



Proceedings of the fourth international molecular pathological epidemiology (MPE) meeting

Peter T. Campbell¹ · Christine B. Ambrosone² · Reiko Nishihara^{3,4,5} · Hugo J. W. L. Aerts⁶ · Melissa Bondy⁷ · Nilanjan Chatterjee⁸ · Montserrat Garcia-Closas⁹ · Marios Giannakis^{10,11} · Jeffrey A. Golden¹² · Yujing J. Heng¹³ · N. Sertac Kip¹⁴ · Jill Koshiol⁹ · X. Shirley Liu¹⁵ · Camila M. Lopes-Ramos⁵ · Lorelei A. Mucci¹⁶ · Jonathan A. Nowak¹² · Amanda I. Phipps¹⁷ · John Quackenbush⁵ · Robert E. Schoen¹⁸ · Lynette M. Sholl¹² · Rulla M. Tamimi^{16,19} · Molin Wang^{5,16,19} · Matty P. Weijnen²⁰ · Catherine J. Wu^{10,11} · Kana Wu⁴ · Song Yao² · Kun-Hsing Yu^{12,21} · Xuehong Zhang¹⁹ · Timothy R. Rebbeck^{10,16} · Shuji Ogino^{3,11,16,22}

Received: 28 September 2018 / Accepted: 27 April 2019 / Published online: 8 May 2019
© Springer Nature Switzerland AG 2019

Abstract

An important premise of epidemiology is that individuals with the same disease share similar underlying etiologies and clinical outcomes. In the past few decades, our knowledge of disease pathogenesis has improved, and disease classification systems have evolved to the point where no complex disease processes are considered homogenous. As a result, pathology and epidemiology have been integrated into the single, unified field of molecular pathological epidemiology (MPE). Advancing integrative molecular and population-level health sciences and addressing the unique research challenges specific to the field of MPE necessitates assembling experts in diverse fields, including epidemiology, pathology, biostatistics, computational biology, bioinformatics, genomics, immunology, and nutritional and environmental sciences. Integrating these seemingly divergent fields can lead to a greater understanding of pathogenic processes. The International MPE Meeting Series fosters discussion that addresses the specific research questions and challenges in this emerging field. The purpose of the meeting series is to: discuss novel methods to integrate pathology and epidemiology; discuss studies that provide pathogenic insights into population impact; and educate next-generation scientists. Herein, we share the proceedings of the Fourth International MPE Meeting, held in Boston, MA, USA, on 30 May–1 June, 2018. Major themes of this meeting included ‘integrated genetic and molecular pathologic epidemiology’, ‘immunology-MPE’, and ‘novel disease phenotyping’. The key priority areas for future research identified by meeting attendees included integration of tumor immunology and cancer disparities into epidemiologic studies, further collaboration between computational and population-level scientists to gain new insight on exposure-disease associations, and future pooling projects of studies with comparable data.

Peter T. Campbell, Christine B. Ambrosone, Timothy R. Rebbeck and Shuji Ogino have contributed equally to this work as the Co-Chairs of the meeting.

Use of Standardized Official Symbols For unambiguous communication, we use HUGO (Human Genome Organisation) Gene Nomenclature Committee-approved official symbols (or root symbols) for genes, gene products, and gene families, including APC, AR, AXIN2, BRAF, CD4, CD8, CTLA4, CTNNA1, EGFR, ERBB2, ERG, ESR1, FOXA1, FDX1, HLA, IDH1, IDH2, IGF1, IGF1R, IGF2, KRAS, MUC1, MKI67, PDCD1, PGR, PTEN, RB1, TERT, TM6SF2, and WNT; all of which are described at www.genenames.org. Each colloquial name is used in parenthesis following its official symbol counterpart. This format enables readers to familiarize themselves with the official symbols for genes and gene products together with common colloquial names.

Extended author information available on the last page of the article

Keywords Molecular pathological epidemiology · Meeting report · Meeting proceedings · Meeting summary · Patho-epidemiology

Abbreviations

AI	Artificial intelligence
AICR	American Institute for Cancer Research
BCAC	Breast cancer association consortium
BCR	B cell receptor
CDR	Complementarity determining region
CIMP	CpG island methylator phenotype
cfDNA	Cell free DNA
ctDNA	Circulating tumor DNA
ddPCR	Droplet digital PCR
eQTL	Expression quantitative trait loci
FFPE	Formalin-fixed paraffin-embedded
GBM	Glioblastoma multiforme
GWAS	Genome-wide association study
HPFS	Health professionals follow-up study
IHC	Immunohistochemistry
MPE	Molecular pathological epidemiology
MSI	Microsatellite instability
NHS	Nurses' Health Study
PCR	Polymerase chain reaction
PRS	Polygenic risk score
RTK	Receptor tyrosine kinase
SNP	Single nucleotide polymorphism
TCGA	The Cancer Genome Atlas
TCR	T cell receptor
TMA	Tissue microarray
WCHS	Women's Circle of Health Study
WCRF	World Cancer Research Fund
WES	Whole exome sequencing

Introduction

A central objective of epidemiology is to investigate why some diseases (or related health events) occur in certain groups of people but not in others, with the view to applying that knowledge to prevent or control those outcomes in the future. Traditionally in this effort, diseases or health events are treated as binary outcomes; for example, a group of people either have or do not have cancer at a given organ site and, if etiology is of interest, rates of that outcome can be compared across levels of a potential exposure. While this approach has identified major causes of morbidity and mortality in human populations that have subsequently supported important public health policy changes, the over-simplified 'yes vs no' nosology misses the essential concept of pathogenic heterogeneity. Most human diseases, and essentially all cancers, are biologically different from one patient to the next. Indeed, many human diseases are complex

processes that occur for reasons that are, in a precise sense, unique to that individual, the result of that person's specific host characteristics (e.g., genome) and a multitude of distinct external factors (e.g., diet, lifestyle, environmental, microbiome). Molecular pathological epidemiology (MPE) is an integrative scientific discipline that examines the interplay of these unique disease, host, and external factors.

One of the main aims of MPE is to investigate potential etiologic/survival factors across strata of molecular characteristics for the disease-of-interest. The underlying premise with an etiologic study in MPE is that diseases that share certain molecular alterations are more likely to share common causes; similarly, for prognosis studies, the general hypothesis is that some external or endogenous factors may influence disease outcomes according to molecular attributes because those factors likely interact with the diseased cells in the local tissue microenvironment. Historically, these molecular classifications were often drawn from the clinic, such as microsatellite instability (MSI) for colorectal cancer and statuses of ESR1 (estrogen receptor 1, ER), PGR (progesterone receptor, PR), and ERBB2 (HER-2) for breast cancer [1–6]. With the recent growth of high-throughput biologic data in large epidemiology studies, disease phenotyping has become more sophisticated and includes tumor sequencing, gene expression, proteomics, and epigenomics.

The International MPE Meeting Series began in April 2013 as a small, local meeting of 10 investigators at the Harvard School of Public Health. Subsequent meetings became larger, with over 150 scientists from more than 16 countries attending each of the Second (December 2014) [7] and Third (May 2016) [8] International MPE meetings. Because MPE is inherently transdisciplinary and it is a relatively new scientific discipline, these meetings gave attendees a rare opportunity to share ideas, methods, successes, and challenges; further, they were an opportunity to help train the next generation of MPE scientists. Herein, we share the proceedings of the Fourth International Molecular Pathological Epidemiology (MPE) Meeting, held in May/June 2018 at Dana-Farber Cancer Institute in Boston, MA, USA. A list of the speaker names, lecture titles, and key references appears in Table 1.

30 May 2018

For the first time in the history of the International MPE Meeting Series, a pre-meeting interactive workshop was held for current and future leaders in MPE. Dr. Reiko Nishihara chaired the session which included panelists from a

Table 1 Summary of podium presentations at the 4th international molecular pathological epidemiology (MPE) meeting in Boston, MA, on 31 May and 1 June, 2018

Session and speaker name	Topic	Key references
<i>Integrated genetic, epidemiologic, and tumor analyses, part 1</i>		
Lorelei Mucci	Integrating tissue biomarkers into prostate cancer epidemiology studies	9–11
Nilanjan Chatterjee	Genetic association testing in the presence of tumor heterogeneity	12, 13
Melissa Bondy	Glioma: Insights from molecular epidemiology	14
<i>Integrated genetic, epidemiologic, and tumor analyses, part 2</i>		
Montserrat Garcia-Closas	Genetic susceptibility to breast cancer subtypes: implications for screening/prevention	12, 13, 15–17
Rulla Tamimi	Leveraging non-tumor tissue sources to understand breast cancer etiology	18–21
Christine Ambrosone	Molecular alterations in breast tumor tissues from African American women: relationships with risk factors	22–24
<i>Integrated genetic, epidemiologic, and tumor analyses, part 3</i>		
Matty Weijenberg	Importance of molecular pathological epidemiology for examining lifestyle and cancer	25–29
Timothy Rebbeck	Molecular signatures of prostate cancer disparities	30, 31
<i>Immunology and immunotherapy/prevention, part 1</i>		
Shuji Ogino	Need for integrative analyses of exposome (including microbiome)-tumor-immune interactions	32–38
X. Shirley Liu	Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response	39–44
Robert ‘Rocky’ Schoen	Prevention of colorectal adenoma with MUC1 vaccine	45
<i>Immunology and immunotherapy/prevention, part 2</i>		
Catherine Wu	Identifying and therapeutically targeting tumor neoantigens	46–49
Marios Giannakis	Genomic mechanisms of immune evasion in colorectal cancer	50–53
<i>Novel disease phenotyping in future medicine and population science, part 1</i>		
Jeffrey Golden	Computational pathology to precision medicine: a working model	54–56
Lynette Sholl	Liquid biopsy for solid tumors: current realities and future possibilities	57, 58
Jan Heng	Comparison of pre-processing methods to compute PAM50 intrinsic subtype in breast cancer and normal tumor-adjacent regions	59–61
<i>Novel disease phenotyping in future medicine and population science, part 2</i>		
John Quackenbush	Using networks to understand the genotype–phenotype connection	62–65
Hugo Aerts	Artificial intelligence in medical imaging	66–69
<i>Special topic lectures</i>		
N. Sertac Kip	How solid is a liquid biopsy?	70–75
Camila Lopes-Ramos	Regulatory networks identify sex differences in colon cancer	76
Molin Wang	Addressing sample selection bias due to tissue availability	77
Song Yao	Population differences in breast tumor immune microenvironment between women of African ancestry and European ancestry	78–81
Kun-Hsing Yu	Integrating lung cancer omics and histopathology profiles for precision medicine	82–86

wide range of career stages and institutions (Drs Christine Ambrosone, Peter Campbell, Montserrat Garcia-Closas, Marios Giannakis, John Quackenbush, and Molin Wang). The workshop consisted of questions from the session chair and audience members with responses from panelists concerning grant writing for transdisciplinary science, transdisciplinary team building, and training opportunities. We discussed common mistakes in grant applications, development of research questions for grants and manuscripts, career development of transdisciplinary expertise, areas of training that are fundamentally important for trainees to pursue MPE research, and tips for successful collaborations across disciplines.

30 May 2018

Session 1: Integrated genetic, epidemiologic, and tumor analyses #1 (Session Chair: Dr. Peter Campbell)

The first speaker of the meeting was Dr. Lorelei Mucci who presented an overview of her group's work on integrating tissue biomarkers into prostate cancer epidemiology studies. She focused on two common molecular subtypes: the androgen-regulated gene fusion *TMPRSS2:ERG* and loss of the tumor suppressor *PTEN*. This work leverages prostate tissue biorepositories nested within the Physicians' Health Study (PHS) and Health Professionals Follow-up Study (HPFS). The first set of studies focused on the high heritability of prostate cancer. Inherited variation within the androgen receptor (*AR*), which regulates *AR* expression, was associated with *ERG*-positive cancer but not *ERG*-negative disease [9]. Of 39 inherited prostate cancer risk loci, ten were differentially associated with risk when stratified by *ERG* status [10]. *ERG*-positive cancers show higher expression of the insulin and IGF1 receptors [11]. In unpublished data, vigorous physical activity was associated with a significantly lower risk of *ERG*-positive disease, whereas there was no association with *ERG*-negative disease. Finally, use of the cholesterol lowering drug, statins, was associated with a substantially lower risk of tumors showing *PTEN* loss. These data highlight the etiologic heterogeneity of prostate cancer, and the opportunities to elucidate discoveries based on integrating tissue biomarkers.

Dr. Nilanjan Chatterjee lectured on the development and utility of a broad mixed-effect two-stage logistic regression model for discovering new breast cancer germline genetic risk loci in the context of tumor heterogeneity [12, 13]. Breast cancers are highly heterogeneous, and it may become quickly inefficient to evaluate each potential risk locus with each different disease sub-type. With genome-wide association study (GWAS) data from the Breast Cancer Association

Consortium (BCAC) that includes nearly 100,000 controls and a little over 100,000 cases, approximately 180 SNPs were discovered for breast cancer risk overall. When extending this work to subtype-specific analyses, a two-stage logistic regression approach was preferred over standard analyses to account for the large number of comparisons, correlations between markers, missing marker data, and other reasons. Case-control and case-case odds ratios were calculated using the mixed-effect two-stage approach, and 11 novel SNPs for subtype-specific breast cancers were identified. One of the more interesting findings from this work was the discovery that a *TP53* SNP was associated with increased risk of luminal breast cancer and with decreased risk of triple negative tumors, probably reflecting the different pathologic mechanisms that drive these different tumor sub-types and potentially underscoring the broad utility of this statistical approach.

In the last lecture of the first session, Dr. Melissa Bondy spoke about GWAS results for gliomas, overall and when stratified by histological and molecular subtypes [14]. Brain tumors, which comprise a highly heterogeneous group of cancers, account for 1–2% of all cancers overall. To better understand germline genetic risk factors for glioma, Dr. Bondy and collaborators pooled data from eight independent glioma GWAS datasets. Their combined meta-analysis identified 13 novel glioma risk loci (five for glioblastoma multiforme (GBM) and eight for non-GBM). Curiously, all but one locus showed significant allele frequency differences between GBM and non-GBM tumors. The only locus consistently associated with glioma risk was at 17p13.1 (*TP53*). Overall, genetic heritability is estimated to account for approximately 1/3 of the population variability in glioma. Further work showed a high degree of concordance, according to germline genetics, between familial and sporadic gliomas, suggesting genetic predisposition is largely the same for both types of disease. Many of the risk SNPs were in DNA-repair or telomere maintenance-related pathways. Future work in this area will include additional, detailed work within core pathways, sequencing efforts, and gene-environment interaction.

Session 2: Integrated genetic, epidemiologic, and tumor analyses #2 (Session Chair: Dr. Kana Wu)

The second morning session continued with the earlier theme of integrated studies in cancer epidemiology. Dr. Montserrat 'Montse' Garcia-Closas's lecture gave an overview of recent developments in risk factor identification for breast cancer according to clinical subtype and the implications for those results on personal risk prediction. Breast cancer is a heterogeneous disease with different survival outcomes. The most aggressive tumors are hormone receptor negative tumors, including triple negative tumors that

are also negative for ERBB2 (HER2) amplification. These phenotypes represent approximately 13% of tumors and are more common in younger women and women of African descent. GWAS have identified nearly 180 germline loci for breast cancer risk overall. More recent work has identified clusters of SNPs associated with specific tumor features, including ESR-1 status and grade. Dr. Garcia-Closas and colleagues have developed Polygenic Risk Scores (PRS) for ESR-1 + and ESR-1-negative specific disease and are integrating them into risk models to identify women at different risks of breast cancer overall, and by subtypes; however, more work on implementation is needed before translating these tools into clinical practice [12, 13, 15–17].

Dr. Rulla Tamimi described the role of early life risk factors on markers in non-tumor tissues. Many risk factors for breast cancer in adults occur early in life. More specifically, there are windows of susceptibility that may be relevant to specific tissue markers of disease processes. Her group has conducted work on early life body size in the Nurses' Health Studies (NHS) that asked women about their body size at ages 5, 10, and 20. They found that larger body size in early life was inversely associated with proliferative benign breast disease and breast cancer risk [18–20]. In a study of normal breast tissue adjacent to benign breast disease lesions, they found women who reported as being heavier early in life (ages 5–10) had a reduction of MKI67 (Ki67) expression [21]. In additional work, her group has examined breast tissue gene expression in the Nurses' Health Studies using a transcriptome array. In preliminary work, they have seen differences in gene expression patterns related to exposures when considering tumor and adjacent normal tissue. Dr. Tamimi echoed a theme that was common during this year's meeting: the need for more consortia work for validation, replication, and new discovery.

Dr. Christine Ambrosone discussed her research focused on understanding the etiology of more aggressive breast cancers in African American women, particularly tumors that do not express ESR1 (estrogen receptor, ER) that are associated with poorer prognosis [22]. She described results from a study of DNA methylation in tumors from African-American and European-American women that found that methylation of a gene important to guiding the luminal phenotype, *FOXA1*, was greatest in ESR1 (ER)-negative breast cancers [23]. Methylation was more common in women who had children and did not breastfeed, suggesting a mechanism for the increased risk of ESR1 (ER)-negative breast cancer with parity, and not breastfeeding [24].

Session 3: Integrated genetic, epidemiologic, and tumor analyses #3 (Session Chair: Dr. Song Yao)

After a 2-h pause in podium presentations for attendees to interact and to view poster presentations, the afternoon sessions began with the third installment of the integrated epidemiology theme.

The first lecture of the afternoon session was presented by Dr. Matty Weijenberg who discussed the role of MPE in supporting lifestyle guidelines for cancer prevention. She illustrated how MPE could contribute to the research directions provided by the third expert report from the World Cancer Research Fund (WCRF)/American Institute for Cancer Research (AICR) on diet, nutrition, physical activity, and cancer. First, she showed examples of how mutational spectra of tumors can provide clues to mechanisms, for example how heme iron intake is associated with specific G to A mutations in *KRAS* genes in colorectal tumors, pointing to the role of alkylating agents [25]. More recent studies suggest that tumor signatures of mutational processes are associated with exogenous mutational processes [26]. Second, she showed how exposure to the Dutch Hunger Winter was associated with a reduced risk of colorectal cancer only in tumors with a CpG-island methylator phenotype (CIMP) [27] and tumors with increasing number of *IGF1P* genes methylated [28]. Third, she showed how investigating subtypes of tumors can reveal previously unknown etiologies. For example, results from a meta-analysis revealed how adherence to a Mediterranean diet is associated specifically with a reduced risk of ESR1 (ER)-negative and PGR (PR)-negative postmenopausal breast cancer [29].

In the final lecture of the 'integrated epidemiology' sessions, Dr. Timothy Rebbeck discussed the role of germline genetics in explaining some disparities for prostate cancer [30, 31]. Prostate cancer has higher incidence and mortality rates in African American men compared to all other race/ethnic groups. Many prostate cancer susceptibility loci have been identified via GWAS for prostate cancer overall. But the contribution of these loci to prostate cancer disparities is unclear. To address this issue, Dr. Rebbeck's group evaluated the population structure of 68 previously identified prostate cancer susceptibility loci by calculating: (1) genetic disparity contribution statistics to quantify the contribution of each SNP to differences in prostate cancer risk across populations, and (2) genetic risk scores that integrate GWAS results with allele frequency data from 45 African and 19 non-African populations. They found that predicted prostate cancer risks were highest for men of West African descent and lowest for men of East Asian descent. These population-level differences were further explained by the out-of-Africa bottleneck and natural selection. Only a few loci seemed to drive the excess prostate cancer risk observed in African American men. Although most prostate cancer

susceptibility loci are evolving neutrally across different race/ethnic groups, there are several instances where alleles have hitchhiked at higher frequencies with adaptive alleles, including alleles for skin pigmentation (at 2q37).

Session 4: Immunology, immunotherapy, and prevention #1 (Session Chair: Dr. Amanda Phipps)

The first lecture in the ‘immunology-MPE’ session was given by Dr. Shuji Ogino who spoke about the broad importance of incorporating immune system data into MPE studies. Despite remarkable advances of cancer immunology in recent years, investigations on the influences of the exposome on tumor-immune interactions lag. To address a substantial gap between cancer immunology and epidemiology, the integrative field of immunology-MPE can investigate influences of the exposome (dietary, lifestyle, environmental, microbial, pharmacological, and other exposures) on tumor-immune interactions [32, 33]. Using epidemiological studies and colorectal cancer cases with data on immune response, tumor molecular pathology, and tissue microorganisms, proof-of-principle immunology-MPE studies provide evidence supporting hypotheses that several exposures influence carcinogenic processes through their influences on tumor-immune interactions [34–38]. For instance, marine omega-3 polyunsaturated fatty acid intake has been associated with a lower risk of colorectal carcinoma containing abundant FOXP3⁺ cells (mostly regulatory T cells) [34] and a prudent dietary pattern has been associated with a lower risk of colorectal carcinoma containing abundant *Fusobacterium nucleatum* [38]. These new insights from immunology-MPE research can provide a possible path for precision immunoprevention and immunotherapy.

Dr. X. Shirley Liu introduced her work in mining and integrating large-scale tumor molecular profiles to inform cancer immunology and immunotherapy. Dr. Liu discussed three algorithms that her laboratory developed to extract useful insights from treatment-naïve RNA-seq samples in The Cancer Genome Atlas (TCGA). First, ‘TIMER’ can estimate immune cell components in tumors [39], and a webserver was created for users to explore immune infiltration across TCGA tumors and to draw inferences on user-provided samples [40]. Second, ‘TRUST’ can assemble T cell receptor (TCR) and B cell receptor (BCR) complementarity-determining regions (CDR3s) from bulk tumor RNA-seq data [41]. When applied to over 10,000 samples in the TCGA, TRUST assembled 3 M TCR CDR3 sequences from tumor RNA-seq samples and revealed associations between tumor infiltrating TCR clonotype diversity and tumor mutational load [42]. TRUST also identified 30 M BCR sequences from TCGA tumor RNA-seq and revealed widespread B cell clonal expansions among other events [43]. Third, ‘TIDE’

software derived gene expression signatures from pretreated tumor specimens to predict patient response to anti-PDCD1 (PD-1) and anti-CTLA4 treatment [44]. TIDE analyses of published immune checkpoint inhibitor trials suggested some tumors are unlikely to respond to anti-PDCD1 (PD-1) or anti-CTLA4 alone. This work indicates that tumor RNA-seq, even on treatment naïve tumors, is cost-effective to inform tumor microenvironment and immunity.

Dr. Robert ‘Rocky’ Schoen described his group’s experiences in developing a vaccine for colorectal cancer. Immunotherapy targeting of antigens that are aberrantly expressed on colon cancers and polyps offers the potential for relatively non-invasive, non-toxic, and prolonged preventive strategies. Whereas vaccines in advanced cancers have had little success, likely because of immunosuppressive tumor microenvironments, vaccines administered in pre-malignant stages when the immune system is still powerful should be more effective. Dr. Schoen and colleagues are testing this hypothesis using the adenomatous polyp-to-colon-cancer pathway. Their target is MUC1, a tumor-associated antigen that is abnormally expressed on polyps and colon cancers. In a pilot study of 39 patients with a history of advanced adenomas, after a series of 3 injections, almost half of the patients showed a twofold ratio increase in anti-MUC1 IgG at week 12 compared to pre-vaccination levels. A booster injection at week 52 resulted in a large increase in IgG, demonstrating a persistent T cell memory response [45]. There was minimal evidence of toxicity. A double-blind randomized trial of 110 patients using a similar vaccination protocol is due to report in late 2019. That trial will also evaluate a clinical endpoint, via assessment of adenoma recurrence.

Session 5: Immunology, immunotherapy, and prevention #2 (Session Chair: Dr. Xuehong Zhang)

In the final session of the day, Dr. Catherine Wu lectured on her work in identifying tumor antigens. Multiple lines of evidence demonstrate that tumor neoantigens are an important class of immunogenic antigens. Neoantigens arise from amino acid changes encoded by somatic mutations in the tumor cell. This work has advanced in recent years due to the availability of next-generation sequencing approaches and the maturation of predictive algorithms [46]. One of the central questions of Dr. Wu’s work is: can a personalized cancer vaccine stimulate anti-tumor immunity in humans? Her group conducted a trial in high-risk melanoma patients. They dosed 6 melanoma patients with up to 20 neoantigens to test for safety, feasibility, and immune response. Across the 6 patients, they observed ~20% CD8 and >60% CD4 T cell responses against the neoantigens [47], all of which were new responses following vaccination. In another trial, Dr. Wu and colleagues tested Neovax in patients with

glioblastoma multiforme, a tumor with a lower somatic mutation rate [48]. Again, they saw circulating neoantigen-specific responses. Her presentation ended with a series of questions that need to be addressed to implement better personalized immunotherapies, including the identification of methods to reduce time from diagnosis to vaccine administration [49].

In the final presentation of the day, Dr. Marios Giannakis presented work on investigating the genomic mechanisms of immune evasion in colorectal cancer. Immune checkpoint blockade has shown activity in approximately 50% of MSI-high colorectal cancers while it is ineffective in microsatellite stable tumors [50, 51]. To better understand the genetic drivers of immune evasion in colorectal cancer, Dr. Giannakis and colleagues integrated next generation sequencing data from over 1,200 tumors with transcriptional and immunohistochemical measures of immune infiltration. The tumor samples for their studies were from TCGA, the NHS, and the HPFS [52]. They demonstrated that WNT-signaling and immune-related genes were significantly mutated in colorectal cancer. They also found frequent inactivating antigen-presentation machinery mutations in MSI-high tumors, and an inverse association between WNT-signaling activity and T cell infiltration in all subtypes of colorectal cancer. Tumors with biallelic disruptive mutations in *APC* or with *AXIN2* super-enhancer hypomethylation had a significantly decreased T cell transcriptional signature [53]. In summary, this work found evidence of immuno-editing through disruptive mutations in antigen-presentation machinery, and of exclusion of an effective immune response through an active WNT-signaling pathway in colorectal cancer. These results shed light to the underlying molecular mechanisms of immune evasion in this disease.

1 June 2018

The second day of the meeting began with award announcements to trainees and early career investigators.

Session 6: Novel disease phenotyping in future medicine and population science #1 (Session Chair: Dr. Jonathan Nowak)

The first lecture of the day was given by Dr. Jeffrey Golden who first shared a few clinical case reports whereby modern tools in molecular pathology aided in more precise diagnoses and superior treatments for patients and almost certainly improved their prognoses. Dr. Golden summarized the importance of computational pathology [54, 55] into five central components: [1] mutation-specific treatment stratification, including clinical trial eligibility; [2] better precision diagnostics; [3] superior prognostication; [4] better targeted

therapies, which may be more effective, less toxic, and more cost-effective; and [5] new biomarker discovery. One of the central challenges for a practicing clinician is in translating the abundance of computational data generated by *omic*s platforms (e.g., Oncopanel) into tractable information and, ultimately, into knowledge of what those mutations mean for the benefit of the patient. Dr. Golden's institution has converted these sorts of information into knowledge—that is, what mutations in those genes mean to the patient and clinician [56]. He used Google maps' system of layered geospatial databases as an analogy for creating relational databases for patients, based on a multitude of patient (e.g., clinical, omic, pharmacy) and external (e.g., electronic medical records, published literature) inputs, that are ultimately led back to actionable decisions for clinicians and their patients.

Dr. Lynette Sholl's lecture focused on recent advancements in liquid biopsy techniques for solid tumors. She started with a clinical case report that highlighted some of the advantages of liquid biopsies over tumor/solid tissues, including: tumor tissues often have limited mass and/or normal cell contamination; there is often a need to repeat a solid tissue biopsy upon relapse; and some tumors are anatomically inaccessible. Liquid biopsies can look for circulating tumor DNA (ctDNA) although their rarity in peripheral blood makes the approach challenging. Current and emerging technologies include commercial allele specific PCR platforms, droplet digital PCR (ddPCR), NGS, and electric field-based measurements [57]. Investigators at her institute decided to invest in ddPCR for ctDNA testing in cancer patients. Using ddPCR for EGFR and KRAS hotspots, they found sensitivity was a direct correlate of the number of metastatic sites, owing to the amount of DNA shed into circulation. Patients with only 1 metastatic site saw sensitivity levels of approximately 60% whereas patients with four or more metastatic sites had nearly 100% sensitivity with ddPCR [58]. Ultimately, liquid biopsy may be used in detection of early relapse after definitive therapy, minimal residual disease testing, or even cancer screening.

Dr. Yujing (Jan) Heng's talk highlighted the importance of gene expression pre-processing methods to obtain reliable and reproducible breast cancer molecular subtype classification by PAM50. Her work compared two established pre-processing methods (i.e., modified median gene centering [59] and subgroup-specific gene centering [60]) to compute molecular subtypes using PAM50 in tumor and tumor-adjacent tissues from participants in the NHS. She reported that although molecular subtypes were highly comparable using either method, the subgroup-specific method tends to classify more cases into more aggressive subtypes. The distribution of molecular subtypes within the NHS/NHS-II was comparable to other population-based studies, and as expected, there were more cancer recurrences in women with Basal-like subtype compared to Luminal A subtype.

Lastly, she showed that the correlation of molecular subtypes classified using PAM50 and immunohistochemical (IHC) surrogates remain poor and more research is needed to refine the IHC definitions to more closely approximate PAM50 subtypes [61].

Session 7: Novel disease phenotyping in future medicine and population science #2 (Session Chair: Dr. Reiko Nishihara)

After a short break, the second session of the day continued with the ‘novel disease phenotyping’ theme. The first lecture of the session was given by Dr. John Quackenbush who described his group’s efforts to infer biological networks from GWAS and gene expression data. To do this, they developed several systems biology algorithms [62, 63]. Dr. Quackenbush’s lecture described how these algorithms were used to better understand networks of GWAS-identified SNPs and gene expression data with outcomes, and how patterns of gene networks differ between phenotypes. Central to this work are the hypotheses that biological systems are driven by complex networks; the structure of the network captures the biology of the system; and, that network structure is conditional, depending on tissue, biological state, and individual. One of the main findings from their work is that biological networks are organized into tight communities such that, in most instances, it is not a single gene controlling a single trait, but a family of genetic variants that influence a process. They also find disease-associated (i.e., GWAS) SNPs map to communities whose genes share functions that are related to the disease, and that most GWAS SNPs are not global hubs in the network, but local hubs in the network. Overall, their work demonstrates how network analysis can take us beyond simple differential expression in understanding disease [64, 65].

Dr. Hugo Aerts described his work at the intersection of radiology, bioinformatics, and data science. He discussed recent work of building Artificial Intelligence (AI) image analysis systems to extract a rich radiomics set and used these features to build biomarkers [66]. He illustrated how technological advances in AI and deep learning are moving imaging modalities into the heart of patient care as imaging can address a critical barrier in precision medicine because solid tumors can be spatially and temporally heterogeneous, and the standard approach to tumor sampling, often invasive needle biopsy, is unable to fully capture the spatial state of the tumor [67, 68]. The main objectives of the talk were to learn about the motivation and methodology of AI technologies in radiology, to learn about the existing and future potential role of radiologic AI with other omics data for precision medicine, and to learn about open-source informatics developments [69].

Session 8: Discussion topic special lectures (Session Chair: Dr. John Quackenbush)

In the final session for podium presentations, the organizing committee selected five submitted abstracts whose topics were thought to provide the broadest interest and potential to generate discussion for the attendees.

Dr. N Sertac Kip lectured on her work in developing a liquid biopsy for tumor detection. Dr. Kip spoke of the limitations to standard tumor profiling (e.g., biopsy/resection), including invasiveness, pain, cost, and lack of sensitivity to mutations unique to the non-biopsied tissue [70]. Further, targeted therapies place tumor cells under selective pressure, thereby triggering clonal progression, which can then be captured [71]. Liquid biopsy is ideal for comparing pre- and post-treatment variants and to optimize sequence of therapy [72]. Lung cancer appears to be an ideal tumor site for this application. Liquid biopsy measures biomarkers that are predictive and prognostic for lung cancer [73]. Coming back to the title of her talk, “How solid is liquid biopsy?”, only two of 34 cases had discordance between solid tumor and liquid biopsy measures, indicating that the liquid biopsy is quite solid [74]. There are some people with mutations in tissue that are not detected in liquid biopsy. Sensitivity also varies by tumor type and stage [75]. Research is rapidly evolving in this area, but the techniques are not yet suitable for wide clinical application and guidelines are currently lacking for standardization of the liquid biopsy results.

Dr. Camila Lopes-Ramos spoke about her work in identifying biological explanations for the higher incidence and mortality rates for men than women from colon cancer. Lifestyle and serological (i.e., sex steroid hormone) differences are often postulated to explain some, but not all, of these observed sex differences. The potential molecular features that drive sex differences are understudied. Her work used both transcript-based and gene regulatory network methods to analyze RNA-seq data from TCGA for colon cancer. They found no meaningful differences between tumors from men and women for gene expression. Next, they examined patient-specific gene regulatory networks and found considerable sex differences in drug and xenobiotic metabolism via cytochrome P450 pathways which were considerably more pronounced in women. This finding was replicated in several independent study samples. This drug metabolism pathway was not associated with survival in men; however, women treated with chemotherapy that had increased targeting (compared to less targeting) of this pathway had considerably better 10-year overall survival [76]. This network-based approach can be applied to explore other etiologic and demographic differences for cancer and other complex diseases.

Dr. Molin Wang lectured on the problem of sample selection bias due to tissue availability and solutions to this problem. Missing data is a common problem for tumor subtype

data. This poses severe statistical challenges for MPE research because the outcome is missing. She reported the percentages of colorectal, breast and ovarian cancer cases with missing tumor marker data in representative national and international cohort studies and consortia. She introduced the definitions and assumptions of possible missingness patterns of the tumor marker data in the MPE context and used the NHS breast and colorectal cancer data to illustrate the problem of sample selection bias in MPE research. She then described the statistical methods available to explore these issues, such as the complete case analysis method, the missing indicator method, inverse probability weighting [77], and the multiple imputation method. They found that the former two methods could lead to biased estimates and the latter two methods were usually most helpful in dealing with the potential selection bias problem. She used a colorectal cancer molecular subtype to illustrate these methods. She described the possible scenarios when the missingness is not at random, and the statistical challenges in these scenarios.

Dr. Song Yao presented recent data from his group on population differences between women of African American descent and European American descent in the breast tumor immune microenvironment. Convincing evidence demonstrates marked differences in systemic immune response between African American and European American populations [78–80], which is also supported in a recent study on circulating cytokine levels [81]. However, data are scarce on population differences in tumor immune microenvironment. With data from the Women's Circle of Health Study (WCHS), they showed that breast tumors from African American women had a significantly stronger presence of tumor-infiltrating lymphocytes than breast tumors from European American women, independent of tumor histopathological features. Using NanoString immune profiling, they confirmed the overall stronger immune infiltration in breast tumors from African American women than those from European American women, and further showed stronger exhausted T cell signatures in tumors from African American women. Their data revealed marked population differences in tumor immune response, which may contribute to some of the observed racial disparities in breast cancer survival.

Dr. Kun-Hsing Yu lectured about his work on integrating lung cancer multi-omics and histopathology images [82]. Lung cancer is the most prevalent cancer worldwide, and histopathological assessment is indispensable for its diagnosis [83]. However, how histopathology findings relate to molecular abnormalities remains largely unknown, and human evaluation of pathology slides do not accurately predict prognosis. To address this gap, his group obtained over 2,100 hematoxylin and eosin stained histopathology whole-slide images, RNA sequencing, and proteomics data of lung

adenocarcinoma and squamous cell carcinoma patients from TCGA, and nearly 300 additional images from the Stanford Tissue Microarray (TMA) Database [84]. They extracted nearly 10,000 quantitative image features and used regularized machine-learning methods to select the top features and to distinguish shorter-term survivors from longer-term survivors with stage I adenocarcinoma or squamous cell carcinoma in the TCGA data set [85]. They successfully validated the survival prediction framework with the TMA cohort [85], identified the cell-cycle regulation and nucleotide binding pathways underpinning tumor cell dedifferentiation [86], and built an integrative histopathology-transcriptomics model to generate better prognostic predictions for stage I adenocarcinoma patients compared with gene expression or histopathology studies alone [86]. These results suggest that automatically derived image features can predict the prognosis of lung cancer patients and therefore contribute to precision oncology.

Conclusions

The Fourth International MPE Meeting assembled over 170 trainees and experts working in the various, diverse scientific disciplines that comprise MPE. As even more sophisticated means of molecular characterization of disease processes enter epidemiologic studies and clinical medicine, the utility and preponderance of MPE principles and methods should continue to expand. As actively discussed in the Fourth International MPE Meeting, new ideas of flexibly shaping and integrating multiple disciplines are further expanding opportunities in biomedical and population sciences [87]. In terms of key recommendations and next steps, integration of tumor immunology into epidemiologic studies and further exploration of disparity research were concluded as high priorities for the field. Additionally, further collaborations between computational and population-level scientists were noted as high priority as were general pooling projects between studies with similar data. We look forward to meeting again at the Fifth International MPE Meeting, tentatively planned for June 2020 in Boston, MA, USA.

Funding Grant support: NIH R35 CA197735 (to S.O.).P30 CA006516 (Dana-Farber Harvard Cancer Center Support Grant; to L.H. Glimcher).1R35 CA220523 (to J.Q.) 1R01 HL111759 (to J.Q.)

References

1. Campbell PT, Curtin K, Ulrich C et al (2008) Mismatch repair polymorphisms and risk of colon cancer, tumor microsatellite instability, and interactions with lifestyle factors. *Gut* 58:661–667

2. Campbell PT, Jacobs ET, Ulrich CM et al (2010) Case-control study of overweight, obesity, and colorectal cancer risk, overall and by tumor microsatellite instability status. *J Natl Cancer Inst* 102:391–400
3. Campbell PT, Newton CC, Newcomb PA et al (2015) Association between body mass index and mortality for colorectal cancer survivors: overall and by tumor molecular phenotype. *Cancer Epidemiol Biomarkers Prev* 24:1229–1238
4. Yang XR, Chang-Claude J, Goode EL et al (2011) Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. *J Natl Cancer Inst* 103:250–263
5. Schmidt ME, Steindorf K, Mutschelknauss E et al (2008) Physical activity and postmenopausal breast cancer: effect modification by breast cancer subtypes and effective periods in life. *Cancer Epidemiol Biomarkers Prev* 17:3402–3410
6. Suzuki R, Ye W, Rylander-Rudqvist T, Saji S, Colditz GA, Wolk A (2005) Alcohol and postmenopausal breast cancer risk defined by estrogen and progesterone receptor status: a prospective cohort study. *J Natl Cancer Inst* 97:1601–1608
7. Ogino S, Campbell PT, Nishihara R et al (2015) Proceedings of the second international molecular pathological epidemiology (MPE) meeting. *Cancer Causes Control* 26:959–972
8. Campbell PT, Rebbeck TR, Nishihara R et al (2017) Proceedings of the third international molecular pathological epidemiology (MPE) meeting. *Cancer Causes Control* 28:167–176
9. Penney KL, Pettersson A, Shui IM et al (2016) Association of prostate cancer risk variants with TMPRSS2:ERG status: evidence for distinct molecular subtypes. *Cancer Epidemiol Biomarkers Prev* 25:745–749
10. Yoo S, Pettersson A, Jordahl KM et al (2014) Androgen receptor CAG repeat polymorphism and risk of TMPRSS2:ERG-positive prostate cancer. *Cancer Epidemiol Biomarkers Prev* 23:2027–2031
11. Pettersson A, Lis RT, Meisner A et al (2013) Modification of the association between obesity and lethal prostate cancer by TMPRSS2:ERG. *J Natl Cancer Inst* 105:1881–1890
12. Michailidou K, Lindstrom S, Dennis J et al (2017) Association analysis identifies 65 new breast cancer risk loci. *Nature* 551:92–94
13. Milne RL, Kuchenbaecker KB, Michailidou K et al (2017) Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat Genet* 49:1767–1778
14. Jacobs DI, Fukumura K, Bainbridge MN et al (2018) Elucidating the molecular pathogenesis of glioma: integrated germline and somatic profiling of a familial glioma case series. *Neuro Oncol* 20:1625–1633
15. Garcia-Closas M, Gunsoy NB, Chatterjee N. (2014) Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. *J Natl Cancer Inst* 106:fju305
16. Maas P, Barrdahl M, Joshi AD et al (2016) Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. *JAMA Oncol* 2:1295–1302
17. Mavaddat N, Pharoah PD, Michailidou K, et al. (2015) Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst* 107:djv036
18. Berkey CS, Rosner B, Tamimi RM et al (2017) Body size from birth through adolescence in relation to risk of benign breast disease in young women. *Breast Cancer Res Treat* 162:139–149
19. Baer HJ, Schnitt SJ, Connolly JL et al (2005) Early life factors and incidence of proliferative benign breast disease. *Cancer Epidemiol Biomarkers Prev* 14:2889–2897
20. Warner ET, Hu R, Collins LC et al (2016) Height and body size in childhood, adolescence, and young adulthood and breast cancer risk according to molecular subtype in the Nurses' Health Studies. *Cancer Prev Res (Phila)* 9:732–738
21. Oh H, Eliassen AH, Beck AH et al (2017) Breast cancer risk factors in relation to estrogen receptor, progesterone receptor, insulin-like growth factor-1 receptor, and Ki67 expression in normal breast tissue. *NPJ Breast Cancer* 3:39
22. Palmer JR, Ambrosone CB, Olshan AF (2014) A collaborative study of the etiology of breast cancer subtypes in African American women: the AMBER consortium. *Cancer Causes Control* 25:309–319
23. Espinal AC, Buas MF, Wang D et al (2017) FOXA1 hypermethylation: link between parity and ER-negative breast cancer in African American women? *Breast Cancer Res Treat* 166:559–568
24. Palmer JR, Viscidi E, Troester MA, et al (2014) Parity, lactation, and breast cancer subtypes in African American women: results from the AMBER Consortium. *J Natl Cancer Inst* 106:dju237
25. Gilsing AM, Fransen F, de Kok TM et al (2013) Dietary heme iron and the risk of colorectal cancer with specific mutations in KRAS and APC. *Carcinogenesis* 34:2757–2766
26. Petljak M, Alexandrov LB (2016) Understanding mutagenesis through delineation of mutational signatures in human cancer. *Carcinogenesis* 37:531–540
27. Hughes LA, van den Brandt PA, de Bruine AP et al (2009) Early life exposure to famine and colorectal cancer risk: a role for epigenetic mechanisms. *PLoS ONE* 4:e7951
28. Simons CC, van den Brandt PA, Stehouwer CD, van Engeland M, Weijenberg MP (2014) Body size, physical activity, early-life energy restriction, and associations with methylated insulin-like growth factor-binding protein genes in colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 23:1852–1862
29. van den Brandt PA, Schulpen M (2017) Mediterranean diet adherence and risk of postmenopausal breast cancer: results of a cohort study and meta-analysis. *Int J Cancer* 140:2220–2231
30. Rebbeck TR, Sankar P (2005) Ethnicity, ancestry, and race in molecular epidemiologic research. *Cancer Epidemiol Biomarkers Prev* 14:2467–2471
31. Lachance J, Berens AJ, Hansen MEB, Teng AK, Tishkoff SA, Rebbeck TR (2018) Genetic hitchhiking and population bottlenecks contribute to prostate cancer disparities in men of African descent. *Cancer Res* 78:2432–2443
32. Ogino S, Nowak JA, Hamada T et al (2018) Integrative analysis of exogenous, endogenous, tumour and immune factors for precision medicine. *Gut* 67:1168–1180
33. Ogino S, Giannakis M (2018) Immunoscore for (colorectal) cancer precision medicine. *Lancet* 391:2084–2086
34. Song M, Nishihara R, Cao Y et al (2016) Marine omega-3 polyunsaturated fatty acid intake and risk of colorectal cancer characterized by tumor-infiltrating T cells. *JAMA Oncol* 2:1197–1206
35. Cao Y, Nishihara R, Qian ZR et al (2016) Regular aspirin use associates with lower risk of colorectal cancers with low numbers of tumor-infiltrating lymphocytes. *Gastroenterology* 151:879–892
36. Liu L, Nishihara R, Qian ZR et al (2017) Association between inflammatory diet pattern and risk of colorectal carcinoma subtypes classified by immune responses to tumor. *Gastroenterology* 153(1517–30):e14
37. Khalili H, Gong J, Brenner H et al (2015) Identification of a common variant with potential pleiotropic effect on risk of inflammatory bowel disease and colorectal cancer. *Carcinogenesis* 36:999–1007
38. Mehta RS, Nishihara R, Cao Y et al (2017) Association of dietary patterns with risk of colorectal cancer subtypes classified by *Fusobacterium nucleatum* in tumor tissue. *JAMA Oncol* 3:921–927
39. Li B, Severson E, Pignon JC et al (2016) Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol* 17:174

40. Li T, Fan J, Wang B et al (2017) TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res* 77:e108–e110
41. Li B, Li T, Wang B et al (2017) Ultrasensitive detection of TCR hypervariable-region sequences in solid-tissue RNA-seq data. *Nat Genet* 49:482–483
42. Li B, Li T, Pignon JC et al (2016) Landscape of tumor-infiltrating T cell repertoire of human cancers. *Nat Genet* 48:725–732
43. Hu X, Zhang J, Wang J et al (2019) Landscape of B cell immunity and related immune evasion in human cancers. *Nat Genet* 51:560–567
44. Jiang P, Gu S, Pan D et al (2018) Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* 24:1550–1558
45. Kimura T, McKolanis JR, Dzubinski LA et al (2013) MUC1 vaccine for individuals with advanced adenoma of the colon: a cancer immunoprevention feasibility study. *Cancer Prev Res (Phila)* 6:18–26
46. Abelin JG, Keskin DB, Sarkizova S et al (2017) Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity* 46:315–326
47. Ott PA, Hu Z, Keskin DB et al (2017) An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* 547:217–221
48. Keskin DB, Anandappa AJ, Sun J et al (2019) Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature* 565:234–239
49. Hu Z, Ott PA, Wu CJ (2018) Towards personalized, tumour-specific, therapeutic vaccines for cancer. *Nat Rev Immunol* 18:168–182
50. Le DT, Uram JN, Wang H et al (2015) PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med* 372:2509–2520
51. Le DT, Durham JN, Smith KN et al (2017) Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* 357:409–413
52. Giannakis M, Mu XJ, Shukla SA et al (2016) Genomic correlates of immune-cell infiltrates in colorectal carcinoma. *Cell Rep* 15:857–865
53. Grasso CS, Giannakis M, Wells DK et al (2018) Genetic mechanisms of immune evasion in colorectal cancer. *Cancer Discov* 8:730–749
54. Louis DN, Feldman M, Carter AB et al (2016) Computational pathology: a path ahead. *Arch Pathol Lab Med* 140:41–50
55. Louis DN, Gerber GK, Baron JM et al (2014) Computational pathology: an emerging definition. *Arch Pathol Lab Med* 138:1133–1138
56. Lennerz JK, McLaughlin HM, Baron JM et al (2016) Health care infrastructure for financially sustainable clinical genomics. *J Mol Diagn* 18:697–706
57. Sacher AG, Paweletz C, Dahlberg SE et al (2016) Prospective validation of rapid plasma genotyping for the detection of EGFR and KRAS mutations in advanced lung cancer. *JAMA Oncol* 2:1014–1022
58. Adalsteinsson VA, Ha G, Freeman SS et al (2017) Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* 8:1324
59. Heng YJ, Lester SC, Tse GM et al (2017) The molecular basis of breast cancer pathological phenotypes. *J Pathol* 241:375–391
60. Zhao X, Rodland EA, Tibshirani R, Plevritis S (2015) Molecular subtyping for clinically defined breast cancer subgroups. *Breast Cancer Res* 17:29
61. Kensler KH, Sankar VN, Wang J et al (2019) PAM50 molecular intrinsic subtypes in the Nurses' Health Study Cohorts. *Cancer Epidemiol Biomarkers Prev* 28:798–806
62. Glass K, Huttenhower C, Quackenbush J, Yuan GC (2013) Passing messages between biological networks to refine predicted interactions. *PLoS ONE* 8:e64832
63. Platig J, Castaldi PJ, DeMeo D, Quackenbush J (2016) Bipartite community structure of eQTLs. *PLoS Comput Biol* 12:e1005033
64. Sonawane AR, Platig J, Fagny M et al (2017) Understanding tissue-specific gene regulation. *Cell Rep* 21:1077–1088
65. Fagny M, Paulson JN, Kuijjer ML et al (2017) Exploring regulation in tissues with eQTL networks. *Proc Natl Acad Sci USA* 114:E7841–E7850
66. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H (2018) Artificial intelligence in radiology. *Nat Rev Cancer* 18:500–510
67. Aerts H (2018) Data science in radiology: a path forward. *Clin Cancer Res* 24:532–534
68. Grossmann P, Stringfield O, El-Hachem N, et al. (2017) Defining the biological basis of radiomic phenotypes in lung cancer. *Elife* 6:e30563
69. Aerts HJ (2016) The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA Oncol* 2:1636–1642
70. Kumar M, Choudhury Y, Ghosh SK, Mondal R (2018) Application and optimization of minimally invasive cell-free DNA techniques in oncogenomics. *Tumour Biol* 40:1010428318760342
71. Liang W, Zhao Y, Huang W, Liang H, Zeng H, He J (2018) Liquid biopsy for early stage lung cancer. *J Thorac Dis* 10:S876–S881
72. Tie J, Gibbs P (2016) Sequencing circulating cell-free DNA: the potential to refine precision cancer medicine. *Clin Chem* 62:796–798
73. Liu C, Yang Y, Wu Y (2018) Recent advances in exosomal protein detection via liquid biopsy biosensors for cancer screening, diagnosis, and prognosis. *AAPS J* 20:41
74. Jung A, Kirchner T (2018) Liquid biopsy in tumor genetic diagnosis. *Dtsch Arztebl Int* 115:169–174
75. Chae YK, Davis AA, Jain S et al (2017) Concordance of genomic alterations by next-generation sequencing in tumor tissue versus circulating tumor DNA in breast cancer. *Mol Cancer Ther* 16:1412–1420
76. Lopes-Ramos CM, Kuijjer ML, Ogino S et al (2018) Gene regulatory network analysis identifies sex-linked differences in colon cancer drug metabolism. *Cancer Res* 78:5538–5547
77. Liu L, Nevo D, Nishihara R et al (2018) Utility of inverse probability weighting in molecular pathological epidemiology. *Eur J Epidemiol* 33:381–392
78. Ye CJ, Feng T, Kwon HK et al (2014) Intersection of population variation and autoimmunity genetics in human T cell activation. *Science* 345:1254665
79. Nedelec Y, Sanz J, Baharian G et al (2016) Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell* 167(657–69):e21
80. Quach H, Rotival M, Pothlichet J et al (2016) Genetic adaptation and neandertal admixture shaped the immune system of human populations. *Cell* 167(643–56):e17
81. Yao S, Hong CC, Ruiz-Narvaez EA et al (2018) Genetic ancestry and population differences in levels of inflammatory cytokines in women: Role for evolutionary selection and environmental factors. *PLoS Genet* 14:e1007368
82. Yu KH WF, Berry GJ, Re C, Altman RB, Snyder M, Kohane IS. (2019) Classifying non-small cell lung cancer histopathology types and transcriptomic subtypes using convolutional neural networks. *bioRxiv* 1:530360
83. Travis WD, Brambilla E, Noguchi M et al (2011) International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol* 6:244–285
84. Marinelli RJ, Montgomery K, Liu CL et al (2008) The stanford tissue microarray database. *Nucleic Acids Res* 36:D871–D877

85. Yu KH, Zhang C, Berry GJ et al (2016) Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun*. 7:12474
86. Yu KH, Berry GJ, Rubin DL, Re C, Altman RB, Snyder M (2017) Association of omics features with histopathology patterns in lung adenocarcinoma. *Cell Syst*. 5(620–7):e3
87. Ogino S, Nowak JA, Hamada T, Milner DA Jr, Nishihara R (2019) Insights into pathogenic interactions among environment, host, and tumor at the crossroads of molecular pathology and epidemiology. *Annu Rev Pathol* 14:83–103

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Peter T. Campbell¹ · Christine B. Ambrosone² · Reiko Nishihara^{3,4,5} · Hugo J. W. L. Aerts⁶ · Melissa Bondy⁷ · Nilanjan Chatterjee⁸ · Montserrat Garcia-Closas⁹ · Marios Giannakis^{10,11} · Jeffrey A. Golden¹² · Yujing J. Heng¹³ · N. Sertac Kip¹⁴ · Jill Koshiol⁹ · X. Shirley Liu¹⁵ · Camila M. Lopes-Ramos⁵ · Lorelei A. Mucci¹⁶ · Jonathan A. Nowak¹² · Amanda I. Phipps¹⁷ · John Quackenbush⁵ · Robert E. Schoen¹⁸ · Lynette M. Sholl¹² · Rulla M. Tamimi^{16,19} · Molin Wang^{5,16,19} · Matty P. Weijenberg²⁰ · Catherine J. Wu^{10,11} · Kana Wu⁴ · Song Yao² · Kun-Hsing Yu^{12,21} · Xuehong Zhang¹⁹ · Timothy R. Rebbeck^{10,16} · Shuji Ogino^{3,11,16,22}

✉ Peter T. Campbell
peter.campbell@cancer.org

✉ Shuji Ogino
Shuji_ogino@dfci.harvard.edu

Christine B. Ambrosone
Christine.Ambrosone@RoswellPark.org

Reiko Nishihara
reiko.nishihara@mail.harvard.edu

Hugo J. W. L. Aerts
Hugo_Aerts@DFCI.HARVARD.EDU

Melissa Bondy
mbondy@bcm.edu

Nilanjan Chatterjee
nchatte2@jhu.edu

Montserrat Garcia-Closas
montserrat.garcia-closas@nih.gov

Marios Giannakis
Marios_Giannakis@dfci.harvard.edu

Jeffrey A. Golden
jagolden@bwh.harvard.edu

Yujing J. Heng
yheng@bidmc.harvard.edu

N. Sertac Kip
sertac.kip@sema4.com

Jill Koshiol
koshiolj@mail.nih.gov

X. Shirley Liu
xslu@jimmy.harvard.edu

Camila M. Lopes-Ramos
camilaramos@hsph.harvard.edu

Lorelei A. Mucci
lmucci@hsph.harvard.edu

Jonathan A. Nowak
janowak@bwh.harvard.edu

Amanda I. Phipps
aiphipps@uw.edu

John Quackenbush
johnq@hsph.harvard.edu

Robert E. Schoen
rschoen@pitt.edu

Lynette M. Sholl
lmsholl@bwh.harvard.edu

Rulla M. Tamimi
nrhmt@channing.harvard.edu

Molin Wang
stmow@channing.harvard.edu

Matty P. Weijenberg
mp.weijenberg@maastrichtuniversity.nl

Catherine J. Wu
cwu@partners.org

Kana Wu
hpkwu@channing.harvard.edu

Song Yao
Song.Yao@RoswellPark.org

Kun-Hsing Yu
Kun-Hsing_Yu@hms.harvard.edu

Xuehong Zhang
xuehong.zhang@channing.harvard.edu

Timothy R. Rebbeck
Timothy_Rebbeck@DFCI.HARVARD.EDU

¹ Behavioral and Epidemiology Research Group, American Cancer Society, 250 Williams Street NW, Atlanta, GA 30303, USA

² Department of Cancer Prevention and Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA

³ Program in MPE Molecular Pathological Epidemiology, Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, 450 Brookline Ave, Room SM1036, Boston, MA 02215, USA

- 4 Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA
- 5 Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
- 6 Departments of Radiation Oncology and Radiology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
- 7 Cancer Prevention and Population Sciences, Dan L. Duncan Comprehensive Cancer Center, Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, TX, USA
- 8 Department of Biostatistics, Bloomberg School of Public Health, Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, MD, USA
- 9 Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
- 10 Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA
- 11 Broad Institute of Harvard & MIT, Cambridge, MA, USA
- 12 Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
- 13 Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA
- 14 Sema4, Mount Sinai Icahn School of Medicine, Genetics & Genomic Sciences and Pathology, Branford, CT, USA
- 15 Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard T.H. Chan School of Public Health, Boston, MA 02215, USA
- 16 Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
- 17 Department of Epidemiology, University of Washington, Seattle, WA, USA
- 18 Departments of Medicine and Epidemiology, The University of Pittsburgh, Pittsburgh, PA, USA
- 19 Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
- 20 Department of Epidemiology, GROW-School for Oncology and Developmental Biology, Maastricht University, Maastricht, The Netherlands
- 21 Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
- 22 Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA, USA