderated by co-chairs

3:30 p.m. **Closing Remarks**

- Hannah Carter, Ph.D., University of California, San Diego
- Cheryl Walker, Ph.D., Baylor College of Medicine
- Ron Johnson, Ph.D., NCI
- Daniel Shaughnessy, Ph.D., NIEHS

Appendix 3: Key Publications

Session 1: Mutational Signatures of Exposure in Cancer

Alexandrov, L.B., et al., *Deciphering signatures of mutational processes operative in human cancer*. Cell Rep, 2013. 3(1): p. 246-259.

Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer*. Nature, 2013. 500(7463): p. 415-421.

Alexandrov, L.B., et al., *The repertoire of mutational signatures in human cancer*. Nature, 2020. 578(7793): p. 94-101.

Kim, Y.A., et al., *Mutational Signatures as Sensors of Environmental Exposures: Analysis of Smoking-Induced Lung Tissue Remodeling.* Biomolecules, 2022. 12(10): p. 1384.

Kim, Y.A., et al., *Mutational Signatures: From Methods to Mechanisms.* Annu Rev Biomed Data Sci, 2021. 4: p. 189-206.

Wojtowicz, D., et al., *Hidden Markov models lead to higher resolution maps of mutation signature activity in cancer.* Genome Med, 2019. 11(1): p. 49.

Huang, X., D. Wojtowicz, and T.M. Przytycka, *Detecting presence of mutational signatures in cancer with confidence.* Bioinformatics, 2018. 34(2): p. 330-337.

Session 2: Other Data Types as Signatures of Exposure in Cancer

Johnson, K.J., et al., *A Transformative Vision for an Omics-Based Regulatory Chemical Testing Paradigm*. Toxicol Sci, 2022. 190(2): p. 127-132.

Corton, J.C., et al., *A Collaborative Initiative to Establish Genomic Biomarkers for Assessing Tumorigenic Potential to Reduce Reliance on Conventional Rodent Carcinogenicity Studies*. Toxicol Sci, 2022. 188(1): p. 4-16.

Gwinn, W.M., et al., *Evaluation of 5-day in vivo rat liver and kidney with high-throughput transcriptomics for estimating benchmark doses of apical outcomes*. Toxicol Sci, 2020. 176(2): p. 343—354.

Ramaiahgari, S.C., et al., *The power of resolution: contextualized understanding of biological responses to liver injury chemicals using high-throughput transcriptomics and benchmark concentration modeling*. Toxicol Sci, 2019. 169(2): p. 553—566.

Dekkers, K.F., et al., *An online atlas of human plasma metabolite signatures of gut microbiome composition.* Nat Commun, 2022. 13(1): p. 5370.

Bokulich, N.A., et al., *Multi-omics data integration reveals metabolome as the top predictor of the cervicovaginal microenvironment.* PLoS Comput Biol, 2022. 18(2): p. e1009876.

Hofseth, L.J., et al., *Early-onset colorectal cancer: initial clues and current views.* Nat Rev Gastroenterol Hepatol, 2020. 17(6): p. 352-364.

Ugai, T., et al., *Is early-onset cancer an emerging global epidemic? Current evidence and future implications.* Nat Rev Clin Oncol, 2022. 19(10): p. 656-673.

Bessonneau, V. and Rudel, R.A., *Mapping the Human Exposome to Uncover the Causes of Breast Cancer.* Int J Environ Res Public Health, 2019. 17(1): p. 189.

Sud, A., Turnbull, C., and Houlston, R, *Will polygenic risk scores for cancer ever be clinically useful?* NPJ Precis Oncol, 2021. 5(1): p. 40.

Yao, S., et al., *Proceedings of the fifth international Molecular Pathological Epidemiology (MPE) meeting.* Cancer Causes Control, 2022. 33(8): p. 1107-1120.

Advancing cancer research with genetic analysis tools. ThermoFisher Scientific, 2022.

Brockway-Lunardi, L., et al., *Early-onset colorectal cancer research: gaps and opportunities.* Colorect Cancer, 2020.

Wang, M., et al., *Strain dropouts reveal interactions that govern the metabolic output of the gut microbiome.* Cell, 2023. 186(13): p. 2839-2852.

Weeden, C.E., et al., *Impact of risk factors on early cancer evolution.* Cell, 2023. 186(8): p. 1541-1563.

Session 3: Computational Challenges and Integrating Multi-Omics to Identify Signatures

Campbell, J.D., et al., *Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas.* Nat Genet, 2016. 48(6): p. 607-616.

Campbell, J.D., et al., *The Case for a Pre-Cancer Genome Atlas (PCGA).* Cancer Prev Res (Phila), 2016. 9(2): p. 119-124.

Campbell, J.D., et al., *Assessment of microRNA differential expression and detection in multiplexed small RNA sequencing data.* RNA, 2015. 21(2): 164-171.

Cancer Genome Atlas Research Network. *Comprehensive molecular profiling of lung adenocarcinoma.* Nature, 2014. 511(7511): p. 543-550.
-Roles: DNA-sequencing team and manuscript coordinator

Perdomo, C., et al., *MicroRNA 4423 is a primate-specific regulator of airway epithelial cell differentiation and lung carcinogenesis.* Proc Natl Acad Sci U S A, 2013. 110(47): p.18946-18951.

Campbell, J.D., et al., *A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK.* Genome Med, 2012. 4(8): p. 67.

Li, S., F.W. Crawford, and M.B. Gerstein, *Using sigLASSO to optimize cancer mutation signatures jointly with sampling likelihood*. Nat Commun, 2020. 11(1): p. 3575.

Wojtowicz, D., et al., *RepairSig: Deconvolution of DNA damage and repair contributions to the mutational landscape of cancer.* Cell Syst, 2021. 12(10): p. 994-1003.

Wojtowicz, D., et al., *DNA Repair Footprint Uncovers Contributions of DNA Repair Mechanism to Mutational Signatures.* Pac Symp Biocomput, 2020. 25: p. 262-273.

Amgalan, B., et al., *Influence network model uncovers relations between biological processes and mutational signatures.* Genome Med, 2023. 15(1): p. 1-15.

Kim, Y.A., et al., *Network-based approaches eludicate differences within APOBEC and clock-like signatures in breast cancer.* Genome Med, 2020. 12(1): p. 52.

Chevalier, A., et al., *The Mutational Signature Comprehensive Analysis Toolkit (musicatk) for the Discovery, Prediction, and Exploration of Mutational Signatures.* Cancer Res, 2021. 81(23): p. 5813-5817.

Hill, W., et al., *Lung adenocarcinoma promotion by air pollutants.* Nature, 2023. 616(7955): p. 159-167.

Duclos, G.E., et al., *Characterizing smoking-induced transcriptional heterogeneity in the human bronchial epithelium at single-cell resolution.* Sci Adv, 2019. 5(12): p. eaaw3413.

Session 4: Challenges in Tracking Signatures of Exposures

Chawanthayatham, S., et al., *Mutational spectra of aflatoxin B_1 in vivo establish biomarkers of exposure for human hepatocellular carcinoma*. Proc Natl Acad Sci USA, 2017. 114(15): p. E3101-E3109.

Armijo, A.L., et al., *Molecular origins of mutational spectra produced by the environmental carcinogen N-nitrosodimethylamine and S_N1 chemotherapeutic agents*. NAR Cancer, 2023. 5(2).

Fedeles, B.I. and J.M. Essigmann, *Mutational spectra provide insight into the mechanisms bridging DNA damage to genetic disease*. DNA Damage, DNA Repair and Disease, Vol. 2, 2020. p. 214-253.

George, S., et al., *Epigenomic reprogramming in iAs-mediated carcinogenesis*. Adv Pharmacol, 2023. 96: p. 319-365.

Saintilnord, W.N., et al., *Chronic Exposure to Cadmium Induces Differential Methylation in Mice Spermatozoa*. Toxicol Sci, 2021. 180(2): p. 262-276.

Eckstein, M., M. Rea, and Y.N. Fondufe-Mittendorf, *Transient and permanent changes in DNA methylation patterns in inorganic arsenic-mediated epithelial-to-mesenchymal transition*. Toxicol Appl Pharmacol, 2017. 331: p. 6-17.

Rea, M., T. Gripshover, and Y.N. Fondufe-Mittendorf, *Selective inhibition of CTCF binding by iAs directs TET-mediated reprogramming of 5-hydroxymethylation patterns in iAs-transformed cells*. Toxicol Appl Pharmacol, 2018. 338: p. 124-133.

Eleazer, R., et al., *PARP1 Regulates Circular RNA Biogenesis though Control of Transcriptional Dynamics*. Cells, 2023. 12(8): p. 1160.

Chang, Z., et al., *The relationship between co-exposure to multiple heavy metals and liver damage*. J Trace Elem Med Biol, 2023. 77: p. 127128.

Dye, C.K., et al., *Maternal DNA methylation signatures of arsenic exposure is associated with adult offspring insulin resistance in the Strong Heart Study.* Environ Int, 2023. 173: p. 107774.

Chernoff, M.B., et al., *Sequencing-based fine-mapping and in silico functional characterization of the 10q24.32 arsenic metabolism efficiency locus across multiple arsenic-exposed populations.* PLoS Genet, 2023. 19(1): p. e1010588.

Zacher, T., et al., *Evaluation of a water arsenic filter in a participatory intervention to reduce arsenic exposure in American Indian communities: The Strong Heart Water Study.* Sci Total Environ, 2023. 862: p. 160217.

Lieberman-Cribbin, W., et al., *Epigenetic Biomarkers of Lead Exposure and Cardiovascular Disease: Prospective Evidence in the Strong Heart Study.* J Am Heart Assoc, 2022. 11(23): p. e026934.

Domingo-Relloso, A., et al., *Arsenic Exposure, Blood DNA Methylation, and Cardiovascular Disease.* Circ Res, 2022. 131(2): p. e51-e69.

Kuo, C.C., et al., *The association of arsenic exposure and arsenic metabolism with all-cause, cardiovascular and cancer mortality in the Strong Heart Study.* Environ Int, 2022. 159: p. 107029.

Session 5: Population-Based Cancer Studies

Goodrich, J.A., et al., *Exposure to perfluoroalkyl substances and risk of hepatocellular carcinoma in a multiethnic cohort*. JHEP Rep, 2022. 4(10): p. 100550.

Shearer, J.J., et al., *Serum Concentrations of Per- and Polyfluoroalkyl Substances and Risk of Renal Cell Carcinoma*. J Natl Cancer Inst, 2021. 113(5): p. 580-587.

Goodrich, J.A., et al., *Metabolic Signatures of Youth Exposure to Mixtures of Per- and Polyfluoroalkyl Substances: A Multi-Cohort Study*. Environ Health Perspect, 2023. 131(2): p. 27005.

Kucab, J.E., et al., *A Compendium of Mutational Signatures of Environmental Agents*. Cell, 2019. 177(4): p. 821-836.

Maitre, L., et al., *Multi-omics signatures of the human early life exposome*. Nat Commun, 2022. 13(1): p. 7024.

Ladd-Acosta, C., *Epigenetic Signatures as Biomarkers of Exposure*. Curr Environ Health Rep, 2015. 2(2): p. 117-125.

Huber, A.R., A.V. Hoeck, and R.V. Boxtel, *The Mutagenic Impact of Environmental Exposures in Human Cells and Cancer: Imprints Through Time*. Front Genet, 2021. 12: p. 760039.

Xue F. and Michels K.B., *Intrauterine factors and risk of breast cancer: a systematic review and meta-analysis of current evidence*. Lancet Oncol, 2007. 8(12): p. 1088-1100.

Carwile, J.L., et al., *Polycarbonate bottle use and urinary bisphenol A concentrations*. Environ Health Perspect, 2009. 117(9): p. 1368-1372.

Carwile, J.L., et al., *Canned soup consumption and urinary bisphenol A: a randomized crossover trial*. Jama, 2011. 306(20): p. 2218-2020.

Michels, K.B., et al., *Recommendations for the design and analysis of epigenome-wide association studies*. Nat Methods, 2013. 10(10): p. 949-955.

Rancourt, R.C., et al., *The prevalence of loss of imprinting of H19 and IGF2 at birth*. FASEB J, 2013. 27(8): p. 3335-3343.

Harris, H.R., et al., *An Adolescent and Early Adulthood Dietary Pattern Associated with Inflammation and the Incidence of Breast Cancer*. Cancer Res, 2017. 77(5): p. 1179-1187.

Gaskins, A.J., et al., *Dairy intake in relation to breast and pubertal development in Chilean girls.* Am J Clin Nutr, 2017. 105(5): p. 1166-1175.

Binder, A.M., et al., *Faster ticking rate of the epigenetic clock is associated with faster pubertal development in girls.* Epigenetics, 2018. 13(1): p. 85-94.

Terry, M.B., et al., *Environmental exposures during windows of susceptibility for breast cancer: a framework for prevention research.* Breast Cancer Res, 2019. 21(1): p. 96.

Yoon, L.S., et al., *Variability in urinary phthalates, phenols, and parabens across childhood and relation to adolescent breast composition in Chilean girls.* Environ Int, 2022. 170: p. 107586.