The Japanese Society of Gastroenterology

CrossMark

REVIEW

Molecular pathological epidemiology: new developing frontiers of big data science to study etiologies and pathogenesis

Tsuyoshi Hamada[1] · NaNa Keum[2] · Reiko Nishihara[1,2,3,4] · Shuji Ogino[1,3,5,6]

Abstract Molecular pathological epidemiology (MPE) is an integrative field that utilizes molecular pathology to incorporate interpersonal heterogeneity of a disease process into epidemiology. In each individual, the development and progression of a disease are determined by a unique combination of exogenous and endogenous factors, resulting in different molecular and pathological subtypes of the disease. Based on "the unique disease principle," the primary aim of MPE is to uncover an interactive relationship between a specific environmental exposure and disease subtypes in determining disease incidence and mortality. This MPE approach can provide etiologic and pathogenic insights, potentially contributing to precision medicine for personalized prevention and treatment. Although breast, prostate, lung, and colorectal cancers have been among the most commonly studied diseases, the MPE approach can be used to study any disease. In addition to molecular features, host immune status and microbiome profile likely affect a disease process, and thus serve as informative biomarkers. As such, further integration of several disciplines into MPE has been achieved (e.g., pharmaco-MPE, immuno-MPE, and microbial MPE), to provide novel insights into underlying etiologic mechanisms. With the advent of high-throughput sequencing technologies, available genomic and epigenomic data have expanded dramatically. The MPE approach can also provide a specific risk estimate for each disease subgroup, thereby enhancing the impact of genome-wide association studies on public health. In this article, we present recent progress of MPE, and discuss the importance of accounting for the disease heterogeneity in the era of big-data health science and precision medicine.

Keywords Colorectal neoplasms · Drug therapy · Immunology · Microbiology · Personalized therapy

T. Hamada and N. Keum contributed equally as co-first authors and R. Nishihara and S. Ogino contributed equally as co-last authors.

✉ Reiko Nishihara
rnishiha@hsph.harvard.edu

✉ Shuji Ogino
shuji_ogino@dfci.harvard.edu

[1] Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, 450 Brookline Ave., Room SM1036, Boston, MA 02215, USA

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

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

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

[5] Division of 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

[6] Department of Oncologic Pathology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA

Abbreviations
25(OH)D	25-hydroxyvitamin D
CIMP	CpG island methylator phenotype
GWAS	Genome-wide association studies
MPE	Molecular pathological epidemiology
MSI	Microsatellite instability
NIH	National Institutes of Health

NSAID   Nonsteroidal anti-inflammatory drug
PUFA   Polyunsaturated fatty acid

Introduction to molecular pathological epidemiology (MPE)

Molecular pathological epidemiology (MPE), which incorporates molecular pathology into epidemiologic research, has emerged as a transdisciplinary field in population health science [1–3]. Conventional epidemiology primarily investigates the relationship between an exposure and a disease entity in population-based cohorts (Fig. 1). This conventional approach assumes that people diagnosed with similar symptoms or disease manifestations represent a homogeneous group (i.e., a single disease entity) and share similar etiologies. However, this reductionist approach might have led to biased [3] or paradoxical findings [4] in which a well-known risk factor was apparently associated with a better prognosis.

In contrast, MPE, by means of applying molecular pathology diagnostics to a disease classification, aims to address inherent heterogeneity in a single traditional disease entity [1, 2]. The MPE paradigm is founded on "the unique disease principle [5]" (or "the unique tumor principle [6, 7]") and "the disease continuum theory [3]". To elaborate, "the unique disease principle [5]" posits that, while people diagnosed with the same disease entity share some similarities, each individual has a unique pathologic process driven by a complex interaction between molecular alterations in cells and the surrounding microenvironment. At the same time, "the disease continuum theory [3]" asserts that people diagnosed with different diseases can have overlapping etiologies and pathogenesis. Moreover, a wide spectrum of inherent factors (e.g., germline genetic variations, sex, ethnicity) and acquired or exogenous factors (e.g., acquired genetic and epigenetic alterations, diet, lifestyle, smoking, medications, microorganisms) can affect the disease process. As a result, significant interpersonal heterogeneity exists in the disease process including initiation, evolution, and progression [3, 5]. In order to address the disease heterogeneity, MPE utilizes molecular pathological signatures that can categorize patients into subgroups [1, 2], so that people in each subgroup share more homogenous etiology and pathogenic process. Through this paradigm shift, MPE enables us to explore whether an exposure forms a differential relationship with disease subgroups classified by molecular biomarkers. Thus, findings from MPE research can provide biological evidence to enhance our understanding of etiologies and pathogenesis of diseases, strengthening evidence for causal relationships [1, 2, 7, 8]. The concept of this unified field of MPE has gained considerable popularity in the literature [9–30].

Based on the growing popularity of molecular pathology assays and increasing recognition of the importance of the precision medicine concept [31, 32], molecular classifications of diseases are widely used in clinical practice.
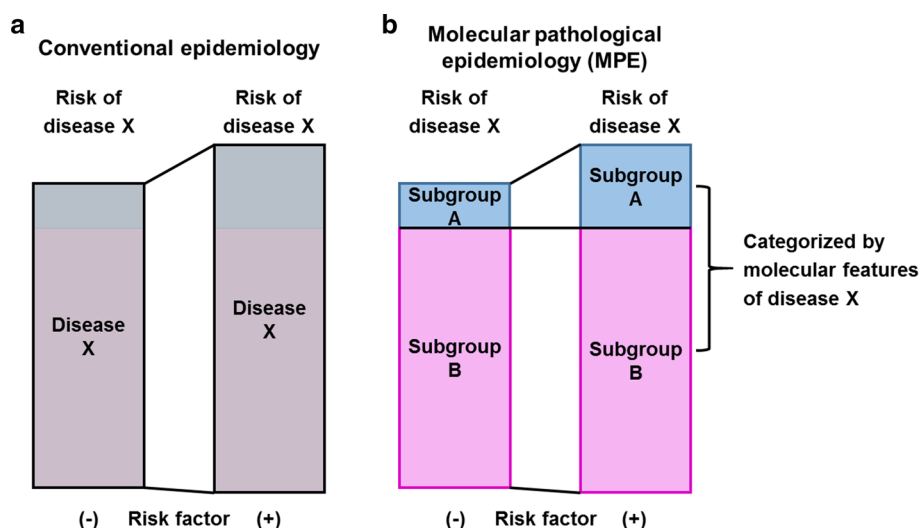

Fig. 1 The paradigm of molecular pathological epidemiology (MPE) research. **a** Scheme of a conventional epidemiology study. The overall association of an exposure with risk of disease X appears to be weak. **b** Scheme of a MPE study. By categorizing disease X into subgroups (A and B) based on molecular pathological features, the significant association of the exposure with risk of subtype A can be revealed. Note that, although we present an example of two disease subgroups for simplicity, more than two disease subgroups can be evaluated in MPE research. Typically, the primary hypothesis in MPE research tests for a difference between the associations of the exposure with subtypes classified by molecular features. MPE, molecular pathological epidemiology

Furthermore, biomedical big data have opened new opportunities to enhance our understanding of disease heterogeneity in humans. Owing to the advent of high-throughput sequencing techniques and analytical methods of genomic data, large-scale biomedical data are increasingly available. Thus, there is a pressing need to establish analytical frameworks to synthesize appropriate knowledge from available big data. Such a need is reflected in the aims of The Big Data to Knowledge (BD2K) and Precision Medicine Initiatives launched by the U.S. National Institutes of Health (NIH), which are in line with the core framework of MPE [31, 32]. The strengths of the unique MPE approach have been well recognized by international symposia [33–35]. The International MPE Meeting Series were established in 2013, and the Second International MPE Meeting was held in 2014 [36], and the Third International MPE Meeting was held in 2016 as a NIH-supported meeting (R13 CA203287, funded by National Cancer Institute, National Human Genome Research Institute, and National Institute of Environmental Health Sciences).

The purposes of this article are to introduce the paradigm of MPE, and to summarize recent advances of MPE. As a major strength of MPE, MPE is a very flexible field that can achieve integrations of other scientific fields such as social science [37], lifecourse epidemiology [38], phamarcoepidemiology, immunology, and microbiology.

Framework of MPE research

The outline of MPE research is illustrated in Fig. 1. MPE has evolved through cancer epidemiology research, owing to early recognition of molecular classification systems and wide availability of tumor tissue specimens. In particular, colorectal cancer has served as a practical model for MPE research [1, 2]. The model of accumulation of genetic and epigenetic alterations was established to explain colorectal carcinogenesis [39]. It has been increasingly evident that colorectal cancer represents a considerably heterogeneous group of neoplasms arising from a unique sequence of genetic and epigenetic alterations in each individual [6, 40–42]. Of note, there is a wide range of etiologic factors for colorectal cancer such as genetic factors, aging, smoking, excessive alcohol consumption, obesity, physical inactivity, dietary factors, diabetes mellitus, inflammatory bowel diseases, and possibly intestinal microbiome [43–46]. To understand specific pathogenic pathways linking diverse etiologic factors to colorectal cancer, it is critical to fully consider potential etiologic heterogeneity. Furthermore, the MPE methodology is readily applicable to research of any human neoplastic [5, 47] and nonneoplastic [48–51] diseases possessing substantial disease heterogeneity [3].

Typical MPE studies are risk analyses or survival analyses that attempt to address heterogeneous relationships between an exposure, and disease incidence or prognosis according to molecular subtypes (Fig. 2) [2]. Based on a hypothesis test for heterogeneity in exposure-disease associations across subgroups defined by molecular markers [3], MPE studies can provide not only risk estimates of incidence, recurrence, or progression tailored to specific subgroups, but also insights into diverse pathogenic pathways [1, 2, 5]. Therefore, findings from MPE studies serve as critical evidence to support the need for personalized management in guiding disease prevention, screening, and treatment [52–59].

A MPE study on obesity/physical activity and colorectal cancer incidence by tumor CTNNB1 (beta-catenin) expression status illustrated a critical role of MPE in enhancing our understanding of host-tumor interactions in the WNT-CTNNB1 signaling pathway [60]. The WNT signaling pathway, when activated, induces the translocation of CTNNB1 into the nucleus, where CTNNB1 promotes the transcription of various genes including growth-promoting genes [61]. The study showed that obesity and low physical activity were associated with a higher risk of CTNNB1-negative colorectal cancer, but not with CTNNB1-positive cancer risk [60]. This finding suggests that colorectal carcinogenesis related to energy imbalance is likely less dependent on the activation of the WNT-CTNNB1 pathway. Consistently, another MPE study found that postdiagnosis physical activity was associated with
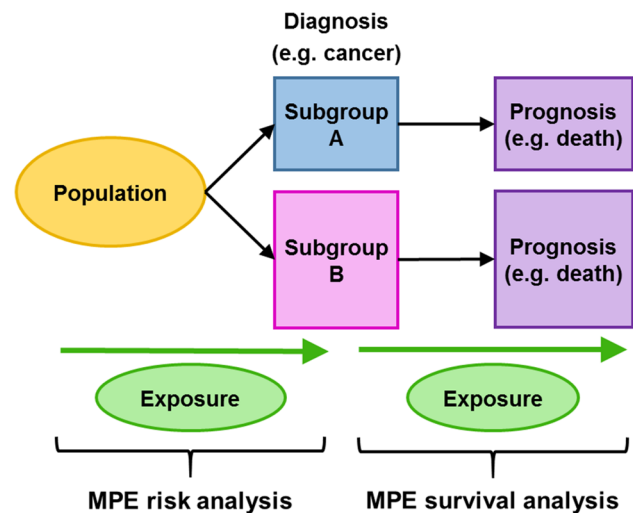

Fig. 2 Risk analysis and survival analysis in molecular pathological epidemiology (MPE) research. MPE risk analysis examines differential associations of a prediagnosis exposure with incidence of disease subtypes defined by molecular features. MPE survival analysis examines differential associations of a prediagnosis or postdiagnosis exposure with prognosis of disease subtypes defined by molecular features. *Arrows* indicate disease process with time. MPE, molecular pathological epidemiology

longer survival only in CTNNB1-negative cancer [62]. In that study, CTNNB1-positive tumors, likely developing due to factors other than energy imbalance, was associated with longer survival only among obese patients [62]. These data imply that the association between energy imbalance-related factors and colorectal cancer survival may vary by activation status of WNT-CTNNB1 signaling. Considering that physical activity and obesity are modifiable lifestyle factors, weight management strategy and exercise program after colorectal cancer diagnosis can be further tailored according to tumor CTNNB1 expression level to effectively improve cancer survival.

Medications have a great potential as chemopreventive agents, and further integration of pharmacoepidemiology into MPE has provided novel insights into mechanistic pathways between common medications and risk of specific colorectal cancer subtypes; this area of investigation has recently been coined "pharmaco-MPE" [3]. Emerging evidence indicates that the pathogenic processes of neoplastic and non-neoplastic diseases are influenced not only by signaling molecules but also by host immune response and microbiota [63]. Thus, disease categorization by factors associated with the immune status and microorganisms has opened new opportunities to examine the disease heterogeneity (immuno-MPE [3] and microbial MPE). In fact, the importance to account for the interplay of tumor molecular features, the gut microbiome, and host factors (e.g., diet, immunity, inflammation) in carcinogenesis process is in line with "the colorectal continuum theory," which proposes that certain molecular features of colorectal cancer may change gradually from the rectum to ascending colon rather than having an abrupt transition at the splenic flexure [64, 65]. In the following sections, we present recent progress and promise of the emerging subfields of MPE (pharmaco-MPE, immuno-MPE, and microbial MPE; Fig. 3). It is the versatile nature (one of the major strengths) of MPE that has enabled the developments of these subfields.

Pharmaco-MPE: integration of pharmacoepidemiology into MPE

Pharmacoepidemiology investigates effects of drugs on disease outcomes and their potential side effects in human populations. Evidence from pharmacoepidemiology research serves as a foundation for chemoprevention and drug therapy. Integration of pharmacoepidemiology and MPE, pharmaco-MPE [3], examines the relationship of a drug with disease incidence or survival according to molecular markers of a disease (Fig. 3). Pharmaco-MPE has particular clinical relevance. No drugs are free from adverse events, and thus it is clinically important to
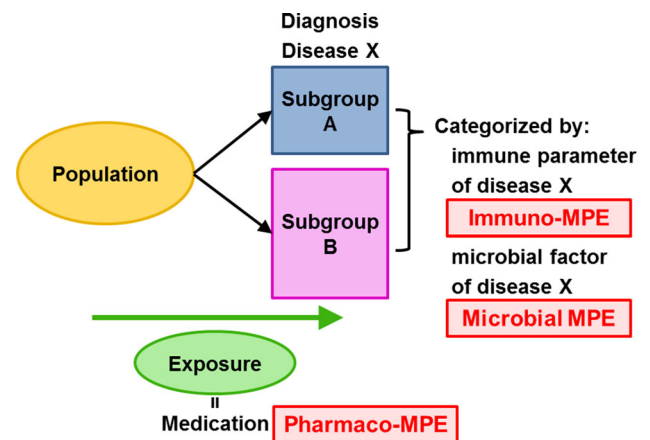

Fig. 3 Further integration of several disciplines into molecular pathological epidemiology (MPE). Pharmaco-MPE integrates pharmacoepidemiology into MPE, where we evaluate differential associations of a medication as an exposure with disease subgroups. Immuno-MPE and microbial MPE integrate MPE with immunology and microbiology, respectively. Diseases are categorized into subtypes by parameters of disease immunity status or microbial profile. *Arrows* indicate disease process with time. MPE, molecular pathological epidemiology

identify target individuals who most likely benefit from use of a particular drug. In fact, pharmaco-MPE has made a striking contribution to not only revealing novel insights into the etiologies and pathogenesis of diseases but also potentially identifying such target populations [52–54, 56, 66–69].

Aspirin, a commonly used nonsteroidal anti-inflammatory drug (NSAID), has been regarded as a promising chemopreventive agent against colorectal cancer incidence and mortality [52–54, 56, 66, 69–75]. In the 2016 recommendation statement, the U.S. Preventive Services Task Force recommends the use of low-dose aspirin for primary prevention of colorectal cancer among adults with a substantial cardiovascular risk [76]. Pharmaco-MPE further refined the inverse relationship between aspirin and colorectal cancer risk, by showing that the association was more evident in tumors with PTGS2 (cyclooxygenase-2) overexpression [52]. This pharmaco-MPE finding suggests that aspirin, as a PTGS (cyclooxygenase) inhibitor, may exert antitumor effects by inhibiting PTGS2 during carcinogenesis. A subsequent study observed that aspirin use was associated with lower incidence of *BRAF*-wild-type colorectal cancer, but not with *BRAF*-mutant cancer risk [56]. This finding led to a hypothesis that *BRAF*-mutated neoplastic cells, by upregulating the MAPK (mitogen-activated protein kinases) pathway, might have resistance to the antitumor effects of aspirin. As illustrated here, the pharmaco-MPE approach has substantially enhanced our understanding of molecular mechanisms underlying the chemopreventive action of aspirin. Another pharmaco-MPE study, by showing that a reduced risk of colorectal

cancer associated with aspirin use was limited to individuals with high expression of HPGD [hydroxyprostaglandin dehydrogenase 15-(NAD), or 15-PGDH], the primary enzyme catabolizing prostaglandins produced by PTGS2 [66], supported potential use of HPGD expression level in normal colorectal mucosa to predict those who would benefit from aspirin chemoprevention.

In survival analyses, pharmaco-MPE provides insights into molecular pathways that modify the effect of a drug on disease progression. Studies have shown that the survival benefit associated with aspirin use may be stronger in cancers with PTGS2 overexpression [53] or *PIK3CA* mutation [54, 69], indicating interactions of aspirin with the prostaglandin or PI3K (phosphatidylinositol-4,5-bisphosphate 3-kinase) signaling pathways in tumor progression. Taken together, the risk and survival analyses incorporating the pharmaco-MPE approach have opened new opportunities to refine regimes for aspirin use in the prevention and treatment of colorectal cancer. It is worth noting that risk factors for cancer incidence are not necessarily consistent with prognostic factors for cancer mortality. Tumor cells continuously interact with the local tumor microenvironment, which consists of extra-cellular matrix, microbiome, and non-neoplastic host cells including inflammatory or immune cells [63]. During tumor progression from earlier to later phases, colonic cells accumulate genomic and epigenomic alterations [63], manifesting different profiles of molecular alterations. It has been shown that neoantigens produced by colorectal cancer cells correlate with T lymphocytic immune response in the tumor microenvironment [77]. Thus, depending on dominant molecular alterations at a stage of carcinogenesis, an interaction between an exposure and tumor molecular markers in the host tumor microenvironment may vary.

Pharmaco-MPE studies have also examined potential heterogeneity in the associations between statin (HMA-CoA reductase inhibitor to lower blood cholesterol level) and colorectal cancer incidence or survival according to tumor molecular subtypes. While statin use was associated with a reduced risk of colorectal cancer with *KRAS* mutation [67], colorectal cancer survival was not related to statin use regardless of *KRAS* mutation status [68].

With growing popularity of high-throughput sequencing, genome-wide association studies (GWAS) can further enrich pharmaco-MPE studies [2, 78–82]. That is, combining knowledge of susceptibility alleles identified by GWAS with that of molecular alterations allows us to examine potential heterogeneity in the drug-disease associations in a more refined manner, providing deeper insights on causality. For instance, in a case–control study, the inverse association between aspirin use and colorectal cancer risk was observed in individuals with the TT

genotype in rs2965667, but not in those with the TA or AA genotypes [79]. Further incorporating molecular markers, the relationship between aspirin use and colorectal cancer risk was examined according to markers jointly defined by rs6983267 genotype and CTNNB1 (beta-catenin) expression [80]. An inverse association between regular aspirin use and colorectal cancer incidence observed among individuals with protective T allele (TT or GT vs. GG genotype) of rs6983267 was further confined to cancer with positive nuclear CTNNB1 expression.

Immuno-MPE: integration of immunology into MPE

Immunology is the study of the immune system and related diseases. Immuno-MPE has been derived by integrating immunology into MPE with the purpose of addressing disease heterogeneity by host immune response (Fig. 3) [3]. Innate and adaptive immunity is a host defense system, and accumulating evidence suggests host immune dysregulation as an underlying etiology for a wide spectrum of human diseases including several types of cancer [63]. In oncology, host immune response to tumors as well as tumor molecular features influences tumor behavior, and serves as an informative biomarker [83–86]. During cancer evolution, cancer cells continuously interact with microenvironment characterized by a complex network across extra-cellular matrix, vascular endothelial cells, and non-neoplastic host cells including immune cells [63]. Therefore, cancer immunology is an interdisciplinary field that requires integrated analyses on host factors, tumor factors, and their interaction [63]. Emerging evidence suggests that activation of immune cells in the tumor microenvironment can be a promising strategy to treat different types of cancer [87–91]. In particular, T cell-mediated immunotherapy has made a breakthrough in cancer treatment by targeting the immune checkpoint pathways related to the PDCD1 (programmed cell death 1, PD-1), CD274 (PDCD1 ligand 1, PD-L1), or CTLA4 proteins [87–91]. As immunotherapy modulates host factors, it is less likely to lead to resistance due to tumor mutations. Along with immunotherapy strategies in cancer treatment, immune modulation can be an attractive strategy for cancer prevention (immunoprevention) [92–94]. A better understanding of host-tumor interactions in the tumor microenvironment would help develop immunoprevention strategies, improve the effectiveness of immunotherapy, and identify patients likely to benefit from immunotherapy and immunoprevention [95, 96].

In immuno-MPE research exploring potential heterogeneity in associations between etiologic factors and disease outcomes by immune parameters, it is of importance

to identify etiologic factors capable of influencing host immune response and to define immune parameters. Accumulating evidence suggests that the immune status may be modulated by a wide variety of epidemiologic factors including diet (e.g., ω-3 polyunsaturated fatty acid [PUFA]), smoking, alcohol, physical activity, obesity, vitamins (e.g., vitamin D), hormones, and common medications (e.g., aspirin, statin) [43, 94, 97–99]. Considering that these factors are readily modifiable, they can be used as immunoprevention strategies. With regard to host immune parameters to sub-classify a particular disease, a single immune cell or a combination of diverse immune cells can be used: e.g., T cells (helper-, memory-, regulatory-, cytotoxic-, or suppressor-T cell), B cell, natural killer cell (NK cell), myeloid-derived suppressor cell, macrophage, neutrophil, eosinophil, and dendritic cell. Additionally, with increasing popularity of immunotherapy in various types of cancer, it is worth investigating other potential markers including the immune checkpoint molecules (e.g., CD274 [PD-L1], PDCD1LG2 [PDCD1 ligand 2, PD-L2], PDCD1 [PD-1], CTLA4 [87–89, 100–103]) and metabolic enzymes (e.g., ARG1, IDO1, TDO2 [104, 105]).

While immuno-MPE is likely to shed light on etiologies and pathogenesis of diseases, there are several challenges. First, epidemiologic studies are limited due to the lack of a large database with comprehensive information on pathological examinations, tumor molecular markers, and immune parameters [63]. Second, while pathological and immunohistochemical examinations of tumor-infiltrating immune cells in tissue sections permit a reliable assessment of host antitumor immune reactivity, pathological methods have not been standardized in terms of specimen types (whole-tissue section or tissue microarray), methods of tissue coring, antibodies for immunohistochemistry, or analytical methods (pathologist's interpretation, or computer-assisted image analysis) [63].

Yet, several immuno-MPE studies have been conducted in relation to colorectal cancer [106–108], identifying potential immunomodulators for cancer immunoprevention [106, 107]. For instance, high plasma 25-hydroxyvitamin D [25(OH)D] level, an indicator of adequate vitamin D status, was associated with a lower risk of colorectal cancer with high-level histopathological immune response, but not with risk of cancer with low-level immune response [106]. The finding supports that antitumor effects of vitamin D may be in part mediated by immune cells that can enzymatically convert 25(OH)D to a bioactive form, 1,25-dihydroxyvitamin D (also known as calcitriol) [106]. Similarly, a reduction in colorectal cancer risk associated with higher marine ω-3 PUFA intake was greater for colorectal cancer with higher FOXP3$^+$ T cell infiltrates [107]. It is speculated that ω-3 PUFA may inhibit regulatory T cell function,

and exert antitumor effects to prevent FOXP3+ T cell-rich cancer [107].

In colorectal cancer, high-level microsatellite instability (MSI) status due to mismatch repair deficiency was characterized by increased neoantigen load, which elicits intense host immune response in the tumor microenvironment [13, 85, 86, 95, 109]. Recently, genomic features of colorectal tumors were shown to be linked to antitumor immunity status. With increasing availability of tumor immunity markers, immune-MPE research is expected to discover new insights into cancer pathogenesis in the context of host-tumor interactions [77].

Microbial MPE: integration of microbiology into MPE

Microbiology is the study of microorganisms, such as bacteria, viruses, archaea, fungi, and protozoa. The colorectum is the most microorganism-rich organ in the human body. To maintain intestinal homeostasis, a complex microflora ecosystem must be under control, and the dysregulation of the intestinal microbial communities may contribute to impaired immunity, chronic inflammation, and carcinogenesis in the colorectum. Indeed, compelling evidence suggests that the gut microbiome is involved in the pathogenesis of various benign and malignant diseases [110–112] including inflammatory bowel diseases and colorectal cancer [113, 114]. Thus, in epidemiologic studies on colorectal cancer, it is important to account for the complex network of the microbiome, intestinal epithelium, and the immune system. Microbial MPE addresses etiologic heterogeneity according to subgroups of colorectal cancer classified by tumor tissue microbial profiling (Fig. 3).

Fusobacterium nucleatum (*F. nucleatum*) has recently gained attention for a potential role in initiating and progressing colorectal cancer [115–118]. Studies have shown that *F. nucleatum* might be associated with molecular features in colorectal adenoma and cancer, including high-level microsatellite instability (MSI) and high-level CpG island methylator phenotype (CIMP) [117–120], as well as with suppression of T cells in the tumor microenvironment [118]. Furthermore, the gut microbiome might also influence the effectiveness of T cell-mediated immunotherapy [121, 122]. Therefore, it is of interest to examine the association between environmental factors (e.g., diet, medications) and colorectal cancer by tumor *F. nucleatum* status. Investigations of viruses and other bacteria, such as *Bifidobacterium*, *Bacteroides*, *Escherichia coli*, and *Campylobacter*, are also warranted in the future. Although technically challenging at this time, comprehensive

assessments of the human microbiome ecosystem along with immune status throughout the body (in relation to disease etiologies and molecular pathologic signatures) will further improve our understanding of disease pathogenesis and evolution.

Future perspectives and conclusions

Based on the unique disease principle [5] and the disease continuum theory [3], MPE has established itself as an evolving research area in epidemiology, enabling us to address potential heterogeneity of the conventional exposure-disease relationship by molecular pathological, immune, or microbial markers of diseases [1–3, 63]. Providing insights into the etiologies and pathogenesis underlying heterogeneous exposure-disease relationships, MPE research serves as a basis for tailored strategies for early detection, prevention, and treatment of diseases [1–3]. Thus, the MPE paradigm is in line with the aim of the NIH Precision Medicine Initiative [31, 32]. The concept and methodology of MPE have been increasingly adopted by cohort studies in various settings [40, 52–58, 60, 62, 69, 123–137].

Currently, biobank/biorepository networks and worldwide collaborative databases are increasingly available for population-based research [8]. In parallel with this trend, there have been great advances in the framework of computational biology, bioinformatics, and genomic medicine. To optimize expanding biomedical data (e.g., genomics, epigenomics, transcriptomics, proteomics, metabolomics, microbiome), it is essential to consider the disease heterogeneity [3]. By categorizing a disease into distinct subgroups based on the unique disease principle, MPE can be a powerful tool to gain novel pathogenic insights and to infer causality from a rich resource of biomedical data.

Integration of several disciplines into MPE has further led to a few evolving subfields within MPE, including pharmaco-MPE, immuno-MPE, and microbial MPE. Further integration across MPE subfields is feasible and expected to become a promising research area in the future. For example, several medications can serve as immunomodulators, altering levels of infiltrating lymphocytes in the tumor microenvironment. Indeed, the integration of pharmaco-MPE and immuno-MPE is one of the greatest achievements in recent MPE research. A recent MPE study examined the association between aspirin use and colorectal cancer incidence by levels of lymphocytic reactions to cancer cells in the tumor microenvironment [138]. Similarly, medications could also influence the gut microbiome through direct antimicrobial effects or alterations in other factors [139, 140]. Therefore, a further integration of pharmaco-MPE and microbial MPE could provide a promising research framework.

Despite all of the abovementioned strengths, there are some challenges in MPE. First, the generation and maintenance of comprehensive tumor molecular databases that are a prerequisite for MPE research demand much effort. As a consequence, MPE analyses limited to individuals with tissue specimens tend to have small sample sizes. To obtain adequate statistical power, large sample sizes of parent cohorts and efforts to obtain as many tissue specimens as possible in the parent cohorts are mandatory [3]. Second, in examining the disease heterogeneity, multiple hypothesis testing is inevitable, which increases false-positive findings. Therefore, it is of particular importance in MPE analyses to form a priori hypotheses based on earlier exploratory findings or on plausible biological mechanisms, and to interpret the results accounting for multiple comparisons [2]. Additionally, unique statistical methods to address the disease heterogeneity have been developed for MPE research [141–147]. Finally, due to the interdisciplinary nature of MPE, one of the most profound challenges in MPE is the paucity of professionals with multidisciplinary expertise across molecular pathology, epidemiology, and biostatistics [148–151]. Multidisciplinary education programs for MPE research in universities and academic institutions could be part of a solution.

In conclusion, the paradigm shift from conventional epidemiology to MPE has opened new opportunities to address the disease heterogeneity, and to provide epidemiologic evidence for molecular pathogenic mechanisms. The MPE research framework is in parallel with the NIH Precision Medicine Initiative, which has emphasized personalized prevention and treatment [31, 32]. Given increasing availability of biomedical data, the disease heterogeneity should be appropriately addressed in order to extract insights into disease etiologies and pathogenesis from invaluable data. The evolving field of MPE can be a core field in the era of big-data health science and precision medicine.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflicts of interest.

References

1. Ogino S, Stampfer M. Lifestyle factors and microsatellite instability in colorectal cancer: the evolving field of molecular pathological epidemiology. J Natl Cancer Inst. 2010;102:365–7.

2. Ogino S, Chan AT, Fuchs CS, et al. Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. Gut. 2011;60:397–411.

3. Ogino S, Nishihara R, VanderWeele TJ, et al. Review article: the role of molecular pathological epidemiology in the study of neoplastic and non-neoplastic diseases in the era of precision medicine. Epidemiology. 2016;27:602–11.

4. Nishihara R, VanderWeele TJ, Shibuya K, et al. Molecular pathological epidemiology gives clues to paradoxical findings. Eur J Epidemiol. 2015;30:1129–35.

5. Ogino S, Lochhead P, Chan AT, et al. Molecular pathological epidemiology of epigenetics: emerging integrative science to analyze environment, host, and disease. Mod Pathol. 2013;26:465–84.

6. Ogino S, Fuchs CS, Giovannucci E. How many molecular subtypes? Implications of the unique tumor principle in personalized medicine. Expert Rev Mol Diagn. 2012;12:621–8.

7. Ogino S, Giovannucci E. Commentary: lifestyle factors and colorectal cancer microsatellite instability–molecular pathological epidemiology science, based on unique tumour principle. Int J Epidemiol. 2012;41:1072–4.

8. Ogino S, Lochhead P, Giovannucci E, et al. Discovery of colorectal cancer PIK3CA mutation as potential predictive biomarker: power and promise of molecular pathological epidemiology. Oncogene. 2014;33:2949–55.

9. Sugimura H. Susceptibility to human cancer: from the perspective of a pathologist. Pathol Int. 2016;66:359–68.

10. Chen IC, Lee KH, Hsu YH, et al. Expression pattern and clinicopathological relevance of the indoleamine 2,3-dioxygenase 1/tryptophan 2,3-dioxygenase protein in colorectal cancer. Dis Markers. 2016;2016:8169724.

11. Jiang MJ, Dai JJ, Gu DN, et al. Aspirin in pancreatic cancer: chemopreventive effects and therapeutic potentials. Biochim Biophys Acta. 2016;1866:163–76.

12. Li YY, Ge QX, Cao J, et al. Association of Fusobacterium nucleatum infection with colorectal cancer in Chinese patients. World J Gastroenterol. 2016;22:3227–33.

13. Li SK, Martin A. Mismatch repair and colon cancer: mechanisms and therapies explored. Trends Mol Med. 2016;22:274–89.

14. Zhou L, Wang K, Li Q, et al. Clinical proteomics-driven precision medicine for targeted cancer therapy: current overview and future perspectives. Expert Rev Proteomics. 2016;13:367–81.

15. Aran V, Victorino AP, Thuler LC, et al. Colorectal cancer: epidemiology, disease mechanisms and interventions to reduce onset and mortality. Clin Colorectal Cancer. 2016;15:195–203.

16. Wang S, Wu S, Meng Q, et al. FAS rs2234767 and rs1800682 polymorphisms jointly contributed to risk of colorectal cancer by affecting SP1/STAT1 complex recruitment to chromatin. Sci Rep. 2016;6:19229.

17. Kuroiwa-Trzmielina J, Wang F, Rapkins RW, et al. 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. 2016.

18. Ohno T, Adachi S, Okuno M, et al. Development of a novel scoring system for predicting the risk of colorectal neoplasia: a retrospective study. PLoS ONE. 2016;11:e0157269.

19. Szylberg L, Janiczek M, Popiel A, et al. Serrated polyps and their alternative pathway to the colorectal cancer: a systematic review. Gastroenterol Res Pract. 2015;2015:573814.

20. Cebola I, Custodio J, Munoz M, et al. Epigenetics override pro-inflammatory PTGS transcriptomic signature towards selective hyperactivation of PGE2 in colorectal cancer. Clin Epigenetics. 2015;7:74.

21. Li W, Qiu T, Ling Y, et al. Molecular pathological epidemiology of colorectal cancer in Chinese patients with KRAS and BRAF mutations. Oncotarget. 2015;6:39607–13.

22. Tillmans LS, Vierkant RA, Wang AH, et al. Associations between environmental exposures and incident colorectal cancer by ESR2 protein expression level in a population-based cohort of older women. Cancer Epidemiol Biomarkers Prev. 2015;24:713–9.

23. Mehta AM, Osse M, Kolkman-Uljee S, et al. Molecular backgrounds of ERAP1 downregulation in cervical carcinoma. Anal Cell Pathol (Amst). 2015;2015:367837.

24. Bishehsari F, Mahdavinia M, Vacca M, et al. Epidemiological transition of colorectal cancer in developing countries: environmental factors, molecular pathways, and opportunities for prevention. World J Gastroenterol. 2014;20:6055–72.

25. Campbell PT, Deka A, Briggs P, et al. Establishment of the cancer prevention study II nutrition cohort colorectal tissue repository. Cancer Epidemiol Biomarkers Prev. 2014;23:2694–702.

26. Hagland HR, Berg M, Jolma IW, et al. Molecular pathways and cellular metabolism in colorectal cancer. Dig Surg. 2013;30:12–25.

27. Buchanan DD, Win AK, Walsh MD, et al. Family history of colorectal cancer in BRAF p. V600E-mutated colorectal cancer cases. Cancer Epidemiol Biomarkers Prev. 2013;22:917–26.

28. Chia WK, Ali R, Toh HC. Aspirin as adjuvant therapy for colorectal cancer–reinterpreting paradigms. Nat Rev Clin Oncol. 2012;9:561–70.

29. Hughes LA, Khalid-de Bakker CA, Smits KM, et al. The CpG island methylator phenotype in colorectal cancer: progress and problems. Biochim Biophys Acta. 2012;1825:77–85.

30. Curtin K, Slattery ML, Samowitz WS. CpG island methylation in colorectal cancer: past, present and future. Patholog Res Int. 2011;2011:902674.

31. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med. 2015;372:793–5.

32. Lyman GH, Moses HL. Biomarker tests for molecularly targeted therapies-the key to unlocking precision medicine. N Engl J Med. 2016;375:4–6.

33. Kuller LH, Bracken MB, Ogino S, et al. The role of epidemiology in the era of molecular epidemiology and genomics: summary of the 2013 AJE-sponsored Society of Epidemiologic Research Symposium. Am J Epidemiol. 2013;178:1350–4.

34. Epplein M, Bostick RM, Mu L, et al. Challenges and opportunities in international molecular cancer prevention research: an ASPO molecular epidemiology and the environment and international cancer prevention interest groups report. Cancer Epidemiol Biomarkers Prev. 2014;23:2613–7.

35. Ogino S. Molecular pathological epidemiology (MPE): overview of its paradigm and wide applicability even without tumor tissue [abstract]. In: Proceedings of the twelfth annual AACR international conference on frontiers in cancer prevention research; 2013 Oct 27–30. National Harbor, MD: Cancer Prev Res (Phila); 2013;6:CN06-01.

36. Ogino S, Campbell PT, Nishihara R, et al. Proceedings of the second international molecular pathological epidemiology (MPE) meeting. Cancer Causes Control. 2015;26:959–72.

37. Nishi A, Milner DA Jr, Giovannucci EL, et al. Integration of molecular pathology, epidemiology and social science for global precision medicine. Expert Rev Mol Diagn. 2016;16:11–23.

38. Nishi A, Kawachi I, Koenen KC, et al. Lifecourse epidemiology and molecular pathological epidemiology. Am J Prev Med. 2015;48:116–9.

39. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61:759–67.
40. Phipps AI, Limburg PJ, Baron JA, et al. Association between molecular subtypes of colorectal cancer and patient survival. Gastroenterology. 2015;148(77–87):e2.
41. Kuipers EJ, Grady WM, Lieberman D, et al. Colorectal cancer. Nat Rev Dis Primers. 2015;1:15065.
42. Martinez-Useros J, Garcia-Foncillas J. Obesity and colorectal cancer: molecular features of adipose tissue. J Transl Med. 2016;14:21.
43. Song M, Garrett WS, Chan AT. Nutrients, foods, and colorectal cancer prevention. Gastroenterology. 2015;148(1244–60):e16.
44. Lee DH, Keum N, Giovannucci EL. Colorectal cancer epidemiology in the Nurses' Health Study. Am J Public Health. 2016;106:1599–607.
45. Strum WB. Colorectal adenomas. N Engl J Med. 2016;374:1065–75.
46. Aran V, Victorino AP, Thuler LC, et al. Colorectal Cancer: Epidemiology, Disease Mechanisms and Interventions to Reduce Onset and Mortality. Clin Colorectal Cancer. 2016.
47. Lochhead P, Chan AT, Giovannucci E, et al. Progress and opportunities in molecular pathological epidemiology of colorectal premalignant lesions. Am J Gastroenterol. 2014;109:1205–14.
48. Field AE, Camargo CA Jr, Ogino S. The merits of subtyping obesity: one size does not fit all. JAMA. 2013;310:2147–8.
49. Ikramuddin S, Livingston EH. New insights on bariatric surgery outcomes. JAMA. 2013;310:2401–2.
50. Reddon H, Gueant JL, Meyre D. The importance of gene-environment interactions in human obesity. Clin Sci (Lond). 2016;130:1571–97.
51. Bam M, Yang X, Zumbrun EE, et al. Dysregulated immune system networks in war veterans with PTSD is an outcome of altered miRNA expression and DNA methylation. Sci Rep. 2016;6:31209.
52. Chan AT, Ogino S, Fuchs CS. Aspirin and the risk of colorectal cancer in relation to the expression of COX-2. N Engl J Med. 2007;356:2131–42.
53. Chan AT, Ogino S, Fuchs CS. Aspirin use and survival after diagnosis of colorectal cancer. JAMA. 2009;302:649–58.
54. Liao X, Lochhead P, Nishihara R, et al. Aspirin use, tumor PIK3CA mutation, and colorectal-cancer survival. N Engl J Med. 2012;367:1596–606.
55. Yamauchi M, Lochhead P, Imamura Y, et al. Physical activity, tumor PTGS2 expression, and survival in patients with colorectal cancer. Cancer Epidemiol Biomarkers Prev. 2013;22:1142–52.
56. Nishihara R, Lochhead P, Kuchiba A, et al. Aspirin use and risk of colorectal cancer according to BRAF mutation status. JAMA. 2013;309:2563–71.
57. Nishihara R, Morikawa T, Kuchiba A, et al. A prospective study of duration of smoking cessation and colorectal cancer risk by epigenetics-related tumor classification. Am J Epidemiol. 2013;178:84–100.
58. Hanyuda A, Kim SA, Martinez-Fernandez A, et al. Survival benefit of exercise differs by tumor IRS1 expression status in colorectal cancer. Ann Surg Oncol. 2016;23:908–17.
59. Frampton M, Houlston RS. Modeling the prevention of colorectal cancer from the combined impact of host and behavioral risk factors. Genet Med. 2016.
60. Morikawa T, Kuchiba A, Lochhead P, et al. Prospective analysis of body mass index, physical activity, and colorectal cancer risk associated with beta-catenin (CTNNB1) status. Cancer Res. 2013;73:1600–10.
61. Korinek V, Barker N, Morin PJ, et al. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC$^{-/-}$ colon carcinoma. Science. 1997;275:1784–7.
62. Morikawa T, Kuchiba A, Yamauchi M, et al. Association of CTNNB1 (beta-catenin) alterations, body mass index, and physical activity with survival in patients with colorectal cancer. JAMA. 2011;305:1685–94.
63. Ogino S, Galon J, Fuchs CS, et al. Cancer immunology–analysis of host and tumor factors for personalized medicine. Nat Rev Clin Oncol. 2011;8:711–9.
64. Yamauchi M, Morikawa T, Kuchiba A, et al. Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. Gut. 2012;61:847–54.
65. Yamauchi M, Lochhead P, Morikawa T, et al. Colorectal cancer: a tale of two sides or a continuum? Gut. 2012;61:794–7.
66. Fink SP, Yamauchi M, Nishihara R, et al. Aspirin and the risk of colorectal cancer in relation to the expression of 15-hydroxyprostaglandin dehydrogenase (HPGD). Sci Transl Med. 2014;6:233re2.
67. Lee JE, Baba Y, Ng K, et al. Statin use and colorectal cancer risk according to molecular subtypes in two large prospective cohort studies. Cancer Prev Res (Phila). 2011;4:1808–15.
68. Ng K, Ogino S, Meyerhardt JA, et al. Relationship between statin use and colon cancer recurrence and survival: results from CALGB 89803. J Natl Cancer Inst. 2011;103:1540–51.
69. Domingo E, Church DN, Sieber O, et al. Evaluation of PIK3CA mutation as a predictor of benefit from nonsteroidal anti-inflammatory drug therapy in colorectal cancer. J Clin Oncol. 2013;31:4297–305.
70. Cao Y, Nishihara R, Wu K, et al. Population-wide impact of long-term use of aspirin and the risk for cancer. JAMA Oncol. 2016;2:762–9.
71. Drew DA, Cao Y, Chan AT. Aspirin and colorectal cancer: the promise of precision chemoprevention. Nat Rev Cancer. 2016;16:173–86.
72. Rothwell PM, Wilson M, Elwin CE, et al. Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. Lancet. 2010;376:1741–50.
73. Rothwell PM, Wilson M, Price JF, et al. Effect of daily aspirin on risk of cancer metastasis: a study of incident cancers during randomised controlled trials. Lancet. 2012;379:1591–601.
74. Li P, Wu H, Zhang H, et al. Aspirin use after diagnosis but not prediagnosis improves established colorectal cancer survival: a meta-analysis. Gut. 2015;64:1419–25.
75. Friis S, Riis AH, Erichsen R, et al. Low-dose aspirin or nonsteroidal anti-inflammatory drug use and colorectal cancer risk: a population-based. case-control study. Ann Intern Med. 2015;163:347–55.
76. Bibbins-Domingo K. Aspirin use for the primary prevention of cardiovascular disease and colorectal cancer: U.S. preventive services task force recommendation statement. Ann Intern Med. 2016;164:836–45.
77. Giannakis M, Mu XJ, Shukla SA, et al. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. Cell Rep. 2016.
78. Garcia-Albeniz X, Nan H, Valeri L, et al. Phenotypic and tumor molecular characterization of colorectal cancer in relation to a susceptibility SMAD7 variant associated with survival. Carcinogenesis. 2013;34:292–8.
79. Nan H, Hutter CM, Lin Y, et al. Association of aspirin and NSAID use with risk of colorectal cancer according to genetic variants. JAMA. 2015;313:1133–42.
80. Nan H, Morikawa T, Suuriniemi M, et al. Aspirin use, 8q24 single nucleotide polymorphism rs6983267, and colorectal cancer according to CTNNB1 alterations. J Natl Cancer Inst. 2013;105:1852–61.
81. Khalili H, Gong J, Brenner H, et al. Identification of a common variant with potential pleiotropic effect on risk of inflammatory

bowel disease and colorectal cancer. Carcinogenesis. 2015;36:999–1007.

82. Garcia-Albeniz X, Rudolph A, Hutter C, et al. CYP24A1 variant modifies the association between use of oestrogen plus progestogen therapy and colorectal cancer risk. Br J Cancer. 2016;114:221–9.

83. Pages F, Berger A, Camus M, et al. Effector memory T cells, early metastasis, and survival in colorectal cancer. N Engl J Med. 2005;353:2654–66.

84. Galon J, Costes A, Sanchez-Cabo F, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science. 2006;313:1960–4.

85. Ogino S, Nosho K, Irahara N, et al. Lymphocytic reaction to colorectal cancer is associated with longer survival, independent of lymph node count, microsatellite instability, and CpG island methylator phenotype. Clin Cancer Res. 2009;15:6412–20.

86. Nosho K, Baba Y, Tanaka N, et al. Tumour-infiltrating T-cell subsets, molecular changes in colorectal cancer, and prognosis: cohort study and literature review. J Pathol. 2010;222:350–66.

87. Tumeh PC, Harview CL, Yearley JH, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature. 2014;515:568–71.

88. Postow MA, Callahan MK, Wolchok JD. Immune checkpoint blockade in cancer therapy. J Clin Oncol. 2015;33:1974–82.

89. Baumeister SH, Freeman GJ, Dranoff G, et al. Coinhibitory pathways in immunotherapy for cancer. Annu Rev Immunol. 2016;34:539–73.

90. Topalian SL, Drake CG, Pardoll DM. Immune checkpoint blockade: a common denominator approach to cancer therapy. Cancer Cell. 2015;27:450–61.

91. Le DT, Uram JN, Wang H, et al. PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med. 2015;372:2509–20.

92. Wojtowicz ME, Dunn BK, Umar A. Immunologic approaches to cancer prevention-current status, challenges, and future perspectives. Semin Oncol. 2016;43:161–72.

93. Finn OJ, Beatty PL. Cancer immunoprevention. Curr Opin Immunol. 2016;39:52–8.

94. Janakiram NB, Mohammed A, Madka V, et al. Prevention and treatment of cancers by immune modulating nutrients. Mol Nutr Food Res. 2016;60:1275–94.

95. Llosa NJ, Cruise M, Tam A, et al. The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. Cancer Discov. 2015;5:43–51.

96. Gentles AJ, Newman AM, Liu CL, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med. 2015;21:938–45.

97. Shiels MS, Katki HA, Freedman ND, et al. Cigarette smoking and variations in systemic immune and inflammation markers. J Natl Cancer Inst. 2014;106:dju294.

98. Zelenay S, van der Veen AG, Bottcher JP, et al. Cyclooxygenase-dependent tumor growth through evasion of immunity. Cell. 2015;162:1257–70.

99. Hussain M, Javeed A, Ashraf M, et al. Aspirin and immune system. Int Immunopharmacol. 2012;12:10–20.

100. Masugi Y, Nishihara R, Yang J, et al. Tumour CD274 (PD-L1) expression and T cells in colorectal cancer. Gut. 2016.

101. Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. N Engl J Med. 2012;366:2443–54.

102. Brahmer JR, Tykodi SS, Chow LQ, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. N Engl J Med. 2012;366:2455–65.

103. Robert C, Thomas L, Bondarenko I, et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. N Engl J Med. 2011;364:2517–26.

104. Muller AJ, DuHadaway JB, Donover PS, et al. Inhibition of indoleamine 2,3-dioxygenase, an immunoregulatory target of the cancer suppression gene Bin1, potentiates cancer chemotherapy. Nat Med. 2005;11:312–9.

105. Antonioli L, Blandizzi C, Pacher P, et al. Immunity, inflammation and cancer: a leading role for adenosine. Nat Rev Cancer. 2013;13:842–57.

106. Song M, Nishihara R, Wang M, et al. Plasma 25-hydroxyvitamin D and colorectal cancer risk according to tumour immunity status. Gut. 2016;65:296–304.

107. Song M, Nishihara R, Cao Y, et al. Marine omega-3 Polyunsaturated Fatty Acid Intake and Risk of Colorectal Cancer Characterized by Tumor-Infiltrating T Cells. JAMA Oncol. 2016.

108. Hanyuda A, Ogino S, Qian ZR, et al. Body mass index and risk of colorectal cancer according to tumor lymphocytic infiltrate. Int J Cancer. 2016;139:854–68.

109. Dudley JC, Lin MT, Le DT, et al. Microsatellite instability as a biomarker for PD-1 blockade. Clin Cancer Res. 2016;22:813–20.

110. Lozupone CA, Stombaugh JI, Gordon JI, et al. Diversity, stability and resilience of the human gut microbiota. Nature. 2012;489:220–30.

111. Amirian ES, Petrosino JF, Ajami NJ, et al. Potential role of gastrointestinal microbiota composition in prostate cancer risk. Infect Agent Cancer. 2013;8:42.

112. Claassen-Weitz S, Wiysonge CS, Machingaidze S, et al. Current knowledge and future research directions on fecal bacterial patterns and their association with asthma. Front Microbiol. 2016;7:838.

113. Ahn J, Sinha R, Pei Z, et al. Human gut microbiome and risk for colorectal cancer. J Natl Cancer Inst. 2013;105:1907–11.

114. Flemer B, Lynch DB, Brown JM, et al. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. Gut. 2016.

115. Kostic AD, Gevers D, Pedamallu CS, et al. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res. 2012;22:292–8.

116. Castellarin M, Warren RL, Freeman JD, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. 2012;22:299–306.

117. Mima K, Nishihara R, Qian ZR, et al. Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. Gut. 2015.

118. Mima K, Sukawa Y, Nishihara R, et al. Fusobacterium nucleatum and T Cells in colorectal carcinoma. JAMA Oncol. 2015;1:653–61.

119. Tahara T, Yamamoto E, Suzuki H, et al. Fusobacterium in colonic flora and molecular features of colorectal carcinoma. Cancer Res. 2014;74:1311–8.

120. Ito M, Kanno S, Nosho K, et al. Association of Fusobacterium nucleatum with clinical and molecular features in colorectal serrated pathway. Int J Cancer. 2015;137:1258–68.

121. Vetizou M, Pitt JM, Daillere R, et al. Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. Science. 2015;350:1079–84.

122. Sivan A, Corrales L, Hubert N, et al. Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. Science. 2015;350:1084–9.

123. Desai NB, Scott SN, Zabor EC, et al. Genomic characterization of response to chemoradiation in urothelial bladder cancer. Cancer. 2016.

124. Dixon SC, Nagle CM, Thrift AP, et al. Adult body mass index and risk of ovarian cancer by subtype: a Mendelian randomization study. Int J Epidemiol. 2016.

125. Mima K, Nishihara R, Yang J, et al. MicroRNA MIR21 (miR-21) and PTGS2 expression in colorectal cancer and patient survival. Clin Cancer Res. 2016;22:3841–8.

126. Sisti JS, Collins LC, Beck AH, et al. Reproductive risk factors in relation to molecular subtypes of breast cancer: results from the nurses' health studies. Int J Cancer. 2016;138:2346–56.

127. Mima K, Nowak JA, Qian ZR, et al. Tumor LINE-1 methylation level and colorectal cancer location in relation to patient survival. Oncotarget. 2016.

128. Hirko KA, Willett WC, Hankinson SE, et al. Healthy dietary patterns and risk of breast cancer by molecular subtype. Breast Cancer Res Treat. 2016;155:579–88.

129. Simons CC, Schouten LJ, Godschalk R, et al. Body size, physical activity, genetic variants in the insulin-like growth factor pathway and colorectal cancer risk. Carcinogenesis. 2015;36:971–81.

130. Campbell PT, Newton CC, Newcomb PA, et al. Association between body mass index and mortality for colorectal cancer survivors: overall and by tumor molecular phenotype. Cancer Epidemiol Biomarkers Prev. 2015;24:1229–38.

131. Hardikar S, Newcomb PA, Campbell PT, et al. Prediagnostic physical activity and colorectal cancer survival: overall and stratified by tumor characteristics. Cancer Epidemiol Biomarkers Prev. 2015;24:1130–7.

132. Simons CC, van den Brandt PA, Stehouwer CD, et al. 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. 2014;23:1852–62.

133. Zhu Y, Yang SR, Wang PP, et al. Influence of pre-diagnostic cigarette smoking on colorectal cancer survival: overall and by tumour molecular phenotype. Br J Cancer. 2014;110:1359–66.

134. Brandstedt J, Wangefjord S, Nodin B, et al. Associations of hormone replacement therapy and oral contraceptives with risk of colorectal cancer defined by clinicopathological factors, beta-catenin alterations, expression of cyclin D1, p53, and microsatellite-instability. BMC Cancer. 2014;14:371.

135. Nishihara R, Wu K, Lochhead P, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. N Engl J Med. 2013;369:1095–105.

136. Bae JM, Kim JH, Cho NY, et al. Prognostic implication of the CpG island methylator phenotype in colorectal cancers depends on tumour location. Br J Cancer. 2013;109:1004–12.

137. Hughes LA, Williamson EJ, van Engeland M, et al. Body size and risk for colorectal cancers showing BRAF mutations or microsatellite instability: a pooled analysis. Int J Epidemiol. 2012;41:1060–72.

138. Cao Y, Nishihara R, Qian ZR, et al. Regular Aspirin Use Associates with Lower Risk of Colorectal Cancers With Low Numbers of Tumor-infiltrating Lymphocytes. Gastroenterology. 2016.

139. Blaser MJ. Antibiotic use and its consequences for the normal microbiome. Science. 2016;352:544–5.

140. Rogers MA, Aronoff DM. The influence of non-steroidal anti-inflammatory drugs on the gut microbiome. Clin Microbiol Infect. 2016;22(178):e1–9.

141. Rosner B, Glynn RJ, Tamimi RM, et al. Breast cancer risk prediction with heterogeneous risk profiles according to breast cancer tumor markers. Am J Epidemiol. 2013;178:296–308.

142. Begg CB, Orlow I, Zabor EC, et al. Identifying etiologically distinct sub-types of cancer: a demonstration project involving breast cancer. Cancer Med. 2015;4:1432–9.

143. Wang M, Kuchiba A, Ogino S. A meta-regression method for studying etiological heterogeneity across disease subtypes classified by multiple biomarkers. Am J Epidemiol. 2015;182:263–70.

144. Wang M, Spiegelman D, Kuchiba A, et al. Statistical methods for studying disease subtype heterogeneity. Stat Med. 2016;35:782–800.

145. Mauguen A, Begg CB. Using the Lorenz curve to characterize risk predictiveness and etiologic heterogeneity. Epidemiology. 2016;27:531–7.

146. Begg CB, Seshan VE, Zabor EC, et al. Genomic investigation of etiologic heterogeneity: methodologic challenges. BMC Med Res Methodol. 2014;14:138.

147. Chatterjee N, Sinha S, Diver WR, et al. Analysis of cohort studies with multivariate and partially observed disease classification data. Biometrika. 2010;97:683–98.

148. Ogino S, King EE, Beck AH, et al. Interdisciplinary education to integrate pathology and epidemiology: towards molecular and population-level health science. Am J Epidemiol. 2012;176:659–67.

149. Kuller LH. Invited commentary: the 21st century epidemiologist—a need for different training? Am J Epidemiol. 2012;176:668–71.

150. Ogino S, Beck AH, King EE, et al. Ogino et Al. Respond to "the 21st century epidemiologist". Am J Epidemiol. 2012;176:672–4.

151. Ogino S, Nishihara R. All biomedical and health science researchers, including laboratory physicians and scientists, need adequate education and training in study design and statistics. Clin Chem. 2016;62:1039–40.