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

Campbell, P. T., Rebbeck, T. R., Nishihara, R., Beck, A. H., Begg, C. B., Bogdanov, A. A., Cao, Y., Coleman, H. G., Freeman, G. J., Heng, Y. J., Huttenhower, C., Irizarry, R. A., Kip, N. S., Michor, F., Nevo, D., Peters, U., Phipps, A. I., Poole, E. M., Qian, Z. R., ... Ogino, S. (2017). Proceedings of the third international molecular pathological epidemiology (MPE) meeting. *Cancer Causes & Control: an international journal of studies of cancer in human populations*. https://doi.org/10.1007/s10552-016-0845-z

Published in:
Cancer Causes & Control: an international journal of studies of cancer in human populations

Document Version:
Peer reviewed version

Proceedings of the Third International Molecular Pathological Epidemiology (MPE) Meeting.

Peter T. Campbell*[1], Timothy R. Rebbeck[2,3], Reiko Nishihara[3,4], Andrew H. Beck[5,6], Colin B. Begg[7], Alexei A. Bogdanov[8], Yin Cao[4,9,10], Helen G. Coleman[11], Gordon J. Freeman[3], Yujing J. Heng[5,6], Curtis Huttenhower[12,13], Rafael A. Irizarry[12,14], N. Sertac Kip[15], Franziska Michor[12,14], Daniel Nevo[2,12], Ulrike Peters[16,17], Amanda I. Phipps[16,17], Elizabeth M. Poole[2,18], Zhi Rong Qian[3], John Quackenbush[12,14], Harlan Robins[16], Peter K. Rogan[19], Martha L. Slattery[20], Stephanie A. Smith-Warner[2,4], Mingyang Song[4,9], Tyler J. VanderWeele[2], Daniel Xia[21], Emily C. Zabor[7], Xuehong Zhang[18], Molin Wang[2], and Shuji Ogino*[2,3,22,23]

*These authors contributed equally to this work.

Author affiliations:

[1]Epidemiology Research Program, American Cancer Society, Atlanta, GA.

[2]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA.

[3]Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA.

[4]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA.

[5]Cancer Research Institute, Beth Israel Deaconess Cancer Center, Boston, MA.

[6]Department of Pathology, Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, MA.

[7]Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, New York, N.Y.

[8]Department of Radiology, University of Massachusetts Medical School, Worcester, MA.

[9]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA.

[10]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA.

[11]Epidemiology & Health Services Research Group, Centre for Public Health, Queens University Belfast, Northern Ireland

[12]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA.

[13]Microbial Systems and Communities, Genome Sequencing and Analysis Program, The Broad Institute, Cambridge, MA.

[14]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA.

[15]Laboratory Medicine and Pathology, Geisinger Health System, Danville, PA.

[16]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA.

[17]Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA.

[18]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

[19]Department of Biochemistry, University of Western Ontario, London, Canada.

[20]University of Utah School of Medicine, Salt Lake City, UT.

[21]Department of Pathology, Brigham and Women's Hospital, Boston, MA.

[22]Division of MPE Molecular Pathological Epidemiology, Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

[23]Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA.

Correspondence:

Peter T Campbell, PhD
Epidemiology Research Program
American Cancer Society
250 Williams Street NW
Atlanta, GA, 30303
Email: peter.campbell@cancer.org
Phone: 404-327-6460
Fax: 404-327-6450

Shuji Ogino, MD, PhD, MS
Division of MPE Molecular Pathological Epidemiology
Brigham and Women's Hospital
450 Brookline Ave, Room SM1036
Boston, MA 02215 USA
617-632-1972; fax 617-582-8558
Shuji_ogino@dfci.harvard.edu

Co-author email addresses:

Martha Slattery - marty.slattery@hsc.utah.edu

Stephanie Smith-Warner - swarner@hsph.harvard.edu

Ulrike Peters - upeters@fredhutch.org

Amanda Phipps - aphipps@fredhutch.org

Harlan Robins - hrobins@fredhutch.org

Gordon Freeman - gordon_freeman@dfci.harvard.edu

Xuehong Zhang - xuehong.zhang@channing.harvard.edu

Timothy R. Rebbeck - Timothy_Rebbeck@DFCI.HARVARD.EDU

Franziska Michor - michor@jimmy.harvard.edu

Alexei Bogdanov - Alexei.Bogdanov@umassmed.edu

Molin Wang - stmow@channing.harvard.edu

Colin Begg - beggc@mskcc.org

Tyler VanderWeele - tvanderw@hsph.harvard.edu

Reiko Nishihara - reiko.nishihara@mail.harvard.edu; rnishiha@hsph.harvard.edu

John Quackenbush - johnq@jimmy.harvard.edu

Rafael Irizarry - rafa@jimmy.harvard.edu

Curtis Huttenhower - chuttenh@hsph.harvard.edu

Elizabeth Poole - nhlip@channing.harvard.edu

Emily C. Zabor - zabore@mskcc.org

Daniel Nevo - nhdne@channing.harvard.edu

Yujing J. Heng – yheng@bidmc.harvard.edu

N Sertac Kip - nskip@geisinger.edu

Daniel Xia – YXIA8@PARTNERS.ORG

Peter Rogan - progan@uwo.ca

Helen G. Coleman - h.coleman@qub.ac.uk

Yin Cao - yic576@mail.harvard.edu

Mingyang Song - msong2@partners.org

Andrew Beck - abeck2@bidmc.harvard.edu

Zhirong Qian – Zhirong_qian@dfci.harvard.edu

Grant support:
NIH
R13 CA203287 (to S.O.: supported by National Cancer Institute, National Human Genome Research Institute, and National Institute of Environmental Health Sciences)
R35 CA197735 (to S.O.)
P30 CA006516 (Dana-Farber Harvard Cancer Center Support Grant; to E.J. Benz;).
Nodal Award (to S.O.) from Dana-Farber Harvard Cancer Center

Abstract

Molecular pathological epidemiology (MPE) is a transdisciplinary and relatively new scientific discipline that integrates theory, methods and resources from epidemiology, pathology, biostatistics, bioinformatics and computational biology. The underlying objective of MPE research is to better understand the etiology and progression of complex and heterogeneous human diseases with the goal of informing prevention and treatment efforts in population health and clinical medicine. Although MPE research has been commonly applied to investigating breast, lung, and colorectal cancers, its methodology can be used to study most diseases. Recent successes in MPE studies include: 1) the development of new statistical methods to address etiologic heterogeneity; 2) the enhancement of causal inference; 3) the identification of previously unknown exposure-subtype disease associations; and 4) better understanding of the role of lifestyle/behavioral factors on modifying prognosis according to disease subtype. Central challenges to MPE include the relative lack of transdisciplinary experts, educational programs, and forums to discuss issues related to the advancement of the field. To address these challenges, highlight recent successes in the field, and identify new opportunities, a series of MPE meetings have been held at the Dana-Farber Cancer Institute in Boston, MA. Herein, we share the proceedings of the Third International MPE Meeting, held in May 2016 and attended by 150 scientists from 17 countries. Special topics included integration of MPE with immunology and health disparity research. This meeting series will continue to provide an impetus to foster further transdisciplinary integration of divergent scientific fields.

Introduction

Molecular pathological epidemiology (MPE) is an integrative scientific discipline that examines the interplay of risk and prognostic factors with pathology tissue-based biomarkers of health and disease in human populations. Although molecular pathology had been integrated into epidemiologic research for decades, it was only in 2010 when the integrative field that unified molecular pathology and epidemiology was first described in the literature (1). Since then, the field of MPE has expanded considerably to advance population health sciences (2-4).

As a sub-discipline of epidemiology, MPE studies are usually drawn from larger prospective cohort (e.g., Nurses' Health Study, Cancer Prevention Study-II) or case-control (e.g., Breast and Colon Cancer Family Registries) studies that are supported to collect and to use pathology specimens. In an MPE paradigm, a potential etiologic factor is assessed with risk of an outcome across strata of molecular characteristics for the disease-of-interest. More recently, MPE resources have matured to allow examination of the independent and joint influences of endogenous/lifestyle/behavioral factors and tissue-based molecular markers on patient prognosis and related outcomes (5). The underlying premise with an incidence study in an MPE paradigm is diseases that have certain molecular perturbations in common are more likely to share a common cause (or causes); similarly, for survival studies, it is postulated that endogenous/lifestyle/behavioral factors differentially influence prognosis according to molecular signatures of the disease because those factors likely interact with the diseased cells in the local microenvironment. To date, MPE has been largely employed

as a method to assess neoplastic disease heterogeneity (e.g., cancers of the colorectum, breast, and lung, in particular); however, MPE methods are broadly applicable to examining any complex disease or health condition (6). The MPE concept as a single, integrative field has obtained increased recognition in recent years (7-16) due to the increased ability to molecularly characterize tumors.

One of the leading opportunities in MPE research is the ability to better predict disease occurrence and prognosis compared to the more conventional disease entities without molecular classification. The evidence linking cigarette smoking and colorectal cancer illustrates this point. Whereas the association between smoking and lung cancer is robust, with relative risks (RRs) that often approach 10 when comparing long-term smokers to non-smokers (17), the link between smoking and colorectal cancer overall is much more modest, with RRs usually below 1.2 (18). Indeed, colorectal cancer was added to the list of smoking-associated cancers only in 2009 (19), more than five decades after the link between smoking and lung cancer mortality was discovered. Part of the obfuscation for this association is caused by tumor heterogeneity between individuals: colorectal cancers do not arise through a singular, homogeneous, canonical pathway. Instead, several major sources of genomic instability, which are not mutually exclusive, contribute to malignant transformation of colorectal epithelial cells: chromosomal instability, microsatellite instability (MSI), epigenomic instability (e.g., CpG island methylator phenotype, or CIMP) and somatically acquired point mutations, indels and copy number alterations (20). Once stratified by MSI or CIMP status, several MPE studies have demonstrated an approximate doubling of risk between smoking and the rarer MSI-high (~15% of colorectal cancers) and CIMP-high (~20% of colorectal

cancers) subtypes of colorectal cancer and quite consistently null associations have been shown for colorectal cancers not bearing those phenotypes (21-24). Beyond identifying previously unknown exposure-subtype disease associations and supporting causality, MPE studies may also identify disease subtypes that benefit from certain behavioral or pharmacologic interventions and discover/validate molecular markers for risk assessment, early detection, prognosis and prediction (6).

MPE studies also have several limitations that are common to observational research in general and to epidemiology in particular when sub-group analyses are performed, including the potential for bias (e.g., selection bias), limited generalizability, low statistical power (and the related issue of low risk estimate precision), multiple testing leading to potentially spurious findings, and the potential for measurement errors of the molecular phenotypes of interest. MPE also faces unique challenges, including the lack of researchers with transdisciplinary expertise in epidemiology, biostatistics, bioinformatics, molecular biology, pathology, and computational biology. To address these challenges, highlight recent successes in the field, and to identify new opportunities, a series of MPE meetings have been held at the Dana-Farber Cancer Institute in Boston, MA. The first was a closed meeting with 10 attendees held in April 2013. The second meeting was open to the international scientific community, with 150 attendees from 16 countries, and held in December 2014 (25). Herein, we share the proceedings of the Third International Molecular Pathological Epidemiology (MPE) Meeting, held in May 2016 and attended by 150 scientists from 17 countries. The session topics, speakers, and essential references are summarized in Table 1.

Day 1: May 12, 2016

The meeting started with co-Chairs Drs. Shuji Ogino and Peter Campbell who welcomed participants and gave a brief history of the MPE meetings as well as an overview of the meeting's theme ('Concepts, Tools and Practice') and goals.

Updates of MPE and MPE Pooling Projects

The first session of the day was on 'Updates for MPE', with Dr. Martha Slattery presenting recent data from the Diet, Activity, and Lifestyle Study (DALS) (26) on microRNA (miRNA) profiling in normal and neoplastic tissue. MiRNAs are small, non-protein-coding RNA molecules that regulate gene expression either by post-transcriptionally suppressing mRNA translation or by causing mRNA degradation. In the study, Dr. Slattery and her colleagues profiled 1893 colorectal cancer/normal paired samples and 290 adenoma tissue samples on the Agilent human miRNA Microarray and she gave a thorough overview of some of the methodologic challenges in dealing with the abundance of data generated by these sorts of platforms, including the difficulty in selecting bioinformatics tools, interpretation of complex results, and validation of key findings. One of the key findings from her data was that miRNAs might influence rectal cancer survival outcomes more-so than colon cancer survival.

Keeping with the theme of colorectal cancer MPE, Dr. Ulrike Peters gave a lecture updating the efforts of the large 'GECCO (Genetics and Epidemiology of Colorectal Cancer Consortium) MPE' project that aims to use targeted sequencing of about 200 genes known to be somatically mutated in 50 genomic regions for copy

number alterations, and bacterial genes. The content for the targeted sequencing panel was selected based on comprehensive analysis of large scale whole exome sequencing data from The Cancer Genome Atlas (TCGA) (27) and efforts in the Nurses' Health Study and Health Professional Follow up study (28, 29). The GECCO consortium aims to test 4200 tumors and 3500 normal tissue samples in addition to harmonizing existing tumor molecular phenotype data (e.g., MSI, CIMP, *BRAF*-mutation, *KRAS*-mutation) from over 10,000 cases. Dr. Peters presented preliminary data from the first 198 cases and discussed the bioinformatics pipeline created for the project. In combination with the expansive clinical, epidemiologic and GWAS data from GECCO, which has been underway since 2009, the addition of tumor molecular phenotype data will create a rich resource for MPE discovery and validation.

Cancer Immunology

The second session focused on cancer immunology (chaired by Dr. Amanda Phipps). Dr. Harlan Robins presented a lecture on learning to read immunological memory using immune-sequencing (30). Part of the impetus for this work is to acquire the ability to detect small neoplastic clones that remain in circulation after treatment, which can indicate disease recurrence. He presented work to show that high clonality was associated with better prognosis for hematologic cancer patients. Targeted sequencing of T- and B-cells is likely a fruitful area for multiple tumors to investigate the impact of epidemiological preventive and risk factors on the immune response of the tumor and impact of the immune response on survival outcomes.

Dr. Gordon Freeman presented a lecture describing how pathological data analysis may help to guide immunotherapy. He showed that approximately 30% of solid tumors and selected hematologic malignancies are positive for CD274 (PD-L1), an immune checkpoint ligand that is expressed by tumor and immune cells. The CD274-PDCD1 (PD-1) immune checkpoint pathway is also becoming recognized as a promising therapeutic target in various tumor types. Pathological analysis of tumor immunity status including immune checkpoint has been applied to population-based research (31-33), which can provide new insights on variations of tumor immunity and influence of various exposures.

The final lecture in the cancer immunology session was given by Dr. Ogino who described how innate and adaptive immunity plays a critical role in health and disease. All pathologic processes involve interactions between multiple cell types, including immune cells, in response to environmental exposures in the tissue microenvironment. Hence integrated analyses of exposures, tumor molecular features, and immune characteristics are important to better understand disease processes such as cancers (34). Epidemiology has had more recent successes with, for example, germline genetics than with immunology, but the opportunities are now rich in this field (29, 35, 36). Dr. Ogino introduced how MPE methods can and should be applied to the study of immune-based disease subtyping and provided an example of an association for high dietary marine omega-3 polyunsaturated fatty acids lowering risk of colorectal cancer with high-density FOXP3+ regulatory T cells (37).

MPE of Cancer Disparities

The next session was focused on the MPE of cancer disparities and chaired by Dr. Xuehong Zhang with one presenter, Dr. Timothy Rebbeck. Dr. Rebbeck's lecture gave an overview of cancer disparities for prostate cancer mortality (38). It is well known that African American men have approximately double the rate of prostate cancer-specific death compared to white men. Because prostate cancer is a disease with a complex, multifactorial etiology, it is clear that the disparity is explained by genomics, biochemistry and physiology, exposure, behavior, and social context, among other factors. Models of cancer etiology and disparities therefore need to consider the potential interaction of all of these factors to understand health and disease. Perhaps more importantly, it may be necessary to use these multiple etiological factors to redefine the disparity. That is, using race or ethnicity in defining disparities has been of value, but it is clear that this classification is misclassified with respect to the entities that may be most relevant for the development and implementation of interventions (e.g., cancer prevention). MPE provides a unique framework around which this complexity can be studied, and therefore provides a useful means to understand cancer disparities for diseases like prostate cancer.

Computational Modelling and Molecular Imaging

The final session of the morning was devoted to 'special topics' with two speakers and moderated by Dr. Rebbeck. Dr. Franziska Michor presented a lecture on computational modeling to analyze single-cell data obtained from biopsy and surgical samples of breast cancer patients (39-41). In collaboration with Dr. Kornelia Polyak from

the Dana-Farber Cancer Institute, Dr. Michor developed evolutionary stochastic modeling techniques to determine the effects of intra-tumor heterogeneity on responses to neoadjuvant chemotherapy. Furthermore, such spatially explicit computational modeling techniques can be used to identify mechanisms of tumor evolution, for instance whether different subtypes of breast cancer have intrinsically different migration rates as identified from clinical specimens. Her methods are also being applied to other cancer and treatment types.

The second 'special topics' lecture was on molecular imaging in future population screening and given by Dr. Alexei Bogdanov. Molecular imaging is the noninvasive *in vivo* investigation of cellular and molecular events involved in normal and pathologic processes (42). While the technology is not yet apt for population-wide screening, some aspects of diagnostic agent (label-free) and label-enabled molecular imaging techniques have evolved to play substantial roles in prospective studies. In those studies, patient safety, compliance, add-on time and cost are of paramount importance. In this regard, high-field strength clinical magnetic resonance imagining (MRI) spectroscopy has a proven record of providing molecular information regarding the levels of key metabolites in normal and cancer tissues to aid with differential diagnostic decisions. An example was offered with early stage prostate cancer and breast cancer detection.

Statistical Advances in MPE

After a two-hour pause in podium presentations for attendees to review poster presentations, the afternoon sessions began with statistical advances in MPE (chaired by Dr. Molin Wang). Dr. Colin Begg lectured on his group's efforts toward developing

sub-type specific models for absolute risk in cancer studies (43). Using examples from bilateral primary breast cancer data, he showed how this resource provides unique insights about cancer heterogeneity that cannot be discerned from traditional case-control or cohort studies.

In the second presentation in this session, Dr. Tyler VanderWeele considered the problem of assessing mechanistic interaction when an outcome is ordinal as is the case with etiologic or outcome heterogeneity. Such mechanistic interaction between two exposures is said to be present when for some individuals a particular outcome level or subclass will occur if both exposures are present but will not if only one of the two exposures is present (44). Conditions for such mechanistic interaction for a binary outcome have been derived previously and do not in general coincide with the presence of an interaction term in a statistical model (45, 46); generally stronger empirical conditions are needed for the conclusion of mechanistic than simply a non-zero product term in a statistical model. New empirical conditions are derived for the setting when the outcome is ordinal. It is shown that the new conditions cannot be derived simply by recoding the ordinal outcomes as a series of dichotomizations. The methods are useful in assessing the types of interactions between exposures that may give rise to etiologic heterogeneity in the study of molecular pathological epidemiology.

Computational Biology

The next session focused on computational biology (chaired by Dr. Reiko Nishihara). Dr. John Quackenbush started off the session with a lecture on patient-specific gene regulatory networks (47). The central hypothesis to this work is that

unique gene regulatory processes define biological phenotypes, including those of populations with diseases. His group's work over the past few years has produced powerful methods for inferring population-level gene regulatory networks, for comparing those networks between phenotypes, and for understanding biological properties of those phenotypes based on the features of the networks. In his lecture, he presented a simple extension to those models that allow one to deduce gene regulatory network models for each individual in a population. Further, he showed that these patient-specific networks are predictive of important biological endpoints, and in many cases are more predictive than widely-used biomarkers. More importantly, these patient-specific networks might provide a path to a more directed and individualized approach to treatment.

In the second lecture in this session, Dr. Rafael Irizarry presented data on high dimensional epigenomic analysis wherein he described some of the statistical and biological challenges related to detecting differentially methylated genomic regions (48). He described the important role of modern statistical techniques in finding regions of the genome that are consistently different between diseased and normal groups and some new challenges that are specifically related to next-generation sequencing data. He also described the importance of considering batch effects which often exist in high-throughput data.

Dr. Curtis Huttenhower gave a lecture on high-precision functional profiling of microbial communities and the human microbiome**.** Human gut microbial dysbioses have been associated with diseases ranging from autism to cancer, but the causative

molecular or ecological mechanisms are unclear (49). He presented end-to-end methodologies for functional surveys of the microbiome in human population studies, beginning with scalable sample collection and including computational tools and downstream statistical analyses. These have, to date, identified potentially causal microbial mechanisms in inflammatory bowel disease and type 1 diabetes mellitus, and he discussed strategies for future applications and open questions in colorectal cancer, nutrition, and the microbiome more broadly in public health.

Proffered Abstracts: Part 1

The last session of the first day, chaired by Dr. Elizabeth Poole, was devoted to proffered papers from submitted abstracts.

Ms. Emily Zabor reviewed statistical methods for evaluating etiologic heterogeneity. One focus of MPE is the classification of diseases into subtypes based on molecular and pathological characteristics, and subsequent application of epidemiologic methods to study the resulting subtypes. A particular interest of epidemiologists is the etiologic heterogeneity of the sub-types, i.e. differences across subtypes with respect to the influence of risk factors. She reviewed a variety of methods that have been proposed to study etiologic heterogeneity, including the standard polytomous regression approach (50), a method that incorporates subtype discovery with a scalar measure of the degree of heterogeneity (51), several two-stage regression approaches that are applicable both to cohort and case-control studies (52, 53), a single-stage regression approach that introduces the concept of an adjusted hazard

ratio (54), and a method that allows for non-mutually exclusive subtype membership (55).

Dr. Daniel Nevo lectured on his work in dealing with missing subtypes using auxiliary case covariates. A competing risks proportional hazard model is often used in the analysis of time-to-disease data to assess risk factor associations on different disease subtypes (56, 57). Cases with missing subtypes are often ignored, and even when these cases are included, the analysis is typically based on a missing-at-random assumption. For example, colorectal tumors that harbor molecular perturbations that are linked to poorer prognosis may be less likely to be accrued, leading to a potential bias of the observed risk factor and subtype associations. He described a method to conduct valid analyses when additional auxiliary variables are measured for cases. The method exploits the fact that distribution of the auxiliary case covariates differs according to the molecular subtype. He illustrated the use of the new method in the analysis of colorectal cancer data from the Nurses' Health Study. The auxiliary covariate was tumor location, which is commonly accrued for most cases. The method used the fact that a proximal tumor location is more likely to occur among MSI-high subtype tumors compared to microsatellite stable subtypes to correct potential bias.

Dr. Yujing Jan Heng gave a lecture on molecular analyses of histopathologic features in breast cancer**.** Her group collected histopathologic annotation of invasive breast cancer cases in the TCGA (58) and integrated TCGA’s molecular data with breast cancer histopathologic annotations to elucidate the molecular basis of common morphologic features. Her study found that certain molecular features in breast cancer

were associated with the PAM50 Basal-like subtype. They also used omics-based multivariate models to assess the association of morphologic signatures with survival in ESR1 (ER-alpha)-positive and ESR1 (ER-alpha)-negative breast cancer using six independent datasets. They identified that a transcriptomic signature of poorly differentiated epithelial tubule formation adds prognostic information in ESR1-positive beyond pathologic assessment of clinical grade.

Day 2: May 13, 2016

Proffered Abstracts: Part 2

The second day of the meeting began with another session for proffered abstracts as well as the announcement of trainee awards at the student/post-doctoral and early career levels (chaired by Dr. N. Sertac Kip).

Dr. Daniel Xia lectured on the role of computational pathology to identify stromal inflammation as a prognostic biomarker in squamous cell carcinoma (SCC) of the lung. His group did high-throughput computational digital image analyses on tissue microarray (TMA) samples of tumors from lung SCC cases, with the goal of identifying epithelial and stromal histologic features associated with survival and they identified a stromal inflammation (SI) score was prognostic for patient survival outcomes. The SI histologic score was positively correlated with the expression of genes involved in the adaptive immune response in TCGA data.

Dr. Peter Rogan gave a lecture on cisplatin response in recurrent bladder cancer with biochemically-inspired machine learning. The ability to predict response to

chemotherapy could help with drug selection and dosing, possibly reduce toxicity, and improve outcomes. Using a machine learning approach (59), models were developed for the prediction of cisplatin chemotherapy response in bladder cancer patients. Thegene expression signatures were validated in TCGA patients. Machine learning experiments identified gene sets that were enriched for genes belonging to DNA repair, anti-oxidative response, and metal binding pathways.

Dr. Helen Coleman presented work on low-dose aspirin, PTGS2 expression and survival in colon cancer patients using data from Northern Ireland. The association between high-dose aspirin use and improved survival after colorectal cancer diagnosis may be more pronounced for patients who have tumors with high prostaglandin endoperoxide synthase 2 (PTGS2, cyclooxygenase-2) expression (60). The interaction between PTGS2 and low-dose aspirin is less clear (61, 62). QuPath image analysis software assessed immunohistochemical expression of PTGS2 in TMAs. Clinical follow-up data were obtained through the Northern Ireland Cancer Registry. Compared to nonusers, low-dose aspirin users had lowered risks of cancer-specific mortality and all-cause mortality. Low dose aspirin use was associated with improved overall survival for tumors that overexpressed PTGS2 but not for tumors with weaker PTGS2 expression.

Dr. Yin Cao gave a lecture on aspirin use and risk of colorectal cancer according to tumor immune reaction. She hypothesized that aspirin use might be associated with lower risk of colorectal cancers that demonstrated less immune response because of aspirin’s immune-enhancing effects. Aspirin use data were collected in the Nurses' Health Study and Health Professionals Follow-up Study (63). The inverse association of

regular aspirin use with colorectal cancer risk differed by the degree of tumor infiltrating lymphocytes (TILs). Compared with non-regular use, regular aspirin-use was associated with lower risk of tumor with low-level TILs, and the strength of the association was dose- and duration-dependent. In contrast, aspirin use was not associated with risk of tumors with intermediate-level or high-level TILs. These results suggest a potential role of host immunity in mediating the chemopreventive effect of aspirin.

Open discussion Part 1: illustration of MPE studies from beginning to end

In the first of two open discussion sessions, Dr. Campbell illustrated how tissue collection can be initiated from within an existing prospective cohort study and how existing MPE resources can be leveraged to build an early career in MPE. In the first part of his lecture, the basic methods and challenges faced in creating a tissue repository from an ongoing prospective cohort study, the Cancer Prevention Study-II, were outlined, including the timeline, methodology and resources required for such endeavors (64). Study participants with and without tissue materials were compared on a series of epidemiologic and clinical factors and few differences were found. One of the main challenges in collecting tissue materials in this context was that surgical tissue materials are usually destroyed by hospitals 10-years after the patient's diagnosis. In the second part of the lecture, he reviewed work that primarily focused on real-world challenges involved with complex MPE studies, notably the vast bioinformatics resources, from the Colon Cancer Family Registry (65-68) as an example toward building an early career in MPE from an existing resource.

Open Discussion Part 2: general issues relevant to MPE

In the second open discussion session, Drs. Heng, Kip and Mingyang Song discussed the paucity of interdisciplinary education and training opportunities in MPE and how they overcame these obstacles in their own careers. Drs. Begg and Wang led an open discussion on study design and statistical challenges for MPE with lots of feedback from audience members on their own experiences. Drs. Andrew Beck, Kip and Zhi Rong Qian discussed opportunities and challenges in the disconnect between pathology (focused on the singular patient) and epidemiology (focused on populations). Drs. Peters and Stephanie Smith-Warner led a discussion on consortia building, based on their experiences with GECCO and various NCI-Cohort Consortium projects. Dr. Rebbeck led the group in a discussion on cancer health disparities and how MPE may address some of these issues. The session came to a close with Drs. Beck, Bogdanov and Poole leading a discussion on emerging technologies and new areas of investigation.

Conclusions

The Third International MPE Meeting successfully assembled 150 trainees and experts working in complementary fields of this rather new scientific discipline. As the heterogeneity of pathogenic processes in human complex diseases becomes better appreciated and understood, the MPE paradigm should become more ubiquitous in the future for many areas of clinical medicine and population health sciences. We look forward to again sharing our experiences, successes and challenges at the Fourth International MPE Meeting, which is planned to be held in the spring of 2018.

References

1. Ogino S, Stampfer M. (2010) Lifestyle factors and microsatellite instability in colorectal cancer: the evolving field of molecular pathological epidemiology. J Natl Cancer Inst. 102: 365-7.
2. Ogino S, Lochhead P, Chan AT, et al. (2013) Molecular pathological epidemiology of epigenetics: emerging integrative science to analyze environment, host, and disease. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc. 26: 465-84.
3. Nishi A, Kawachi I, Koenen KC, Wu K, Nishihara R, Ogino S. (2015) Lifecourse epidemiology and molecular pathological epidemiology. Am J Prev Med. 48: 116-9.
4. Nishi A, Milner DA, Jr., Giovannucci EL, et al. (2016) Integration of molecular pathology, epidemiology and social science for global precision medicine. Expert Rev Mol Diagn. 16: 11-23.
5. Ogino S, Chan AT, Fuchs CS, Giovannucci E. (2011) Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. Gut. 60: 397-411.
6. Ogino S, Nishihara R, VanderWeele TJ, et al. (2016) Review Article: The Role of Molecular Pathological Epidemiology in the Study of Neoplastic and Non-neoplastic Diseases in the Era of Precision Medicine. Epidemiology. 27: 602-11.
7. Curtin K, Slattery ML, Samowitz WS. (2011) CpG island methylation in colorectal cancer: past, present and future. Pathology research international. 2011: 902674.
8. Bishehsari F, Mahdavinia M, Vacca M, Malekzadeh R, Mariani-Costantini R. (2014) Epidemiological transition of colorectal cancer in developing countries: environmental factors, molecular pathways, and opportunities for prevention. World J Gastroenterol. 20: 6055-72.
9. Jiang MJ, Dai JJ, Gu DN, Huang Q, Tian L. (2016) Aspirin in pancreatic cancer: chemopreventive effects and therapeutic potentials. Biochim Biophys Acta.
10. Kuipers EJ, Grady WM, Lieberman D, et al. (2015) Colorectal cancer. Nat Rev Dis Primers. 1: 15065.
11. Martinez-Useros J, Garcia-Foncillas J. (2016) Obesity and colorectal cancer: molecular features of adipose tissue. Journal of translational medicine. 14: 21.
12. Kuroiwa-Trzmielina J, Wang F, Rapkins RW, et al. (2016) SNP rs16906252C>T is an expression and methylation quantitative trait locus associated with an increased risk of developing MGMT-methylated colorectal cancer. Clin Cancer Res.
13. Chia WK, Ali R, Toh HC. (2012) Aspirin as adjuvant therapy for colorectal cancer--reinterpreting paradigms. Nat Rev Clin Oncol. 9: 561-70.
14. Hughes LA, Khalid-de Bakker CA, Smits KM, et al. (2012) The CpG island methylator phenotype in colorectal cancer: progress and problems. Biochim Biophys Acta. 1825: 77-85.
15. Buchanan DD, Win AK, Walsh MD, et al. (2013) Family history of colorectal cancer in BRAF p.V600E-mutated colorectal cancer cases. Cancer Epidemiol Biomarkers Prev. 22: 917-26.
16. Figueiredo JC, Lewinger JP, Song C, et al. (2011) Genotype-environment interactions in microsatellite stable/microsatellite instability-low colorectal cancer: results from a genome-wide association study. Cancer Epidemiol Biomarkers Prev. 20: 758-66.
17. Lee PN, Forey BA, Coombs KJ. (2012) Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. BMC Cancer. 12: 385.
18. Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. (2008) Smoking and colorectal cancer: a meta-analysis. JAMA. 300: 2765-78.
19. Secretan B, Straif K, Baan R, et al. (2009) A review of human carcinogens--Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. Lancet Oncol. 10: 1033-4.
20. Jass JR. (2007) Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology. 50: 113-30.

21. Poynter JN, Haile RW, Siegmund KD, et al. (2009) Associations between smoking, alcohol consumption, and colorectal cancer, overall and by tumor microsatellite instability status. Cancer Epidemiol Biomarkers Prev. 18: 2745-50.
22. Slattery ML, Curtin K, Anderson K, et al. (2000) Associations Between Cigarette Smoking, Lifestyle Factors, and Microsatellite Instability in Colon Tumors. J Natl Cancer Inst. 92: 1831-6.
23. Limsui D, Vierkant RA, Tillmans LS, et al. (2010) Cigarette smoking and colorectal cancer risk by molecularly defined subtypes. J Natl Cancer Inst. 102: 1012-22.
24. Nishihara R, Morikawa T, Kuchiba A, et al. (2013) A prospective study of duration of smoking cessation and colorectal cancer risk by epigenetics-related tumor classification. Am J Epidemiol. 178: 84-100.
25. Ogino S, Campbell PT, Nishihara R, et al. (2015) Proceedings of the second international molecular pathological epidemiology (MPE) meeting. Cancer Causes Control. 26: 959-72.
26. Curtin K, Samowitz WS, Ulrich CM, et al. (2011) Nutrients in folate-mediated, one-carbon metabolism and the risk of rectal tumors in men and women. Nutr Cancer. 63: 357-66.
27. Cancer Genome Atlas N. (2012) Comprehensive molecular characterization of human colon and rectal cancer. Nature. 487: 330-7.
28. Giannakis M, Hodis E, Jasmine Mu X, et al. (2014) RNF43 is frequently mutated in colorectal and endometrial cancers. Nat Genet. 46: 1264-6.
29. Giannakis M, Mu XJ, Shukla SA, et al. (2016) Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. Cell Rep.
30. Howie B, Sherwood AM, Berkebile AD, et al. (2015) High-throughput pairing of T cell receptor alpha and beta sequences. Sci Transl Med. 7: 301ra131.
31. Masugi Y, Nishihara R, Yang J, et al. (2016) Tumour CD274 (PD-L1) expression and T cells in colorectal cancer. Gut.
32. Mima K, Nishihara R, Nowak JA, et al. (2016) MicroRNA MIR21 and T Cells in Colorectal Cancer. Cancer Immunol Res. 4: 33-40.
33. Mima K, Sukawa Y, Nishihara R, et al. (2015) Fusobacterium nucleatum and T Cells in Colorectal Carcinoma. JAMA Oncol. 1: 653-61.
34. Ogino S, Galon J, Fuchs CS, Dranoff G. (2011) Cancer immunology--analysis of host and tumor factors for personalized medicine. Nat Rev Clin Oncol. 8: 711-9.
35. Song M, Nishihara R, Wang M, et al. (2016) Plasma 25-hydroxyvitamin D and colorectal cancer risk according to tumour immunity status. Gut. 65: 296-304.
36. Hanyuda A, Ogino S, Qian ZR, et al. (2016) Body mass index and risk of colorectal cancer according to tumor lymphocytic infiltrate. Int J Cancer. 139: 854-68.
37. 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.
38. Tan DS, Mok TS, Rebbeck TR. (2016) Cancer Genomics: Diversity and Disparity Across Ethnicity and Geography. J Clin Oncol. 34: 91-101.
39. Almendro V, Cheng YK, Randles A, et al. (2014) Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. Cell Rep. 6: 514-27.
40. Almendro V, Kim HJ, Cheng YK, et al. (2014) Genetic and phenotypic diversity in breast tumor metastases. Cancer Res. 74: 1338-48.
41. Janiszewska M, Liu L, Almendro V, et al. (2015) In situ single-cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2-positive breast cancer. Nat Genet. 47: 1212-9.
42. Bogdanov AA, Jr., Dixon AJ, Gupta S, et al. (2016) Synthesis and Testing of Modular Dual-Modality Nanoparticles for Magnetic Resonance and Multispectral Photoacoustic Imaging. Bioconjug Chem. 27: 383-90.

43. Begg CB, Orlow I, Zabor EC, et al. (2015) Identifying Etiologically Distinct Sub-Types of Cancer: A Demonstration Project Involving Breast Cancer. Cancer Med. 4: 1432-9.
44. VanderWeele TJ, Robins JM. (2007) The identification of synergism in the sufficient-component-cause framework. Epidemiology. 18: 329-39.
45. VanderWeele TJ. (2009) Sufficient cause interactions and statistical interactions. Epidemiology. 20: 6-13.
46. Vanderweele TJ. (2010) Sufficient cause interactions for categorical and ordinal exposures with three levels. Biometrika. 97: 647-59.
47. Glass K, Quackenbush J, Spentzos D, Haibe-Kains B, Yuan GC. (2015) A network model for angiogenesis in ovarian cancer. BMC Bioinformatics. 16: 115.
48. Jaffe AE, Irizarry RA. (2014) Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. 15: R31.
49. Huttenhower C, Knight R, Brown CT, et al. (2014) Advancing the microbiome research community. Cell. 159: 227-30.
50. Dubin N, Pasternack BS. (1986) Risk assessment for case-control subgroups by polychotomous logistic regression. Am J Epidemiol. 123: 1101-17.
51. Begg CB, Zabor EC, Bernstein JL, Bernstein L, Press MF, Seshan VE. (2013) A conceptual and methodological framework for investigating etiologic heterogeneity. Stat Med. 32: 5039-52.
52. Chatterjee N, Sinha S, Diver WR, Feigelson HS. (2010) Analysis of cohort studies with multivariate and partially observed disease classification data. Biometrika. 97: 683-98.
53. Wang M, Kuchiba A, Ogino S. (2015) A Meta-Regression Method for Studying Etiological Heterogeneity Across Disease Subtypes Classified by Multiple Biomarkers. Am J Epidemiol. 182: 263-70.
54. Rosner B, Glynn RJ, Tamimi RM, et al. (2013) Breast cancer risk prediction with heterogeneous risk profiles according to breast cancer tumor markers. Am J Epidemiol. 178: 296-308.
55. Schildkraut JM, Iversen ES, Akushevich L, et al. (2013) Molecular signatures of epithelial ovarian cancer: analysis of associations with tumor characteristics and epidemiologic risk factors. Cancer Epidemiol Biomarkers Prev. 22: 1709-21.
56. Nevo D, Zucker DM, Tamimi RM, Wang M. (2016) Accounting for measurement error in biomarker data and misclassification of subtypes in the analysis of tumor data. Stat Med.
57. Wang M, Spiegelman D, Kuchiba A, et al. (2016) Statistical methods for studying disease subtype heterogeneity. Stat Med. 35: 782-800.
58. Cancer Genome Atlas N. (2012) Comprehensive molecular portraits of human breast tumours. Nature. 490: 61-70.
59. Dorman SN, Baranova K, Knoll JH, et al. (2016) Genomic signatures for paclitaxel and gemcitabine resistance in breast cancer derived by machine learning. Mol Oncol. 10: 85-100.
60. Chan AT, Ogino S, Fuchs CS. (2009) Aspirin use and survival after diagnosis of colorectal cancer. JAMA. 302: 649-58.
61. Reimers MS, Bastiaannet E, Langley RE, et al. (2014) Expression of HLA class I antigen, aspirin use, and survival after a diagnosis of colon cancer. JAMA internal medicine. 174: 732-9.
62. Domingo E, Church DN, Sieber O, et al. (2013) Evaluation of PIK3CA mutation as a predictor of benefit from nonsteroidal anti-inflammatory drug therapy in colorectal cancer. J Clin Oncol. 31: 4297-305.
63. 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.
64. Campbell PT, Deka A, Briggs P, et al. (2014) Establishment of the cancer prevention study II nutrition cohort colorectal tissue repository. Cancer Epidemiol Biomarkers Prev. 23: 2694-702.

65. 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.
66. 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: 1229-38.
67. Newcomb PA, Baron J, Cotterchio M, et al. (2007) Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. Cancer Epidemiol Biomarkers Prev. 16: 2331-43.
68. Phipps AI, Ahnen DJ, Campbell PT, et al. (2014) Family history of colorectal cancer is not associated with colorectal cancer survival regardless of microsatellite instability status. Cancer Epidemiol Biomarkers Prev. 23: 1700-4.

Table 1: Summary of podium presentations at the 3rd International Molecular Pathological Epidemiology (MPE) Meeting in Boston, MA on May 12 and 13, 2016.

Session and speaker name	Main topic addressed	Relevant references
Updates of MPE and MPE Pooling Projects		
Martha Slattery	MicroRNA and colorectal cancer	26
Ulrike Peters	Update on a large MPE project for colorectal cancer	27-29
Cancer Immunology		
Harlan Robins	Immunological memory in hematologic cancers	30
Gordon Freeman	Immune checkpoint in hematologic cancers	31-33
Shuji Ogino	Immuno-MPE of colorectal cancer	29, 34-37
MPE of Cancer Disparities		
Timothy Rebbeck	MPE of cancer disparities in prostate cancer	38
Special Topics: Computational Modelling and Molecular Imaging		
Franziska Michor	Computational modelling of breast cancer	39-41
Alexei Bogdanov	Molecular imaging of breast cancer	42
Statistical Advances in MPE		
Colin Begg	Approaches for identifying molecular subtypes of breast cancer	43
Tyler VanWeele	Causal interactions for outcome heterogeneity	44-46
Computational Biology		
John Quackenbush	Gene networks in patients and populations	47
Rafael Irizarry	High dimensional epigenomic analysis	48
Curtis Huttenhower	Microbial communities and the microbiome	49
Proffered abstracts: part 1		
Emily Zabor	Statistical methods for identifying heterogeneity	50-55
Daniel Nevo	Dealing with missing sub-types in analyses	56, 57
Yujing Jan Heng	Integrative analysis of breast cancer	58
Proffered abstracts: part 2		
Daniel Xia	Prognostic biomarkers for lung cancer	
Peter Rogan	Machine learning for prognosis of bladder cancer	59
Helen Coleman	Aspirin and tissue expression markers in colon cancer survival	60-62
Yin Cao	Aspirin and the tumor immune reaction in colorectal cancer	63
Open discussion Part 1: illustration of MPE studies from beginning to end		
Peter Campbell	Building an early career (and resources) in MPE studies	64-68
Open Discussion Part 2: general issues relevant to MPE		
Jan Heng, Sertac Kip, Mingyang Song	Interdisciplinary education and training	
Colin Begg, Molin Wang	Study design and statistical methods	
Andrew Beck, Sertac Kip, Zhi Rong Qian	Opportunities and challenges in pathology in epidemiologic research	
Ulrike Peters, Stephanie Smith-Warner	MPE consortium building	
Timothy Rebbeck	Opportunities and challenges in health disparities and MPE	
Andrew Beck, Alexei Bogdanov, Liz Poole	Emerging technologies, new areas of investigation, and our future	