Published in final edited form as:

J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. 2014 ; 32(2): 121–158. doi: 10.1080/10590501.2014.907460.

Toxicogenomics and Cancer Susceptibility: Advances with Next-Generation Sequencing

Baitang Ning[1], **Zhenqiang Su**[1], **Nan Mei**[1], **Huixiao Hong**[1], **Helen Deng**[2], **Leming Shi**[1,3], **James C. Fuscoe**[1], and **William H. Tolleson**[1]

[1]National Center for Toxicological Research, Food and Drug Administration, Jefferson, Arkansas, USA

[2]Arkansas Department of Health and Human Service, Little Rock, Arkansas, USA

[3]Center for Pharmacogenomics, School of Pharmacy, Fudan University, Pudong District, Shanghai, China

Abstract

The aim of this review is to comprehensively summarize the recent achievements in the field of toxicogenomics and cancer research regarding genetic-environmental interactions in carcinogenesis and detection of genetic aberrations in cancer genomes by next-generation sequencing technology. Cancer is primarily a genetic disease in which genetic factors and environmental stimuli interact to cause genetic and epigenetic aberrations in human cells. Mutations in the germline act as either high-penetrance alleles that strongly increase the risk of cancer development, or as low-penetrance alleles that mildly change an individual's susceptibility to cancer. Somatic mutations, resulting from either DNA damage induced by exposure to environmental mutagens or from spontaneous errors in DNA replication or repair are involved in the development or progression of the cancer. Induced or spontaneous changes in the epigenome may also drive carcinogenesis. Advances in next-generation sequencing technology provide us opportunities to accurately, economically, and rapidly identify genetic variants, somatic mutations, gene expression profiles, and epigenetic alterations with single-base resolution. Whole genome sequencing, whole exome sequencing, and RNA sequencing of paired cancer and adjacent normal tissue present a comprehensive picture of the cancer genome. These new findings should benefit public health by providing insights in understanding cancer biology, and in improving cancer diagnosis and therapy.

Keywords

Cancer; carcinogenesis; genomics; toxicogenomics; environmental exposure; next generation sequencing

Address correspondence to Baitang Ning, Ph.D., National Center for Toxicological Research, 3900 NCTR Road, HFT100, Jefferson, AR 72079, USA. baitang.ning@fda.hhs.gov.

INTRODUCTION

All cancers result from genetic and epigenetic aberrations [1]. Genetic aberrations refer to inherited germline mutations and induced somatic mutations, which include single base mutations, indels (insertions and deletions), gene copy number variations, and gene/chromosomal rearrangements [2]. Some inherited germline mutations predispose individuals to particular types of cancer. These inherited variants can be categorized as either (i) rare, high-penetrance alleles and dominant mutations that strongly increase the risk of cancer development or (ii) common and low-penetrance alleles that mildly change an individuals' susceptibility to cancer [3]. In contrast to germline mutations, which are transmitted passively through inheritance, somatic mutations are acquired among the cells of the body sporadically over the lifetime of the individual. Somatic mutations arise from DNA damage caused by exposure to endogenous or exogenous mutagens or resulting from spontaneous errors in DNA replication or repair. Some of these somatic mutations, termed driver mutations, are involved in the development or progression of the cancer; driver mutations confer growth advantages to cancer cells. Other somatic mutations, referred to as passenger mutations, happen to be present in cancer cells but do not contribute to the cancer phenotype [2]. Epigenetics has been defined as "heritable changes in gene expression that are not due to any alteration in the DNA sequence" [4]. Such epigenetic changes include his-tone modifications, DNA methylation pattern changes, and alterations of non-coding RNA (ncRNA) expression. Examples of epigenetic alterations in cancer include: (1) progressive hypomethylation of total DNA in tumor cells followed by hypermethylation of CpG-islands of tumor suppressor genes [4]; (2) alterations in the patterns of ncRNA expression that play important roles in the regulation of genes related to cell growth, cell proliferation, cell differentiation and cell death [5]; (3) global down-regulation of the expression of microRNA in comparison to counterpart normal tissues, with the ability to classify some poorly differentiated tumors by microRNA signatures [6]; and (4) histone modifications that alter chromatin structure and influence the expression of important tumor suppressor genes [7]. Genetic and epigenetic alterations in the expression of genes that regulate cell growth, cell division, cell differentiation, cell apoptosis, and other biological functions determine the cancer phenotype. In fact, it has been reported that more than 500 gene transcripts were differentially expressed at statistically significant levels between normal cells and cancer cells [8].

Advances in biotechnologies will now allow the detection of more mutations, epigenetic alterations, and gene expression changes in cancer cells. The accurate, economical, and rapid identification of mutations, gene expression profiles, and epigenetic alterations in cancer genome has led to insights in understanding cancer biology and to improved cancer diagnosis and therapy. Over the past several decades germline mutation detection approaches have dramatically improved from low throughput, low resolution methods like restriction fragment length polymorphism (RFLP) analyses [9] to next-generation sequencing (NGS) technology, which provides very high resolution genetic information relevant to carcinogenesis [10]. The power of molecular epidemiological studies of different types of cancer has evolved from single marker association analysis using RFLP to multiple marker association analysis using TaqMan assays [11], and to genome-wide association studies

(GWAS) using microarray technologies [12]. Coincident with the announcement in November, 2013 that the Illumina MiSeqDx was the first NGS platform to receive marketing authorization by the FDA, Collins and Hamburg described the vast potential of high-throughput sequencing technology to revolutionize biomedical science and clinical medicine [13]. Currently, NGS methods can provide a much higher level of molecular detail with a relatively low-cost, enabling the comprehensive analysis of human and cancer genomes [14]. Whole genome sequencing of paired cancer and adjacent normal tissue presents a comprehensive picture of the cancer genome. Furthermore, NGS provides the opportunity to study the role of mutational signatures among cancers that can be associated with probable etiologies, such as exposure to genotoxic agents or defects in DNA repair [15]. Comparing somatic and germline mutations at base-pair resolution reveals precisely how the tumor genome is different from the normal genome and sheds light on the mechanisms of carcinogenesis [16]. Several international consortia, such as The Cancer Genome Atlas, the Cancer Genome Project (at the Wellcome Trust Sanger Institute), the International Cancer Genome Consortium, and Catalogue of Somatic Mutations in Cancer (COSMIC) have made tremendous efforts to identify cancer markers and mutations. Mutations, translocations, and potential therapeutic targets have been identified in numerous cancer subtypes by these consortia using NGS technology [17–19]. Microarray technology has made it possible to profile the expression of hundreds to thousands of genes in tumor tissues, allowing classification of cancers (such as breast cancer) into clinic subtypes, and predicting cancer recurrence and response to different treatment protocols [20]. NGS provides much higher resolution and increased depth of cancer gene expression by offering details of transcriptional boundaries, differential expressions, mutations, rare transcripts, and aberrations of alternative splicing [21]. Additional molecular changes associated with cancer can be detected using NGS. Castle and colleagues [22] utilized NGS methods to determine copy number differences for nuclear, mitochondrial, and telomeric DNA sequences with high accuracy. Likewise, NGS is a powerful tool for detection of epigenetic changes that, in turn, alter gene expression. For example, chromatin immunoprecipitation combined with NGS (ChIP-Seq) was used to define epigenome maps, which revealed underlying mechanisms of acquired drug resistance in breast cancer during endocrine therapy [23].

PERSONALIZED ONCOLOGY AND NGS—COLORECTAL CANCER AS AN EXAMPLE

A great wealth of information regarding the biology and treatment of colorectal cancer has been gathered through decades of intensive research. Information gathered from molecular epidemiological studies of colorectal cancer offer valuable perspectives on the need for NGS technologies to enhance future toxicogenomic studies and it also provides rationales for new applications of NGS-based methods in personalized medicine to treat this disease. The capabilities provided by NGS technology to gather and evaluate vast amounts of information on individual patient specimens at the molecular level provides an unprecedented opportunity to leverage what has been learned through past research to increase our understanding of gene-environment interactions in human colorectal cancer and to improve patient care.

Colorectal cancer affects over one million people globally and 5%–10% of these cases can be associated with the inheritance of high penetrance cancer susceptibility alleles that are typically transmitted as autosomal dominant traits [24]. Genes associated with hereditary nonpolyposis colorectal cancer/Lynch syndrome (*MLH1, MSH2, MSH6, PMS2* DNA mismatch repair genes) [25], familial adenomatous polyposis (*APC*) [26], and hamartomatous polyposis/Peutz-Jeghers syndrome (*STK11*) [27] are each high penetrance genes associated with colorectal cancer. Variants of DNA polymerase-delta (*POLD*) and axis inhibitor 2 (*AXIN2*) genes transmitted through the germline are also associated with colorectal cancer with high penetrance, but these are more rare in the population [28]. Particular inherited variant alleles for APC (*APC* I307K*), transforming growth factor beta receptor type 1 (*TGFBR1* 6Ala*), methylene tetrahydrofolate reductase (*MTHFR* 677V*), and H-Ras (*HRAS1* VNTR*) are examples of alleles that are much more common in the population (5%–7%, 14%, 32%–43%, 1%–6% reported frequencies, respectively) but exhibit lower penetrance for colorectal cancer [relative risk (95% CI) 1.5–2.2 (1.21–2.07); 1.20 (1.01–1.43); 0.76 (0.62–0.92 protective); and 2.50 (1.54–4.05), respectively] (reviewed in [28]). Chromosomal instability is common in colorectal cancer [29]. Associations between telomere shortening, chromosomal instability, and increased susceptibility to colorectal tumors and other types of epithelial cancers are apparent in telomerase-deficient and p53-null mice [30]. The presence of colorectal cancer susceptibility traits and evidence of chromosomal instability or telomere dysfunction can be assessed simultaneously in patient samples using NGS methods.

In addition to the preceding list of genetic traits, dietary and environmental factors impact the risk of developing colorectal cancer. Exogenous risk factors of this type may increase the frequency of sporadic mutations resulting from DNA damage. Heterocyclic amines are well-known colon carcinogens that act in this way [31, 32]. Other exogenous nongenotoxic risk factors can affect cell growth, death, or differentiation processes in ways that lead to the selection of mutations that provide growth and survival advantages to mutant cells. Inflammatory stimuli that activate cyclooxygenase-2 and nuclear factor kappaB signaling are believed to promote colorectal cancer via this mechanism [33]. Or, exogenous factors may induce epigenetic changes, such as DNA methylation, that favor development of a cancerous phenotype. Dietary folate deficiency was associated ($p = 0.01$) with increased *APC* promoter methylation among colorectal cancers in a report by Gay and coworkers [34].

The consumption of well-done cooked red meat is considered an exogenous risk factor for colorectal cancer [35] because it is a source of heterocyclic amines including 2-amino-3,4,8-trimethylimidazo[4,5-f]quinoxaline 2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline, 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) which damages DNA leading to mutations [36]. Protein and DNA adducts were detected in the colons and in the blood of human volunteers exposed to labeled PhiP at levels found in the diet [37]. Human metabolism plays a role in the activation and the detoxification of heterocyclic amines relevant to colorectal cancer. Human cytochrome P450 isoforms CYP1A2, CYP1A1, CYP1B1, and CYP2A6 catalyze *N*-oxidation of heterocyclic amines and the resulting hydroxylamine derivatives thus produced are activated by acetyltransferase (NAT1, NAT2) or sulfotransferase (SULT1A1) activity to form electrophilic *N*-acetyloxy or *N*-sulfonyloxy

esters that react with DNA [32, 38]. The inheritance of alleles that provide enhanced catalytic activity for enzymes responsible for the metabolic activation of heterocyclic amines has been associated with increased risks for colorectal cancer [32, 39, 40]. Similarly, inheritance of alleles that result in reduced activity for enzymes involved in detoxification, including glutathione *S*-transferase (*GSTA1***B*) [41] or UDP-glucuronosyltransferase (*UDP1A1*-3279 GG/TG in Caucasians) [42], are also linked to increased risk for colorectal cancer. Exposure to heterocyclic amines results in a diagnostic mutational signature that involves G:C single base pair deletions, particularly within 5′-GGGA-3′ motifs, along with G:C → T:A transversions [43]. The status of each gene involved in heterocyclic amine metabolism within a tumor sample can be determined using NGS and, by comparison with normal adjacent tissue, the presence of signature mutations associated with exposure to heterocyclic amines can also be identified within colorectal tumors.

In addition to the inherited traits described previously that influence the risk for colorectal cancer, certain genes have been found to be targeted by induced, or acquired, somatic mutations in colorectal cancer. Somatic mutations or epigenetic silencing of the *KRAS*, *DCC, TP53,* and *CDKN1B* genes are important in colorectal cancer (reviewed in [44]), along with epigenetic silencing of *MLH1* followed by mutations in *TGFBR2*, *BRAF*, and *BAX* genes and chromosome 18q deletions involving the *SMAD4* and *DCC* tumor suppressor genes (reviewed in [45]). Because NGS techniques can provide a comprehensive analysis of genetic and epigenetic aberrations, induced mutations and epigenetic changes involving all genes associated with colorectal cancer can be accessed directly. If environmental exposures for study populations are also known, molecular fingerprints for colorectal cancer biology may be obtained from enhanced molecular epidemiologic studies using NGS-based strategies.

Important medical decisions regarding patient care choices could be improved by results obtained using NGS methods. Two treatment modalities approved by the U.S. Food and Drug Administration for metastatic colorectal cancer involve the anti-epidermal growth factor receptor (EGFR) antibodies panitumumab, used alone, and cetuximab in combination with irinotecan, 5-fluorouracil, and leucovorin (FOLFIRI). Reported response rates to cetuximab and panitumumab monotherapies for metastatic colorectal cancer are limited to 8%–10% [46, 47] in patient populations that have not been screened for *KRAS* mutation status. Use of either anti-EGFR antibody requires genetic screening [48] because response rates are much lower for colorectal tumors bearing mutant *KRAS* genes [49, 50]. Increased detection sensitivity of *KRAS* mutation status provides enhanced predictive confidence for anti-EGFR antibody therapy [51, 52]. Furthermore, when expression of PTEN and the mutation status of *KRAS, BRAF,* and *PIK3CA* were determined simultaneously, up to 70% of responders could be identified [53]. These observations are consistent with a hypothesis presented by Parsons and Myers [54] that most, if not all, colon tumors contain small subpopulations of cells harboring undetected *KRAS* mutations and that outgrowth of these *KRAS* mutant tumor cells leads to relapse following anti-EGFR antibody therapy. Although preliminary studies with mutant *KRAS* colorectal tumor cells showed sensitivity to the combination of the MEK inhibitor AZD6244 and cetuximab [55], acquired resistance to AZD6244 developed in the human colon cancer cell lines HCT116 and COLO205 that was

conferred through selective amplification of the *KRAS* ^{G13D} and *BRAF* ^{V600E} oncogenes, respectively, that they harbored [56]. Thus, future studies are needed to determine whether co-targeting the RAS/RAF/MEK/ERK signaling pathway simultaneously with anti-EGFR antibodies will lead to improved long term survival of metastatic colorectal cancer patients. The combination therapy for colorectal cancer of 5-fluorouracil (5-FU) with leucovorin, a folate derivative, depletes dTTP pools via inhibition of the *de novo* thymidylate synthesis pathway. In addition, 5-FU acts by also interfering with ribosomal RNA processing via the uridine monophosphate kinase (UMPK)-dependent misincorporation of fluorouridine triphosphate into pre-ribosomal RNA transcripts. Interestingly, variant alleles affecting the 5′- and 3′-untranslated regions of the thymidylate synthase gene (*TYMS*) were found to be protective for colorectal cancer progression to advanced stages (OR = 0.5, 95% CI: 0.3–0.9) [57], and overexpression of thymidylate synthase has been associated with resistance to fluorouracil plus leucovorin therapy in colorectal cancer [58]. Other markers associated with either response or resistance to fluoropyrimidine therapy for colorectal cancer include the expression of UMPK [59], dihydropyrimidine dehydrogenase, thymidine phosphorylase [60], and orotate phosphoribosyltransferase [61]. Variants forms of the *UGT1A1* gene (reviewed in [62]) have been associated with adverse reactions to irinotecan therapy for metastatic colorectal cancer (severe neutropenia and diarrhea), although the Evaluation of Genomic Applications in Practice and Prevention Working Group found that further studies are warranted before recommending routine *UGT1A1* genotyping for irinotecan therapy [63]. Powerful NGS technologies allow the expression levels and mutation status to be evaluated simultaneously for each genetic marker associated with sensitivity or resistance to candidate chemotherapeutic strategies for colorectal cancer and for known markers of adverse treatment outcomes.

GENETIC VARIANTS AND ENVIRONMENTAL EXPOSURE

Most human illnesses are outcomes of the interactions between inherited genetic factors and modifiable environmental factors [64]. Similarly, most cancer cases are the consequence of the interaction of genetic variants and environmental factors. It is believed that the genetic factors by themselves contribute to the development of approximately 5% of all cancers [3]. For example, familial breast cancer results from a predisposition due to genetic variants. However, only 5%–10% of all breast cancers run in families. Even in familial breast cancer, well-known mutations, such as mutations in BRCA1 and BRCA2 genes, account for only approximately 20% of the familial risk. The development of sporadic breast cancers, accounting for the rest of all breast cancers, might be related to the interaction of genetic variation and environmental exposures, such as radiation, chemical carcinogens, and xenoestrogens [65]. An individual may inherit a genetic make-up that predisposes them to the development of cancers, yet the onset of the cancer and the severity of the cancer are usually modified by exposure to environmental factors. In the biological DNA-RNA-protein-biofunction dogma, many studies have shown that environmental factors, such as nutrients, pharmaceuticals, air pollutants, carcinogens, and pesticides, have impacts on genomic events, such as gene mutations, gene transcription, histone modifications, DNA methylation,

and DNA repair, which ultimately affect disease phenotypes such as cancer, diabetes, allergies, and infertility [66].

The risk of developing cancer from environmental exposures may depend on particular genetic variants carried by individuals. A clear example is the greater risk of cutaneous melanoma as a result of ultraviolet (UV) exposure for individuals that harbor xeroderma pigmentosum variant (*XP-V*) alleles. Compared to normal individuals, *XP-V* carriers have an increased risk (1000-fold higher) of developing cutaneous melanoma after exposure to UV light (or sun-exposure) [67] because the defective form of human DNA polymerase-eta is unable to perform damage-bypass replication of UV-damaged DNA [68]. In terms of affecting the risk of cancer development, a less dramatic gene-environment interaction in affecting the risk of cancer development is illustrated by variants in the Cockayne syndrome B protein (*ERCC6*). ERCC6 is a base-excision repair enzyme playing an important role in DNA repair. The association between a common variant (rs3793784: C > G) in *ERCC6* and lung cancer risk was examined in a case-control study with 1000 cases and 1000 controls. A 1.76-fold ($p < 1.0\times10^{-7}$) excess risk of developing lung cancer was found for the -6530CC carriers compared with noncarriers. Through gene-environment interaction analysis, it was found that the −6530CC allele interacts with smoking to heighten lung cancer risk. An odds ratio of 8.87 (95% CI 5.74–13.71) for developing lung cancer was found among heavy smokers, suggesting the importance of environmental factors in the development of genetic variant-related diseases [69]. Colorectal cancer provides another example of the importance of gene-environment interaction. The risk/protective factors for colorectal cancer include family history, consumption of red meat, fat, vegetables, fruits, fiber, alcohol, cigarettes, micronutrients, and anti-inflammatory drugs. Genetic mutations in genes involved in the *APC* (adenomatous polyposis coli)-*β*-catenin-Tcf (T-cell factor) pathway, DNA mismatch repair pathway, and xenobiotic metabolizing pathway are significantly associated with the incidence of colorectal cancer, and these genetic factors interacting with environmental factors intensify the risk [70].

Molecular epidemiological studies have revealed that many environmental factors may interact with genetic variants to affect the risk of cancer development. Humans are continually exposed to harmful environmental factors or hazardous life styles and have evolved metabolic pathways and cellular controls that can minimize the biological impact introduced by hazardous environmental agents. Such pathways and controls involve the metabolic activation or detoxification of carcinogenic chemicals, controlling the cell cycle, cell differentiation and cell death, and DNA repair [71]. All the genes involved in these environmental response pathways are subject to genetic variability, which in turn can change the efficiency of these biological functions. Most cancers seem to be polygenic, arising from complex interactions among several genes or genetic variants, and gene-environment interactions. With the exception of some high penetrance, dominant alleles (such as BRCA1 discussed above, and others discussed next), susceptibility genes or genetic variants are usually insufficient to cause disease. Although these low penetrance genes or genetic variants only have relatively small impacts on increasing disease risk (e.g., only increase the risk by several fold), they could contribute a significant effect on the incidence of cancer development in a human population owing to their relative high allele frequency [3]. In

general, a specific individual's risk for developing a cancer is dependent on the interaction of environmental exposures to carcinogenic agents within the context of that person's inherited set of cancer susceptibility or resistance traits. Table 1 lists some examples of known genotypic markers associated with particular types of cancers and environmental risk factors.

Although some of the more abundant genetic variants may pose lower individual risk or lower penetrance than others, it has been shown that these low penetrance genetic traits contribute to substantial differences in cancer susceptibility among populations. Moreover, gene–environment interactions may intensify the risks of cancer development among carriers of these low penetrance traits, thus they are important for the determination of population risk.

On the other hand, numerous cancer genes have been identified as germline-inherited tumor predisposition genes with very high penetrance. Well-known examples of inherited cancer susceptibility traits include *RB1* for retinoblastoma [72], *TP53* for colorectal cancer [73], and for the familial syndrome of breast cancer, sarcomas, and other neoplasms [74], *BRCA1* [75] and *BRCA2* [76] for breast cancer and ovarian cancer, *RET* for familial medullary thyroid carcinoma [77], *MET* for hereditary papillary renal carcinoma [78], *APC* for colorectal cancer [79], *CDKN2A* for melanoma [80], and *KRAS* for bladder, lung [81], colorectal, and other cancers [82]. Mutations in these genes confer high risks of cancer development among carriers. Compared to low penetrance genetic variants, the high penetrance germline mutations have the following characteristics: they cause familial cancers; their allele frequencies are generally rare in the population; their risks are sufficient to increase the incidence of a cancer in the affected families; and, because powerful activating mutations are already present in the germline, genotoxic environmental factors play a smaller role in carcinogenesis among these families [83].

EPIGENETIC ALTERATIONS AND ENVIRONMENTAL EXPOSURE

While cancer is considered a genetic disease as discussed, it also can be considered as an epigenetic disease. Epigenetics refers to stably maintained or altered gene expression patterns, without changing the underlying DNA sequence, via DNA methylation [84], chromatin modification [85], and ncRNA modulation [86]. The epigenetic mechanisms that control gene expression are based on covalent modifications of DNA and/or protein molecules, or modulation of protein translation. These mechanisms are susceptible to environmental challenges such as chemicals or metabolites derived from the diet, nutritional deficiencies, and exposure to other xenobiotics [87]. The dynamics of DNA methylation/de-methylation, chromatin status, and modulation of protein translation by ncRNA each function in a precisely controlled manner to maintain cellular homeostasis by fine-tuning gene expression patterns in normal cells. However, dramatic changes of epigenetic status may result in the activation of oncogenes, inactivation of tumor suppressor genes, induction of chromosomal instability and mutations, abnormal expression and translation of genes, and production of aneuploidy, all being pivotal biological events in the carcinogenic process [88, 89].

DNA methylation is achieved by addition of a methyl group to the cytosine in CpG dinucleotides, catalyzed by DNA methyltransferases. The genome of the cancer cells is usually characterized by aberrations in DNA methylation: hypermethylation in the CpG island of tumor suppressor genes and hypomethylation in the global genome [88]. In many cancers, hypermethylation at CpG islands in the promoter regions of tumor suppressor genes is usually a major cancer-predisposing event. Hypermethylation provides a mechanism to inactivate a variety of genes involved in normal biological functions that also function to repress cancer. Hypermethylated targets associated with cancer include tumor suppressor genes (e.g., *VHL*, *p15*, *p16*, *p73* and *BRCA1)*, DNA-repair genes (e.g., *hMLH1* and *MGMT*), drug metabolizing genes (e.g., *GSTP1*), cytokine receptor genes (e.g., *SOCS1*), and apoptotic signal genes (e.g.,*TMS1*) [4]. The hypermethylation pattern is cancer-specific, with different types of cancer utilizing different sets of hypermethylated genes [4].

In contrast to hypermethylation of CpG islands in tumor suppressor genes, the overall genome of cancer cells is globally hypomethylated. It has been found that cancer cell genomes have only 20%–60% of the methylation level as their normal counterparts [90, 91]. Hypomethylation commonly affects the exonic and intronic regions of genes, and has biological impacts on carcinogenesis through mechanisms which include the increase of chromosomal instability, reactivation of transposable elements, loss of imprinting, recombination and chromosome rearrangement, and the promotion of aneuploidy [88]. However, a comparison of methylation patterns in the colon cancer cell line SW48 to those of primary colon cells revealed that the methylation patterns were markedly similar between the primary cells and the transformed cells, “suggesting that aberrant methylation of CpG island promoters in malignancy might be less frequent than previously hypothesized” [92].

In eukaryotes, DNA molecules are associated with small proteins including histones, which are the fundamental components of chromatin. Chromatin remodeling controlled by chemical modifications of the histones is important for many biological processes [93]. Dynamic structural changes of chromatin affect the “packing” of DNA chains. Highly condensed chromatin can exclude the transcriptional machinery, thus influencing the expression of genes, DNA replication and repair, chromosome condensation and segregation, and apoptosis [94, 95]. Highly regulated modifications of histones, including histone acetylation/deacetylation, phosphorylation/dephosphorylation, and methylation/demethylation are involved in pivotal cellular events, such as proliferation and differentiation, and are thus modifiers of the carcinogenesis process [96]. For example, the dynamics of histone acetylation, which results in a more open, accessible chromatin state, is precisely regulated by histone acetyltransferases (HAT) and histone deacetylases (HDAC). Impaired enzymatic activities of HAT and/or HDAC are associated with the development of cancers: mutated/truncated HATs have been found in colorectal, gastric tumors and other epithelial cancers; translocation-generated HDACs have been identified in leukemia and lymphoma [97].

ncRNA is another epigenetic modulator for gene expression that can operate through two mechanisms: translational repression by microRNA (miRNA) and mRNA degradation by small interfering RNA (siRNA). These ncRNAs are important players in gene regulatory

networks. Consequently, mutations and or genetic variants residing in ncRNA sequences, as well as differential expression of ncRNA molecules among individuals, could lead to a substantial change in phenotypes. Such changes may have a significant influence on quantitative traits, including the development of cancer and other diseases [98].

As a complex process, carcinogenesis is a result of aberrant cellular processes caused by both genetic and epigenetic events. Epigenetic machinery, including DNA methylation, histone modification, and ncRNA expression, can be modulated by environmental exposure such as dietary chemicals and pollutants. Dietary factors, such as vitamins and micronutrients, are critical for the dynamic balance of epigenetic patterns that keep cells healthy. For example, the one-carbon metabolic pathway is utilized to generate *S*-adenosylmethionine, the methyl donor necessary for DNA methyltransferase activity. Choline, methionine, folate, vitamin B_{12}, vitamin B_6, and riboflavin are strategic dietary factors involved in the efficient maintenance of methyl donor pools and healthy DNA methylation status [99]. Deficiencies of diet-provided substrates and cofactors in one-carbon metabolism may therefore result in impaired DNA methylation which is linked to increased risks for neural tube defects, cardiovascular diseases and cancers [100]. Xenobiotic stressors, including heavy metals, water disinfection byproducts, air pollutants, endocrine-disrupting toxicants, and cigarette smoke, can also affect epigenetic components. Table 2 summarizes impact of environmental chemicals in terms of their effects on epigenetic alterations.

GENE EXPRESSION AND TOXICOGENOMICS

The merging of conventional toxicological research and functional genomics resulted in the emergence of toxicogenomics after the introduction of DNA microarray (i.e., transcriptomics) technology and, more recently, NGS [101]. Toxicogenomics has now incorporated other high throughput omic technologies such as proteomics and metabonomics. Using bioinformatics software and databases, the raw data are analyzed and meaningful biological information such as gene expression changes and alterations in biological pathways/functions are obtained. Therefore, toxicogenomics has become an important sub-discipline in the field of toxicology, with its primary goals to understand the relationship between environmental stress and human disease susceptibility, such as cancer susceptibility; exploring the molecular mechanisms of environmental mutagens and/or carcinogens; and identifying potential biomarkers of disease and toxicity, including mutagenicity and carcinogenicity [102].

Toxicology has traditionally relied on in vivo and in vitro models to study the adverse effects of chemicals, including cellular, biochemical, and molecular mechanisms of action. Toxicogenomics has been used to understand toxicant-induced effects in vivo and in vitro and to address challenges that are difficult to overcome by conventional toxicology methods. Among the omic technologies, microarray-based toxicogenomics methods have been exploited most extensively for mechanistic studies because they can evaluate the relative expression of thousands of genes conveniently and cheaply to identify early biomarkers of toxicity and disease [103]. Recently, next-generation sequencing (i.e., RNA-Seq), a newly developed technology, has been used in toxicogenomics studies and it has significantly

accelerated genomic research and discovery. Compared to DNA microarrays, RNA-Seq may provide more sensitivity in detecting genes with low expression levels [104, 105].

Genotoxic carcinogens cause genetic errors via molecular modifications of DNA that lead to mutations that become fixed following DNA replication. Because of the association between DNA damage and cancer development, genetic toxicology data have been used for hazard identification and cancer risk assessment. A multitude of toxicogenomics studies have been performed to evaluate a number of environmental mutagens and/or carcinogens with both traditional toxicity endpoints and genomics changes. Generally, there are two ways to perform toxicogenomics studies: (1) starting with conventional toxicological research and then focusing on omics approaches to detect systematic biological effects, and (2) starting with the omics study followed by conventional toxicological research to interrogate molecular mechanisms. NGS technologies have been used in both approaches [105–108].

Approaches for investigating simple chemically-induced toxicity or carcinogenicity are well established. For example, aristolochic acid is a potent human nephrotoxin and carcinogen, and the International Agency for Research on Cancer classified it as a Group 1 human carcinogen [109]. DNA adduct formation and mutagenicity by aristolochic acid were detected in the kidney, spleen, and liver of Big Blue transgenic rats gavaged with 0, 0.1, 1.0, and 10.0 mg/kg body weight aristolochic acid for 3 months [110, 111]. Treatments with aristolochic acid resulted in three major DNA adducts in the tissues tested and the induced mutations in the *cII* gene exhibited linear dose-responses to aristolochic acid. A:T → T:A transversion was identified as the predominant mutation in aristolochic acid-treated rats [110, 112]. A dose-dependent induction of H-*Ras* mutation fraction at codon 61 having CAA → CTA mutation in liver and kidney was also determined [113]. Using microarray analysis, it was revealed that gene expression alterations in cancer-related pathways were more significant in kidney than in liver. The genes involved in the biological processes related to defense response, apoptosis, and immune response were significantly altered by aristolochic acid exposure in kidney, but not in liver [114]. Later, the same set of RNA samples were analyzed by NGS using an Illumina Genome Analyzer II [105]. When gene ranking was compared between NGS and microarray, the overlap of the differentially expressed genes was about 40%, because the dynamic range of the NGS platform is greater than that of the microarray technology. However, when the common gene ontology terms generated from the differentially expressed genes were compared for the two methods, the biological interpretation was largely consistent between the NGS and microarray data [105]. NGS technologies are capable of generating more data and providing additional insight into the mechanisms involved in toxicity and carcinogenesis associated with environmental mutagen/carcinogen exposure.

It is a challenge to determine the mechanisms of toxicity induced by mixtures containing many chemical components, such as diesel exhaust, herbal dietary supplements, and cigarette smoke condensates. Tobacco smoke contains more than 4000 chemicals, of which >200 chemicals are toxic to humans and >50 of them are recognized as known or probable human carcinogens [115, 116]. Tobacco smoking is a major public health problem that threatens the lives of one billion people and tobacco use is estimated to kill more than five-

million people worldwide each year [117]. Using the NGS technology (i.e., the SOLiD platform), Pleasance and colleagues [107] sequenced a small cell lung cancer cell line NCI-H209 to investigate the mutational burden associated with tobacco smoking. Overall, 22,910 somatic substitutions were observed across the NCI-H209 genome, illustrating the power of NGS to detect the many complex mutational signatures associated with tobacco smoke-induced cancers [107]. Beane and associates [106] evaluated total RNA from bronchial airway epithelial cells of current smokers using both microarray and RNA-Seq methods and observed a significant correlation between the RNA-Seq gene expression data and Affymetrix microarray data generated from the same samples. The RNA-Seq technology detected more smoking- and cancer-related gene expression differences, and NGS data has the potential to provide greater detail for information concerning mutations that may provide insights into the biological changes associated with smoking and lung cancer [106].

ADVANCE OF NGS TECHNOLOGIES

NGS technologies and platforms have rapidly evolved over the past eight years with a number of innovations and developments in sequencing chemistries, microfabrication and imaging to increase speed, throughput, and accuracy, and decrease costs of sequencing [118, 119]. There are currently four companies including Illumina, Life Technologies, Roche/454, and Pacific Biosciences offering a variety of NGS platforms in the NGS market. Also, many new technologies with promising features, such as nanopore sequencing, are under development and it is likely that commercialization will take place in a few years [120]. Different NGS platforms may rely on dissimilar sequencing chemistries and techniques, but they all share the technical strategy of miniaturization and parallelization of individual sequencing chemical reactions to boost sequencing speed and throughput [121]. Roche/454 was the first to commercialize their NGS platforms, but Illumina, which developed the first short read sequencer, is currently dominating the NGS market. Each NGS technology and platform has its own pros and cons, and may fit different analysis applications. Table 3 provides a summary of the characteristics of commercially available and near-to-release NGS platforms. In this review, we will not discuss each NGS technology and related platforms in detail as most of them have been well-reviewed elsewhere [121, 122]. We will focus on recently developed technologies and small-scale, low-cost platforms including Pacific Biosciences PacBio RS II, Illumina MiSeq, Life Technologies IonTorrent, and Oxford nanopore sequencing.

The Pacific Biosciences PacBio RS II is the latest upgrade of its first commercial NGS sequencer PacBio RS which was released in 2011. The new system produces longer reads and offers higher sequence throughput than the original instrument. But both platforms utilize the company’s zero-mode waveguide (ZMW) technology [123] to conduct single molecule real-time sequencing (SMRT) [124]. The sequencing by synthesis is conducted on a small plastic SMRT cell. Each SMRT cell contains thousands of ZMWs. A ZWM is a nanophotonic confinement structure that can be used to detect fluorescence signals from single nucleotide incorporation. In each ZMW, a single active DNA polymerase enzyme is immobilized with a single molecule of single-stranded DNA template. During sequencing by synthesis, four nucleotides labeled with different fluorescent dyes are supplied to the SMRT

cell. When a nucleotide is incorporated, the fluorescent dye is cleaved off as a part of a natural DNA synthesis process. The fluorescent signal is detected and the base call is made in real time. The cleaved fluorescent dye molecule then diffuses out of the observation area of the ZMW. The SMRT sequencing can generate very long reads (average read length >4 kilobases [125] with relatively low throughput and has been used successfully for *de novo* genome sequencing [125], transcriptome sequencing [126], targeted resequencing [127, 128], methylation detection [129], and in vitro diagnostics [130].

The Illumina MiSeq employed the same technologies for sequencing as those used in the earlier HiSeq systems but is aimed at smaller laboratories and clinical diagnostic applications. In contrast to the HiSeq series of platforms, the MiSeq is a lower throughput but fast-turnaround instrument in which cluster generation, bridge PCR amplification, sequencing by synthesis with fluorescently labeled reversible-terminator nucleotides, and data analysis were integrated and streamlined without user intervention needed [131]. Such a design not only provides rapid, cost-effective NGS analysis but also offers much more flexibility for applications such as clone checking, amplicon sequencing, targeted transcript sequencing, small genome resequencing, *de novo* sequencing, and small RNA sequencing [132–134].

The Life Technologies Ion Torrent semiconductor sequencing is based on the detection of protons that are released when nucleotides are incorporated into growing DNA strands during sequencing-by-synthesis reactions. Libraries are constructed by the fragmentation of DNA or cDNA, and fragments are then linked to specific adapter sequences to form sequencing templates which are clonally amplified with emulsion PCR. The amplified products are subsequently loaded onto an Ion Torrent chip which contains millions of ion-sensitive field-effect transistor (ISFET) sensors that allow parallel detection of multiple sequencing reactions [135, 136]. Sequencing is primed from a specific position in adapter sequences and each of the four nucleotides is supplied sequentially. If a specific type of nucleotide is incorporated, protons are released and pH is altered and measured with an ISFET sensor. Although pH signal is proportional to the number of bases incorporated, the determination of homopolymer bases introduces major sequence errors as false insertions or deletions (Indels). Without the needs for modified nucleotides and optical signal detection, the Ion Torrent sequencing process is much faster and its instruments, chips, and reagents are less expensive than the other platforms. Recently, Life Technologies has released the Ion PGM sequencer with Ion 314/316/318 chips and the Ion Proton sequencer with Ion PI chips to provide cheap and rapid sequencing.

The Oxford Nanopore strand sequencing is based on the detection of electric current changes as a single DNA sequence strand passes through a nanopore, which is a nanoscale hole formed by proteins and set in an electrically resistant membrane bilayer [137]. A constant voltage is set across the bilayer membrane which creates an ionic current passing through nanopores. If a DNA strand passes through a nanopore, each of the four nucleotides induces a different type of electric current change which can be translated into sequence data. By monitoring electric current changes in real time, the base call can be made for a single DNA molecule. Oxford Nanopore developed two sequencing systems, that is,

MiniION and GridION, and planned to commercialize at the end of 2013. The MiniION is a disposable self-contained device for real-time single molecule sensing sequencing and can be run directly with a laptop or desktop computer through a USB port. The GridION is a scalable system consisting of single or multiple nodes. Each GridION node works with a disposable, self-contained cartridge for sequencing experiments. Currently, the Nanopore DNA sequencing platform can be used to produce very long read-lengths with average lengths of 4200 to 8500 bases [138]. It also has the potential to be adapted for RNA, protein, and other polymer sequencing.

Advancements in NGS technologies have enabled a number of sequencing-based analytical approaches that were not affordable previously or even technically feasible. Over the past few years, NGS-based approaches have been widely investigated and implemented in a variety of biological and biomedical research settings including *de novo* genome sequencing [139], detection of genetic variation [140], quantification of gene expression [105], discovery and annotation of genes and transcripts [141], and analysis of gene regulation [142]. Recently, the release of low cost, smaller-scale, and easy-to-use NGS platforms such as MiSeq, Ion Torrent, and Nanopore sequencing, make NGS technologies and NGS-based approaches more accessible and will speed up the application of NGS technologies in clinical settings [120]. Table 4 provides a summary of NGS-based approaches and their potential applications.

LIMITATIONS AND CHALLENGES IN NGS

Although NGS technologies are very powerful when applied appropriately to certain research and clinical applications, many investigations have revealed that there are significant technical drawbacks that limit the utility of NGS technologies in certain situations. Beyond the high cost of NGS instruments and reagents, the storage, management, analysis, and interpretation of NGS data are also posing challenges. For example, a study [143] in which NGS was tailored to detect RNA-editing events reported that an astonishingly large number of exonic mismatches were discovered between mRNA sequences and corresponding DNA sequences from the same individuals and showed that many of these RNA sequence changes were not produced by any known RNA regulatory mechanisms. This report questioned the basic assumption in molecular biology that DNA and RNA of the same individuals are nearly identical. However, this report has been criticized by several groups who reanalyzed the data and found that the majority of those mismatches found in the report can be explained either by systematic technical errors in NGS technology or by analytical artifacts [144–147]. Thus, it is crucial to analyze NGS raw sequencing data properly, such as FASTQ files, and translate them into final variant calls. In addition, the single-base resolution capability of NGS-based approaches in the detection of sequence variation can be compromised by the biases, artifacts, or errors from NGS-related protocols, platforms, sequencing depth, data analysis pipelines, and reference genomes or transcriptomes, resulting in both high rates of false positive and false negative results. Bearing this fact in mind will help avoid pitfalls in identifying true variants from various artifacts generated by NGS technologies.

APPLICATION OF NGS IN CANCER GENOMICS AND EPIGENOMICS

Advances in NGS technology made possible new approaches to cancer research. With much faster speed, higher throughput, and lower cost, the NGS platforms provide tools to identify the genetic mutations, epigenomic alterations and transcriptome changes occurring in cancer cells and to establish a framework for understanding the complexity and heterogeneity of cancers. Cancer genome sequencing is an endeavor to sequence homogeneous or heterogeneous groups of cancer cells to characterize DNA or RNA molecules for discovery of their mutations and gene expression variations. A careful cancer genome sequencing strategy not only performs sequencing analysis on primary cancer cells/tissues directly but also analyzes paired adjacent or distal normal tissue, as well as tumor-related fibroblast/stromal cells and metastatic tumor tissues. The strategy is not limited to DNA sequencing of the whole genome, but also includes exome sequencing or target gene sequencing, transcriptome sequencing or RNA-Seq, and sequencing analysis of the epigenome. Single base-resolution sequencing allows for the characterization of DNA or RNA molecules for the discovery of cellular aberrations including sequence variants, structural rearrangements (such as chromosomal translocation and gene fusion), and variations in gene expression, gene copy numbers, and epigenetic status [148].

Various databases are being developed to organize the massive amount of cancer genome sequence data being generated. The Cancer Gene Census (http://cancer.sanger.ac.uk/cancergenome/projects/census/) has documented 513 genes (as of December 2013) with a wide variety of mutations that are associated with cancers. These mutations include chromosomal translocations, gene fusions, gene amplifications, large deletions, frameshift mutations, and missense mutations. Approximately 90% of these are dominant mutations occurring in oncogenes, thus one allele is sufficient to activate an encoded protein (oncogene) to cause cancer; whereas 10% of these mutations occur in tumor suppressor genes and act in a recessive manner [2]. Launched in 2004, the COSMIC (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/) is an online database that curates somatic mutations found in human cancers from the scientific literature and large-scale experimental screens from the Cancer Genome Project (http://www.sanger.ac.uk/research/projects/cancergenome/). The COSMIC database presents complex phenotype-specific mutation data in a statistical manner. As of October 2013 (version 67), COSMIC has documented over 1,592,109 mutations, 9190 gene fusions, 7584 genomic rearrangements, and 422,314 copy number aberrations that affected 25,606 genes from 947,213 tumor samples. These findings were identified from 17,731 scientific papers. COSMIC has the capability to annotate somatic mutations across the whole genome. With data from large-scale systematic candidate gene screening, whole genome sequencing, and exome sequencing, COSMIC has curated genomic information from 7954 whole genomes. For example, by sequence analysis of coding exons of 21,416 protein coding genes and 1664 miRNAs in 100 primary breast cancer samples, somatic copy number changes and mutations in the coding exons of protein-coding genes were documented. To exclude inherited sequence variation, paired normal DNAs from the same individuals were also sequenced. In such a way, multiple mutational signatures and driver mutations were identified in at least 40 cancer genes. Novel cancer genes were also identified and included *AKT2, ARID1B,*

CASP8, CDKN1B, MAP3K1, MAP3K13, NCOR1, SMARCD1, and *TBX3*, demonstrating the genetic diversity, heterogeneity, and complexity of primary breast cancer [149].

Sponsored by the National Cancer Institute and the National Human Genome Research Institute, The Cancer Genome Atlas is a project aimed at cataloguing genetic mutations in 20–25 major cancer types, including glioblastoma, breast, colorectal, stomach, ovarian, prostate, and lung, using high-throughput genome sequencing technology and bioinformatics approaches (http://cancergenome.nih.gov/). The definition of mutations in the various cancers will allow a better understanding of the biology and pathology of cancer, leading to improved cancer diagnoses, treatments, and prevention. Identification of mutations in glioblastoma, ovarian, colorectal, lung, head and neck, multiple myeloma, chronic lymphocytic leukemia, large B-cell lymphoma, and many other cancers have been accomplished. For instance, Lawrence and colleagues reported a great heterogeneity of mutations in cancer-associated genes [150]. Among 27 cancer types, with a sequencing data set of 3083 tumor–normal pairs, including 2957 pairs by whole-exome sequencing and 126 pairs by whole-genome sequencing, a total of 373,909 non-silent coding mutations were found. The average mutation frequency in the exome was approximately 4.0/Mb per sample, and the median mutation frequency was 44 nonsilent coding mutations per sample, or 1.5/Mb per sample. Furthermore, across different cancer types and among different patients with the same cancer type, the exome mutation frequency and spectrum were extraordinarily variable. The lowest mutation frequencies, as low as 0.1/Mb (approximately one change across the entire exome) were found in pediatric cancers. In contrast, the frequency of mutations in the exomes of melanoma and lung cancer were more than 100/Mb, more than 1000-fold higher than in pediatric cancers. Notably, the highest mutation frequencies appear to be associated with extensive exposure to environmental carcinogens, such as ultraviolet radiation in the melanoma patients, and tobacco smoke in lung cancer patients [150]. In addition, among lung cancer patients, smokers have 10 times more somatic mutations than nonsmokers, indicating the involvement of environmental carcinogens in the pathogenesis of lung cancer [151].

Based on the large international efforts described, as well as others, it has become possible to start defining a comprehensive cancer genome landscape that will aid our understanding of this complex disease process and help in the diagnosis, treatment, and prevention of cancer. This cancer genome landscape for common types of human cancer has revealed previously unanticipated complexities in the patterns of somatic alterations in cancers [152, 153]. The landscape of common human cancer suggests that approximately 140 genes are frequently altered in many types of cancer. The altered genes conferring a selective growth advantage are defined as driver genes and the mutations in these genes are driver mutations. Driver mutations contribute greatly to individuals' cancer susceptibility. Owing to the knowledge provided by NGS technology, it is realized that a cancer can be initiated by 2–8 driver mutations in driver genes [152, 153].

While whole-exome sequencing has been widely used in cancer genome analysis, this strategy is unable to efficiently detect gene fusions, which are important drivers of several types of blood cancers. Sequencing of the transcriptome (RNA-Seq), however, is a powerful

tool to detect fusion genes, in addition to somatic mutations, novel transcripts and alternative splicing variants expressed in tumor cells. Furthermore, it can be used to analyze gene expression profiles at a low cost and high sensitivity [154]. RNA-Seq has been used to identify gene fusion events in many cancer types, such as breast cancer [155], lung cancer [156], prostate cancer [157], and glioblastoma [158, 159]. For example, using RNA-Seq data, Shah et al [159] were able to identify gene fusion events in approximately 30%–50% of 185 glioblastoma multiforme tumors. The majority of fusions were located at the chromosomal 7p11 and 12q14–15 regions. Further analysis demonstrated complex genomic rearrangements in these affected chromosomal regions that result in aberrant *EGFR* gene amplification of a tyrosine kinase domain, suggesting the importance of gene fusion in glioblastoma. On the other hand, the use of RNA-Seq to profile gene expression patterns in tumor-normal pairs may be able to provide a more comprehensive view that comprises the overall outcome resulting from the interaction between an individual's variable genetic makeups, the epigenetic regulation, environmental exposures, and mRNA splicing patterns. For example, by comparing of RNA-Seq profiles between of 17 breast cancer samples and matched normal breast tissues, Horvath and colleagues [160] identified a total of 4847 genetic and functional novel variations, including mutations, mRNA splicing changes, and alterations of gene expression. Their results indicated that the most prevalent novel variants associated with cancer related genes included *ESRP2, GBP1, TPP1, MAD2L1BP, GLUD2,* and *SLC30A8,* which are pivotal in breast cancer tumorigenesis [160].

NGS technology is also being used for surveying DNA methylation, mapping of transcription factor binding sites, and measuring histone binding sites [161, 162]. NGS can be applied to obtain the whole genome DNA methylation status at base-resolution level through bisulphite sequencing [163]. Genome-wide transcription factor binding sites can be analyzed using chromatin immunoprecipitation-mediated massively parallel DNA sequencing (ChIP-seq) [164]. Similarly, ChIP-seq can be used to reveal genome-wide histone modifications. These methods are advanced in comparison to array-based interrogation for discovery, since they make no *a priori* assumptions of where to interrogate the genome. As a high-throughput tool, NGS has added a new avenue to revolutionize our appreciation of how chromatin and DNA methylation are involved in the pathology of tumorigenesis.

Gene mutations that may drive cancer-relevant changes in the epigenome have been found by NGS technology. For example, in a comprehensive study, using whole-genome sequencing or whole-exome sequencing, combined with RNA-Seq, miRNA sequencing and DNA-methylation profiling, the Cancer Genome Atlas Research Network analyzed genomes from 200 cases of de novo acute myeloid leukemia (AML) [165]. DNA methylation changes were identified at 160,519 CpG loci across the AML genomes, which is 42% of sites tested. Among these methylation-altered loci, 67% of them had increased methylation, and 33% of them had decreased methylation. Further analysis indicated that some genetic mutations were highly associated with DNA methylation status. A significant decrease of DNA methylation was associated with *MLL* fusions or co-occurring *NPM1, DNMT3A,* and *FLT3* mutations. In a small portion of samples (7/200), genetic variants were identified in some miRNA genes. The miR-142 gene was highly expressed in AML samples, indicating

significant epigenomic alterations in AML [165]. In another study, genome-wide methylated CpG islands were profiled by NGS. It was found that the regulatory regions of many genes are hypermethylated and several repeated elements are hypomethylated in melanoma cell lines. By integrative analysis of DNA methylation data with RNA-Seq data, a gene co-expression network significantly related to melanoma was determined, which could partially explain the molecular pathogenesis of the disease [166].

FUTURE PERSPECTIVES

Multiple factors such as genetic variants, environmental exposure, and lifestyle are all involved in cancer etiology. Genome-wide association studies (GWAS) based on the hypothesis “common disease-common variants” have identified many susceptibility loci in the human genome for a variety of cancers using genotyping microarray technologies. The first wave of GWAS was successful in understanding cancer etiology. But the findings have not fulfilled the expectations of the scientific community because the cancer-associated genetic variants only explained a very small portion of cancer risk. As described in this article, research efforts to understand cancer susceptibility have been shifting to ascertaining genetic-environmental interactions and to identifying rare genetic variants with high penetrance as well as somatic variants. These new efforts have been made possible by comprehensive, rapid and relatively low-cost next-generation sequencing technologies. Significant progress in the following areas is expected to be achieved in the next few years.

Additional rare genetic variants with high penetrance will be identified that confer a portion of total cancer susceptibility, maybe larger than that explained by the common genetic variants that have been identified from GWAS. NGS technologies, unlike genotyping microarrays, enable interrogating all genetic variants (known and unknown, common and rare) by sequencing the whole genomes of cancer patients. However, challenges to distinguish the causal variants from the large number of apparently novel genetic variants present by chance in any human genome will need to be solved before true genetic variants can be identified.

Somatic mutations resulting from the interaction of genetic and environmental factors may be involved in the development of cancers by conferring growth advantages to cancer cells. Genotyping microarray technologies are hypothesis based, that is the genetic mutations that can be detected are known before they can be designed in the microarrays. Some, if not most, somatic mutations that confer cancer susceptibility may not be presently known so they will not be used in the microarray design. However, detection of genetic variants by NGS is hypothesis-free and, in principle, all somatic mutations in cancer patient genomes can be detected by NGS. Therefore, more somatic mutations that contribute to cancer risk are expected to be identified in future cancer genetic studies using NGS.

Understanding interactions between genetic and environmental factors will provide insight into the etiology of cancers, as well as susceptibilities, possibly allowing the prevention of cancers. Environmental factors can affect the expression of the genome through alterations in the epigenome. Application of NGS technology to the analysis of DNA methylation,

chromatin modification, and ncRNA will allow better definition of the role of the environment in cancer development. Coupling the genomic and epigenomic information may ultimately give the insights needed to reduce the human costs of this prominent disease.

To better understand cancer etiology, not only NGS data but also clinical data, lifestyle records, healthcare records, genealogy records and other types of information need to be analyzed and integrated. Analysis and integration of these data, as well as interpretation of the results, are the key steps for applying NGS technologies in ascertaining cancer susceptibility and translating this knowledge into public health actions. In the near future, powerful infrastructures will be developed to support the storage, access, analysis and management of the huge amount and diversity of data. Powerful bioinformatics tools will be developed for identification of genetic variants and genetic-environmental interactions that contribute to cancer risk.

With the expected progress described, we can look forward to a better understanding of cancer susceptibility that will lead to better diagnosis, treatments, and ultimately prevention.

Acknowledgments

The views presented in this article do not necessarily reflect those of the U.S. Food and Drug Administration.

References

1. Pfeifer GP, Besaratinia A. Mutational spectra of human cancer. Hum Genet. 2009; 125:493–506. [PubMed: 19308457]
2. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009; 458:719–724. [PubMed: 19360079]
3. Perera FP. Environment and cancer: who are susceptible? Science. 1997; 278:1068–1073. [PubMed: 9353182]
4. Esteller M. Epigenetics in cancer. N Engl J Med. 2008; 358:1148–1159. [PubMed: 18337604]
5. He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet. 2004; 5:522–531. [PubMed: 15211354]
6. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, et al. MicroRNA expression profiles classify human cancers. Nature. 2005; 435:834–838. [PubMed: 15944708]
7. Kondo Y. Epigenetic cross-talk between DNA methylation and histone modifications in human cancers. Yonsei Med J. 2009; 50:455–463. [PubMed: 19718392]
8. Zhang L, Zhou W, Velculescu VE, Kern SE, Hruban RH, Hamilton SR, Vogelstein B, Kinzler KW. Gene expression profiles in normal and cancer cells. Science. 1997; 276:1268–1272. [PubMed: 9157888]
9. Kan YW, Dozy AM. Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. Proc Natl Acad Sci U S A. 1978; 75:5631–5635. [PubMed: 281713]
10. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008; 26:1135–1145. [PubMed: 18846087]
11. Hartikainen JM, Tuhkanen H, Kataja V, Eskelinen M, Uusitupa M, Kosma VM, Mannermaa A. Refinement of the 22q12–q13 breast cancer–associated region: evidence of TMPRSS6 as a candidate gene in an eastern Finnish population. Clin Cancer Res. 2006; 12:1454–1462. [PubMed: 16533768]

12. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet. 2007; 39:870–874. [PubMed: 17529973]
13. Collins FS, Hamburg MA. First FDA authorization for next-generation sequencer. N Engl J Med. 2013; 369:2369–2371. [PubMed: 24251383]
14. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet. 2008; 24:133–141. [PubMed: 18262675]
15. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013; 500:415–421. [PubMed: 23945592]
16. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature. 2012; 486:405–409. [PubMed: 22722202]
17. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabe RR, et al. International network of cancer genome projects. Nature. 2010; 464:993–998. [PubMed: 20393554]
18. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013; 502:333–339. [PubMed: 24132290]
19. Ledford H. Big science: The cancer genome challenge. Nature. 2010; 464:972–974. [PubMed: 20393534]
20. Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? Nat Rev Cancer. 2007; 7:545–553. [PubMed: 17585334]
21. Kaur H, Li JJ, Bay BH, Yung LY. Investigating the antiproliferative activity of high affinity DNA aptamer on cancer cells. PLoS One. 2013; 8:e50964. [PubMed: 23341879]
22. Castle JC, Biery M, Bouzek H, Xie T, Chen R, Misura K, Jackson S, Armour CD, Johnson JM, Rohl CA, Raymond CK. DNA copy number, including telomeres and mitochondria, assayed using next-generation sequencing. BMC Genomics. 2010; 11:244. [PubMed: 20398377]
23. Magnani L, Stoeck A, Zhang X, Lanczky A, Mirabella AC, Wang TL, Gy-orffy B, Lupien M. Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. Proc Natl Acad Sci U S A. 2013; 110:E1490–1499. [PubMed: 23576735]
24. Abdel-Rahman WM, Mecklin JP, Peltomaki P. The genetics of HNPCC: application to diagnosis and screening. Crit Rev Oncol Hematol. 2006; 58:208–220. [PubMed: 16434208]
25. Steinke V, Engel C, Buttner R, Schackert HK, Schmiegel WH, Propping P. Hereditary nonpolyposis colorectal cancer (HNPCC)/Lynch syndrome. Deutsches Arzteblatt International. 2013; 110:32–38. [PubMed: 23413378]
26. Half E, Bercovich D, Rozen P. Familial adenomatous polyposis. Orphanet Journal of Rare Diseases. 2009; 4:22. [PubMed: 19822006]
27. Gammon A, Jasperson K, Kohlmann W, Burt RW. Hamartomatous polyposis syndromes. Best Practice & Research Clinical Gastroenterology. 2009; 23:219–231. [PubMed: 19414148]
28. de la Chapelle A. Genetic predisposition to colorectal cancer. Nat Rev Cancer. 2004; 4:769–780. [PubMed: 15510158]
29. Migliore L, Migheli F, Spisni R, Coppede F. Genetics, cytogenetics, and epigenetics of colorectal cancer. Journal of Biomedicine & Biotechnology. 2011; 2011:792362. [PubMed: 21490705]
30. Artandi SE, Chang S, Lee SL, Alson S, Gottlieb GJ, Chin L, DePinho RA. Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. Nature. 2000; 406:641–645. [PubMed: 10949306]
31. Fu Z, Shrubsole MJ, Li G, Smalley WE, Hein DW, Chen Z, Shyr Y, Cai Q, Ness RM, Zheng W. Using gene-environment interaction analyses to clarify the role of well-done meat and heterocyclic amine exposure in the etiology of colorectal polyps. Am J Clin Nutr. 2012; 96:1119–1128. [PubMed: 23015320]
32. Nowell S, Coles B, Sinha R, MacLeod S, Luke Ratnasinghe D, Stotts C, Kadlubar FF, Ambrosone CB, Lang NP. Analysis of total meat intake and exposure to individual heterocyclic amines in a case-control study of colorectal cancer: contribution of metabolic variation to risk. Mutat Res. 2002; 506–507:175–185.

33. Kraus S, Arber N. Inflammation and colorectal cancer. Current Opinion in Pharmacology. 2009; 9:405–410. [PubMed: 19589728]
34. Gay LJ, Mitrou PN, Keen J, Bowman R, Naguib A, Cooke J, et al. Dietary, lifestyle and clinicopathological factors associated with APC mutations and promoter methylation in colorectal cancers from the EPIC-Norfolk study. J Pathol. 2012; 228:405–415. [PubMed: 22864938]
35. Aune D, Chan DS, Vieira AR, Navarro Rosenblatt DA, Vieira R, Greenwood DC, Kampman E, Norat T. Red and processed meat intake and risk of colorectal adenomas: a systematic review and meta-analysis of epidemiological studies. Cancer Causes Control. 2013; 24:611–627. [PubMed: 23380943]
36. Sinha R, Kulldorff M, Chow WH, Denobile J, Rothman N. Dietary intake of heterocyclic amines, meat-derived mutagenic activity, and risk of colorectal adenomas. Cancer Epidemiol Biomarkers Prev. 2001; 10:559–562. [PubMed: 11352869]
37. Dingley KH, Curtis KD, Nowell S, Felton JS, Lang NP, Turteltaub KW. DNA and protein adduct formation in the colon and blood of humans after exposure to a dietary-relevant dose of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine. Cancer Epidemiol Biomarkers Prev. 1999; 8:507–512. [PubMed: 10385140]
38. Turesky RJ. Formation and biochemistry of carcinogenic heterocyclic aromatic amines in cooked meats. Toxicol Lett. 2007; 168:219–227. [PubMed: 17174486]
39. Lang NP, Butler MA, Massengill J, Lawson M, Stotts RC, Hauer-Jensen M, Kadlubar FF. Rapid metabolic phenotypes for acetyltransferase and cytochrome P4501A2 and putative exposure to food-borne heterocyclic amines increase the risk for colorectal cancer or polyps. Cancer Epidemiol Biomarkers Prev. 1994; 3:675–682. [PubMed: 7881341]
40. Hein DW, Doll MA, Fretland AJ, Leff MA, Webb SJ, Xiao GH, Devanaboyina US, Nangju NA, Feng Y. Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. Cancer Epidemiol Biomarkers Prev. 2000; 9:29–42. [PubMed: 10667461]
41. Coles B, Nowell SA, MacLeod SL, Sweeney C, Lang NP, Kadlubar FF. The role of human glutathione S-transferases (hGSTs) in the detoxification of the food-derived carcinogen metabolite N-acetoxy-PhIP, and the effect of a polymorphism in hGSTA1 on colorectal cancer risk. Mutat Res. 2001; 482:3–10. [PubMed: 11535243]
42. Girard H, Butler LM, Villeneuve L, Millikan RC, Sinha R, Sandler RS, Guillemette C. UGT1A1 and UGT1A9 functional variants, meat intake, and colon cancer, among Caucasians and African-Americans. Mutat Res. 2008; 644:56–63. [PubMed: 18675828]
43. Lynch AM, Gooderham NJ, Davies DS, Boobis AR. Genetic analysis of PHIP intestinal mutations in MutaMouse. Mutagenesis. 1998; 13:601–605. [PubMed: 9862191]
44. Ishibe N, Freedman AN. Understanding the interaction between environmental exposures and molecular events in colorectal carcinogenesis. Cancer Invest. 2001; 19:524–539. [PubMed: 11458819]
45. Bacolod MD, Barany F. Molecular profiling of colon tumors: the search for clinically relevant biomarkers of progression, prognosis, therapeutics, and predisposition. Ann Surg Oncol. 2011; 18:3694–6700. [PubMed: 21347779]
46. Van Cutsem E, Peeters M, Siena S, Humblet Y, Hendlisz A, Neyns B, et al. Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. J Clin Oncol. 2007; 25:1658–1664. [PubMed: 17470858]
47. Jonker DJ, O'Callaghan CJ, Karapetis CS, Zalcberg JR, Tu D, Au HJ, et al. Cetuximab for the treatment of colorectal cancer. N Engl J Med. 2007; 357:2040–2048. [PubMed: 18003960]
48. Allegra CJ, Jessup JM, Somerfield MR, Hamilton SR, Hammond EH, Hayes DF, McAllister PK, Morton RF, Schilsky RL. American Society of Clinical Oncology provisional clinical opinion: testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. J Clin Oncol. 2009; 27:2091–2096. [PubMed: 19188670]

49. Lievre A, Bachet JB, Le Corre D, Boige V, Landi B, Emile JF, et al. KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. Cancer Res. 2006; 66:3992–3995. [PubMed: 16618717]
50. Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, Juan T, Sikorski R, Suggs S, Radinsky R, Patterson SD, Chang DD. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. J Clin Oncol. 2008; 26:1626–1634. [PubMed: 18316791]
51. Molinari F, Felicioni L, Buscarino M, De Dosso S, Buttitta F, Malatesta S, et al. Increased detection sensitivity for KRAS mutations enhances the prediction of anti-EGFR monoclonal antibody resistance in metastatic colorectal cancer. Clin Cancer Res. 2011; 17:4901–4914. [PubMed: 21632860]
52. Bando H, Yoshino T, Tsuchihara K, Ogasawara N, Fuse N, Kojima T, et al. KRAS mutations detected by the amplification refractory mutation system-Scorpion assays strongly correlate with therapeutic effect of cetuximab. Br J Cancer. 2011; 105:403–406. [PubMed: 21730978]
53. Sartore-Bianchi A, Di Nicolantonio F, Nichelatti M, Molinari F, De Dosso S, Saletti P, et al. Multi-determinants analysis of molecular alterations for predicting clinical benefit to EGFR-targeted monoclonal antibodies in colorectal cancer. PLoS One. 2009; 4:e7287. [PubMed: 19806185]
54. Parsons BL, Myers MB. Personalized cancer treatment and the myth of KRAS wild-type colon tumors. Discovery Medicine. 2013; 15:259–267. [PubMed: 23636143]
55. Misale S, Yaeger R, Hobor S, Scala E, Janakiraman M, Liska D, et al. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. Nature. 2012; 486:532–536. [PubMed: 22722830]
56. Little AS, Balmanno K, Sale MJ, Newman S, Dry JR, Hampson M, Edwards PA, Smith PD, Cook SJ. Amplification of the driving oncogene, KRAS or BRAF, underpins acquired resistance to MEK1/2 inhibitors in colorectal cancer cells. Science Signaling. 2011; 4:ra17. [PubMed: 21447798]
57. Curtin K, Ulrich CM, Samowitz WS, Bigler J, Caan B, Potter JD, Slattery ML. Thymidylate synthase polymorphisms and colon cancer: associations with tumor stage, tumor characteristics and survival. Int J Cancer. 2007; 120:2226–2232. [PubMed: 17290389]
58. Johnston PG, Lenz HJ, Leichman CG, Danenberg KD, Allegra CJ, Danenberg PV, Leichman L. Thymidylate synthase gene and protein expression correlate and are associated with response to 5-fluorouracil in human colorectal and gastric tumors. Cancer Res. 1995; 55:1407–1412. [PubMed: 7882343]
59. Humeniuk R, Menon LG, Mishra PJ, Gorlick R, Sowers R, Rode W, et al. Decreased levels of UMP kinase as a mechanism of fluoropyrimidine resistance. Molecular Cancer Therapeutics. 2009; 8:1037–1044. [PubMed: 19383847]
60. Soong R, Shah N, Salto-Tellez M, Tai BC, Soo RA, Han HC, et al. Prognostic significance of thymidylate synthase, dihydropyrimidine dehydrogenase and thymidine phosphorylase protein expression in colorectal cancer patients treated with or without 5-fluorouracil-based chemotherapy. Ann Oncol. 2008; 19:915–919. [PubMed: 18245778]
61. Kinoshita M, Kodera Y, Hibi K, Nakayama G, Inoue T, Ohashi N, Ito Y, Koike M, Fujiwara M, Nakao A. Gene expression profile of 5-fluorouracil metabolic enzymes in primary colorectal cancer: potential as predictive parameters for response to fluorouracil-based chemotherapy. Anticancer Res. 2007; 27:851–856. [PubMed: 17465211]
62. Palomaki GE, Bradley LA, Douglas MP, Kolor K, Dotson WD. Can UGT1A1 genotyping reduce morbidity and mortality in patients with metastatic colorectal cancer treated with irinotecan? An evidence-based review. Genet Med. 2009; 11:21–34. [PubMed: 19125129]
63. Evaluation of Genomic Applications in P, Prevention Working G. Recommendations from the EGAPP Working Group: can UGT1A1 genotyping reduce morbidity and mortality in patients with metastatic colorectal cancer treated with irinotecan? Genet Med. 2009; 11:15–20. [PubMed: 19125128]
64. Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet. 2005; 6:287–298. [PubMed: 15803198]

65. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. Nat Genet. 2003; 33(Suppl): 238–244. [PubMed: 12610533]

66. Edwards TM, Myers JP. Environmental exposures and gene regulation in disease etiology. Environ Health Perspect. 2007; 115:1264–1270. [PubMed: 17805414]

67. van Steeg H, Kraemer KH. Xeroderma pigmentosum and the role of UV-induced DNA damage in skin cancer. Mol Med Today. 1999; 5:86–94. [PubMed: 10200950]

68. Masutani C, Kusumoto R, Yamada A, Dohmae N, Yokoi M, Yuasa M, Araki M, Iwai S, Takio K, Hanaoka F. The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. Nature. 1999; 399:700–704. [PubMed: 10385124]

69. Lin Z, Zhang X, Tuo J, Guo Y, Green B, Chan CC, Tan W, Huang Y, Ling W, Kadlubar FF, Lin D, Ning B. A variant of the Cockayne syndrome B gene ERCC6 confers risk of lung cancer. Hum Mutat. 2008; 29:113–122. [PubMed: 17854076]

70. Potter JD. Colorectal cancer: molecules and populations. J Natl Cancer Inst. 1999; 91:916–932. [PubMed: 10359544]

71. Olden K, Wilson S. Environmental health and genomics: visions and implications. Nat Rev Genet. 2000; 1:149–153. [PubMed: 11253655]

72. Weinberg RA. The retinoblastoma protein and cell cycle control. Cell. 1995; 81:323–330. [PubMed: 7736585]

73. Baker SJ, Fearon ER, Nigro JM, Hamilton SR, Preisinger AC, Jessup JM, et al. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. Science. 1989; 244:217–221. [PubMed: 2649981]

74. Malkin D, Li FP, Strong LC, Fraumeni JF, Nelson CE, Kim DH, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science. 1990; 250:1233–1238. [PubMed: 1978757]

75. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC. Linkage of early-onset familial breast cancer to chromosome 17q21. Science. 1990; 250:1684–1689. [PubMed: 2270482]

76. Bodmer WF, Bailey CJ, Bodmer J, Bussey HJ, Ellis A, Gorman P, et al. Localization of the gene for familial adenomatous polyposis on chromosome 5. Nature. 1987; 328:614–616. [PubMed: 3039373]

77. Figlioli G, Landi S, Romei C, Elisei R, Gemignani F. Medullary thyroid carcinoma (MTC) and RET proto-oncogene: mutation spectrum in the familial cases and a meta-analysis of studies on the sporadic form. Mutat Res. 2013; 752:36–44. [PubMed: 23059849]

78. Schmidt LS, Nickerson ML, Angeloni D, Glenn GM, Walther MM, Albert PS, et al. Early onset hereditary papillary renal carcinoma: germline missense mutations in the tyrosine kinase domain of the met proto-oncogene. J Urol. 2004; 172:1256–1261. [PubMed: 15371818]

79. Peltomaki P, Aaltonen LA, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, et al. Genetic mapping of a locus predisposing to human colorectal cancer. Science. 1993; 260:810–812. [PubMed: 8484120]

80. Cannon-Albright LA, Goldgar DE, Meyer LJ, Lewis CM, Anderson DE, Fountain JW, et al. Assignment of a locus for familial melanoma, MLM, to chromosome 9p13-p22. Science. 1992; 258:1148–1152. [PubMed: 1439824]

81. Der CJ, Krontiris TG, Cooper GM. Transforming genes of human bladder and lung carcinoma cell lines are homologous to the ras genes of Harvey and Kirsten sarcoma viruses. Proc Natl Acad Sci U S A. 1982; 79:3637–3640. [PubMed: 6285355]

82. Kranenburg O. The KRAS oncogene: past, present, and future. Biochim Biophys Acta. 2005; 1756:81–82. [PubMed: 16269215]

83. Shields PG, Harris CC. Cancer risk and low-penetrance susceptibility genes in gene-environment interactions. J Clin Oncol. 2000; 18:2309–2315. [PubMed: 10829052]

84. Razin A, Riggs AD. DNA methylation and gene function. Science. 1980; 210:604–610. [PubMed: 6254144]

85. Jenuwein T, Allis CD. Translating the histone code. Science. 2001; 293:1074–1080. [PubMed: 11498575]

86. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011; 12:861–874. [PubMed: 22094949]
87. Nystrom M, Mutanen M. Diet and epigenetics in colon cancer. World J Gastroenterol. 2009; 15:257–263. [PubMed: 19140224]
88. Esteller M, Herman JG. Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. J Pathol. 2002; 196:1–7. [PubMed: 11748635]
89. Croce CM. Causes and consequences of microRNA dysregulation in cancer. Nat Rev Genet. 2009; 10:704–714. [PubMed: 19763153]
90. Lapeyre JN, Becker FF. 5-Methylcytosine content of nuclear DNA during chemical hepatocarcinogenesis and in carcinomas which result. Biochem Biophys Res Commun. 1979; 87:698–705. [PubMed: 454420]
91. Smith IM, Mydlarz WK, Mithani SK, Califano JA. DNA global hypomethylation in squamous cell head and neck cancer associated with smoking, alcohol consumption and stage. Int J Cancer. 2007; 121:1724–1728. [PubMed: 17582607]
92. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet. 2005; 37:853–862. [PubMed: 16007088]
93. Oommen AM, Griffin JB, Sarath G, Zempleni J. Roles for nutrients in epigenetic events. J Nutr Biochem. 2005; 16:74–77. [PubMed: 15681164]
94. Baroux C, Pien S, Grossniklaus U. Chromatin modification and remodeling during early seed development. Curr Opin Genet Dev. 2007; 17:473–479. [PubMed: 18029170]
95. Wang GG, Allis CD, Chi P. Chromatin remodeling and cancer, Part I: Covalent histone modifications. Trends Mol Med. 2007; 13:363–372. [PubMed: 17822958]
96. Cohen I, Poreba E, Kamieniarz K, Schneider R. Histone modifiers in cancer: friends or foes? Genes Cancer. 2011; 2:631–647. [PubMed: 21941619]
97. Marks P, Rifkind RA, Richon VM, Breslow R, Miller T, Kelly WK. Histone deacetylases and cancer: causes and therapies. Nat Rev Cancer. 2001; 1:194–202. [PubMed: 11902574]
98. Mattick JS, Makunin IV. Non-coding RNA. Hum Mol Genet. 2006; 15(Spec1):R17–29. [PubMed: 16651366]
99. Sneider TW, Teague WM, Rogachevsky LM. S-adenosylmethionine: DNA-cytosine 5-methyltransferase from a Novikoff rat hepatoma cell line. Nucleic Acids Res. 1975; 2:1685–1700. [PubMed: 171625]
100. Stover PJ, Garza C. Bringing individuality to public health recommendations. J Nutr. 2002; 132:2476S–2480S. [PubMed: 12163715]
101. Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA. Microarrays and toxicology: the advent of toxicogenomics. Mol Carcinog. 1999; 24:153–159. [PubMed: 10204799]
102. Waters MD, Fostel JM. Toxicogenomics and systems toxicology: aims and prospects. Nat Rev Genet. 2004; 5:936–948. [PubMed: 15573125]
103. Casciano DA, Woodcock J. Empowering microarrays in the regulatory setting. Nat Biotechnol. 2006; 24:1103. [PubMed: 16964221]
104. McHale CM, Zhang L, Thomas R, Smith MT. Analysis of the transcriptome in molecular epidemiology studies. Environ Mol Mutagen. 2013; 54:500–517. [PubMed: 23907930]
105. Su Z, Li Z, Chen T, Li QZ, Fang H, Ding D, et al. Comparing next-generation sequencing and microarray technologies in a toxicological study of the effects of aristolochic Acid on rat kidneys. Chem Res Toxicol. 2011; 24:1486–1493. [PubMed: 21834575]
106. Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, et al. Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq. Cancer Prev Res (Phila). 2011; 4:803–817. [PubMed: 21636547]
107. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature. 2010; 463:184–190. [PubMed: 20016488]

108. Besaratinia A, Li H, Yoon JI, Zheng A, Gao H, Tommasi S. A high-throughput next-generation sequencing-based method for detecting the mutational fingerprint of carcinogens. Nucleic Acids Res. 2012; 40:e116. [PubMed: 22735701]
109. Grosse Y, Baan R, Straif K, Secretan B, El Ghissassi F, Bouvard V, Benbrahim-Tallaa L, Guha N, Galichet L, Cogliano V. A review of human carcinogens-Part A: pharmaceuticals. Lancet Oncol. 2009; 10:13–14. [PubMed: 19115512]
110. Mei N, Arlt VM, Phillips DH, Heflich RH, Chen T. DNA adduct formation and mutation induction by aristolochic acid in rat kidney and liver. Mutat Res. 2006; 602:83–91. [PubMed: 17010389]
111. McDaniel LP, Elander ER, Guo X, Chen T, Arlt VM, Mei N. Mutagenicity and DNA adduct formation by aristolochic acid in the spleen of Big Blue(R) rats. Environ Mol Mutagen. 2012; 53:358–368. [PubMed: 22508110]
112. Chen L, Mei N, Yao L, Chen T. Mutations induced by carcinogenic doses of aristolochic acid in kidney of Big Blue transgenic rats. Toxicol Lett. 2006; 165:250–256. [PubMed: 16764999]
113. Wang Y, Meng F, Arlt VM, Mei N, Chen T, Parsons BL. Aristolochic acid-induced carcinogenesis examined by ACB-PCR quantification of H-Ras and K-Ras mutant fraction. Mutagenesis. 2011; 26:619–628. [PubMed: 21642617]
114. Chen T, Guo L, Zhang L, Shi L, Fang H, Sun Y, Fuscoe JC, Mei N. Gene expression profiles distinguish the carcinogenic effects of aristolochic acid in target (kidney) and non-target (liver) tissues in rats. BMC Bioinformatics. 2006; 7(Suppl 2):S20.
115. Husgafvel-Pursiainen K. Genotoxicity of environmental tobacco smoke: a review. Mutat Res. 2004; 567:427–445. [PubMed: 15572289]
116. Andreoli C, Gigante D, Nunziata A. A review of in vitro methods to assess the biological activity of tobacco smoke with the aim of reducing the toxicity of smoke. Toxicol in Vitro. 2003; 17:587–594. [PubMed: 14599449]
117. World Health Organization. WHO Report on the Global Tobacco Epidemic. World Health Organization. 2009. http://whqlibdoc.who.int/publications/2009/9789241563918engfull.pdf
118. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. Cell. 2013; 155:27–38. [PubMed: 24074859]
119. Su Z, Ning B, Fang H, Hong H, Perkins R, Tong W, Shi L. Next-generation sequencing and its applications in molecular diagnostics. Expert Review of Molecular Diagnostics. 2011; 11:333–343. [PubMed: 21463242]
120. Eisenstein M. Oxford Nanopore announcement sets sequencing sector abuzz. Nat Biotechnol. 2012; 30:295–296. [PubMed: 22491260]
121. Su, Z., Ning, B., Fang, H., Hong, H., Perkins, R., Tong, W., Shi, L. General, Applied and Systems Toxicology. John Wiley & Sons; West Sussex, UK: 2009. Next-Generation Sequencing: A Revolutionary Tool for Toxicogenomics.
122. Metzker ML. Sequencing technologies—the next generation. Nat Rev Genet. 2010; 11:31–46. [PubMed: 19997069]
123. Korlach J, Marks PJ, Cicero RL, Gray JJ, Murphy DL, Roitman DB, Pham TT, Otto GA, Foquet M, Turner SW. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. Proc Natl Acad Sci U S A. 2008; 105:1176–1181. [PubMed: 18216253]
124. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature Methods. 2013; 10:563–569. [PubMed: 23644548]
125. Koren S, Harhay GP, Smith TP, Bono JL, Harhay DM, McVey SD, Radune D, Bergman NH, Phillippy AM. Reducing assembly complexity of microbial genomes with single-molecule sequencing. Genome Biology. 2013; 14:R101. [PubMed: 24034426]
126. Sharon D, Tilgner H, Grubert F, Snyder M. A single-molecule long-read survey of the human transcriptome. Nat Biotechnol. 2013; 31:1009–1014. [PubMed: 24108091]

127. Guo X, Zheng S, Dang H, Pace RG, Stonebraker JR, Jones CD, et al. Genome Reference and Sequence Variation in the Large Repetitive Central Exon of Human MUC5AC. Am J Respir Cell Mol Biol. 2014; 50(1):223–232. [PubMed: 24010879]
128. Carneiro MO, Russ C, Ross MG, Gabriel SB, Nusbaum C, DePristo MA. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. BMC Genomics. 2012; 13:375. [PubMed: 22863213]
129. Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, et al. Genome-wide mapping of methylated adenine residues in pathogenic Escherichia coli using single-molecule real-time sequencing. Nat Biotechnol. 2012; 30:1232–1239. [PubMed: 23138224]
130. Smith CC, Wang Q, Chin CS, Salerno S, Damon LE, Levis MJ, et al. Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukaemia. Nature. 2012; 485:260–263. [PubMed: 22504184]
131. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012; 13:341. [PubMed: 22827831]
132. Nguyen-Dumont T, Teo ZL, Pope BJ, Hammet F, Mahmoodi M, Tsimiklis HS, et al. Hi-Plex for high-throughput mutation screening: application to the breast cancer susceptibility gene PALB2. BMC Medical Genomics. 2013; 6:48. [PubMed: 24206657]
133. Rutvisuttinunt W, Chinnawirotpisan P, Simasathien S, Shrestha SK, Yoon IK, Klungthong C, Fernandez S. Simultaneous and complete genome sequencing of influenza A and B with high coverage by Illumina MiSeq Platform. J Virol Methods. 2013; 193:394–404. [PubMed: 23856301]
134. Williams ST, Foster PG, Littlewood DT. The complete mitochondrial genome of a turbinid vetigastropod from MiSeq Illumina sequencing of genomic DNA and steps towards a resolved gastropod phylogeny. Gene. 2014; 533:38–47. [PubMed: 24120625]
135. Cheng CS, Rai K, Garber M, Hollinger A, Robbins D, Anderson S, et al. Semiconductor-based DNA sequencing of histone modification states. Nature Communications. 2013; 4:2672.
136. Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. PLoS Computational Biology. 2013; 9:e1003031. [PubMed: 23592973]
137. Stoddart D, Heron AJ, Mikhailova E, Maglia G, Bayley H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. Proc Natl Acad Sci U S A. 2009; 106:7702–7707. [PubMed: 19380741]
138. Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW, Robinson ER, Pallen MJ. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat Rev Microbiol. 2012; 10:599–606. [PubMed: 22864262]
139. Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM, et al. De novo genome sequencing and comparative genomics of date palm (Phoenix dactylifera). Nat Biotechnol. 2011; 29:521–527. [PubMed: 21623354]
140. Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, Jin X, et al. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell. 2012; 151:1431–1442. [PubMed: 23260136]
141. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. Nature. 2012; 489:101–108. [PubMed: 22955620]
142. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012; 489:91–100. [PubMed: 22955619]
143. Li M, Wang IX, Li Y, Bruzel A, Richards AL, Toung JM, Cheung VG. Widespread RNA and DNA sequence differences in the human transcriptome. Science. 2011; 333:53–58. [PubMed: 21596952]
144. Kleinman CL, Majewski J. Comment on “Widespread RNA and DNA sequence differences in the human transcriptome”. Science. 2012; 335:1302. author reply 02.

145. Lin W, Piskol R, Tan MH, Li JB. Comment on “Widespread RNA and DNA sequence differences in the human transcriptome”. Science. 2012; 335:1302. author reply 02.

146. Pickrell JK, Gilad Y, Pritchard JK. Comment on “Widespread RNA and DNA sequence differences in the human transcriptome”. Science. 2012; 335:1302. author reply 02.

147. Schrider DR, Gout JF, Hahn MW. Very few RNA and DNA sequence differences in the human transcriptome. PLoS One. 2011; 6:e25842. [PubMed: 22022455]

148. Wong KM, Hudson TJ, McPherson JD. Unraveling the genetics of cancer: genome sequencing and beyond. Annu Rev Genomics Hum Genet. 2011; 12:407–430. [PubMed: 21639794]

149. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature. 2012; 486:400–404. [PubMed: 22722201]

150. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; 499:214–218. [PubMed: 23770567]

151. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell. 2012; 150:1121–1134. [PubMed: 22980976]

152. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science. 2013; 339:1546–1558. [PubMed: 23539594]

153. Watson IR, Takahashi K, Futreal PA, Chin L. Emerging patterns of somatic mutations in cancer. Nat Rev Genet. 2013; 14:703–718. [PubMed: 24022702]

154. Yoshida K, Sanada M, Ogawa S. Deep sequencing in cancer research. Jpn J Clin Oncol. 2013; 43:110–115. [PubMed: 23225907]

155. Robinson DR, Kalyana-Sundaram S, Wu YM, Shankar S, Cao X, Ateeq B, et al. Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. Nat Med. 2011; 17:1646–1651. [PubMed: 22101766]

156. Wang R, Hu H, Pan Y, Li Y, Ye T, Li C, et al. RET fusions define a unique molecular and clinicopathologic subtype of non-small-cell lung cancer. J Clin Oncol. 2012; 30:4352–4359. [PubMed: 23150706]

157. Pflueger D, Terry S, Sboner A, Habegger L, Esgueva R, Lin PC, et al. Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing. Genome Res. 2011; 21:56–67. [PubMed: 21036922]

158. Singh D, Chan JM, Zoppoli P, Niola F, Sullivan R, Castano A, et al. Transforming fusions of FGFR and TACC genes in human glioblastoma. Science. 2012; 337:1231–1235. [PubMed: 22837387]

159. Shah N, Lankerovich M, Lee H, Yoon JG, Schroeder B, Foltz G. Exploration of the gene fusion landscape of glioblastoma using transcriptome sequencing and copy number data. BMC Genomics. 2013; 14:818. [PubMed: 24261984]

160. Horvath A, Pakala SB, Mudvari P, Reddy SD, Ohshiro K, Casimiro S, et al. Novel insights into breast cancer genetic variance through RNA sequencing. Sci Rep. 2013; 3:2256. [PubMed: 23884293]

161. Meaburn E, Schulz R. Next generation sequencing in epigenetics: insights and challenges. Semin Cell Dev Biol. 2012; 23:192–199. [PubMed: 22027613]

162. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. Nat Rev Cancer. 2011; 11:726–734. [PubMed: 21941284]

163. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009; 462:315–322. [PubMed: 19829295]

164. Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, Batzoglou S, Myers RM, Sidow A. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nature Methods. 2008; 5:829–834. [PubMed: 19160518]

165. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013; 368:2059–2074. [PubMed: 23634996]

166. Li JL, Mazar J, Zhong C, Faulkner GJ, Govindarajan SS, Zhang Z, et al. Genome-wide methylated CpG island profiles of melanoma cells reveal a melanoma coregulation network. Sci Rep. 2013; 3:2962. [PubMed: 24129253]

167. McLemore TL, Adelberg S, Liu MC, McMahon NA, Yu SJ, Hubbard WC, et al. Expression of CYP1A1 gene in patients with lung cancer: evidence for cigarette smoke-induced gene expression in normal lung tissue and for altered gene regulation in primary pulmonary carcinomas. J Natl Cancer Inst. 1990; 82:1333–1339. [PubMed: 2380990]

168. Murakami Y, Saigo K, Takashima H, Minami M, Okanoue T, Brechot C, Paterlini-Brechot P. Large scaled analysis of hepatitis B virus (HBV) DNA integration in HBV related hepatocellular carcinomas. Gut. 2005; 54:1162–1168. [PubMed: 16009689]

169. Bressac B, Kew M, Wands J, Ozturk M. Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. Nature. 1991; 350:429–431. [PubMed: 1672732]

170. Bell DA, Taylor JA, Paulson DF, Robertson CN, Mohler JL, Lucier GW. Genetic risk and carcinogen exposure: a common inherited defect of the carcinogen-metabolism gene glutathione S-transferase M1 (GSTM1) that increases susceptibility to bladder cancer. J Natl Cancer Inst. 1993; 85:1159–1164. [PubMed: 8320745]

171. Marcus PM, Hayes RB, Vineis P, Garcia-Closas M, Caporaso NE, Autrup H, et al. Cigarette smoking, N-acetyltransferase 2 acetylation status, and bladder cancer risk: a case-series meta-analysis of a gene-environment interaction. Cancer Epidemiol Biomarkers Prev. 2000; 9:461–467. [PubMed: 10815690]

172. Chen J, Giovannucci E, Kelsey K, Rimm EB, Stampfer MJ, Colditz GA, Spiegel-man D, Willett WC, Hunter DJ. A methylenetetrahydrofolate reductase polymorphism and the risk of colorectal cancer. Cancer Res. 1996; 56:4862–4864. [PubMed: 8895734]

173. Ma J, Stampfer MJ, Giovannucci E, Artigas C, Hunter DJ, Fuchs C, Willett WC, Selhub J, Hennekens CH, Rozen R. Methylenetetrahydrofolate reductase polymorphism, dietary interactions, and risk of colorectal cancer. Cancer Res. 1997; 57:1098–1102. [PubMed: 9067278]

174. Chen J, Giovannucci E, Hankinson SE, Ma J, Willett WC, Spiegelman D, Kelsey KT, Hunter DJ. A prospective study of methylenetetrahydrofolate reductase and methionine synthase gene polymorphisms, and risk of colorectal adenoma. Carcinogenesis. 1998; 19:2129–2132. [PubMed: 9886567]

175. Kadlubar FF, Butler MA, Kaderlik KR, Chou HC, Lang NP. Polymorphisms for aromatic amine metabolism in humans: relevance for human carcinogenesis. Environ Health Perspect. 1992; 98:69–74. [PubMed: 1486865]

176. Rampersaud GC, Kauwell GP, Hutson AD, Cerda JJ, Bailey LB. Genomic DNA methylation decreases in response to moderate folate depletion in elderly women. Am J Clin Nutr. 2000; 72:998–1003. [PubMed: 11010943]

177. Angrisano T, Sacchetti S, Natale F, Cerrato A, Pero R, Keller S, et al. Chromatin and DNA methylation dynamics during retinoic acid-induced RET gene transcriptional activation in neuroblastoma cells. Nucleic Acids Res. 2011; 39:1993–2006. [PubMed: 20952403]

178. Brunaud L, Alberto JM, Ayav A, Gerard P, Namour F, Antunes L, Braun M, Bronowicki JP, Bresler L, Gueant JL. Effects of vitamin B12 and folate deficiencies on DNA methylation and carcinogenesis in rat liver. Clin Chem Lab Med. 2003; 41:1012–1019. [PubMed: 12964806]

179. Zhu H, Wang X, Shi H, Su S, Harshfield GA, Gutin B, Snieder H, Dong Y. A genome-wide methylation study of severe vitamin D deficiency in African American adolescents. J Pediatr. 2013; 162:1004–9. e1. [PubMed: 23219444]

180. Davis CD, Uthus EO, Finley JW. Dietary selenium and arsenic affect DNA methylation in vitro in Caco-2 cells and in vivo in rat liver and colon. J Nutr. 2000; 130:2903–2909. [PubMed: 11110844]

181. Fang M, Chen D, Yang CS. Dietary polyphenols may affect DNA methylation. J Nutr. 2007; 137:223S–228S. [PubMed: 17182830]

182. Dobosy JR, Fu VX, Desotelle JA, Srinivasan R, Kenowski ML, Almassi N, Weindruch R, Svaren J, Jarrard DF. A methyl-deficient diet modifies histone methylation and alters Igf2 and H19 repression in the prostate. Prostate. 2008; 68:1187–1195. [PubMed: 18459101]

183. Shen J, Montecino M, Lian JB, Stein GS, Van Wijnen AJ, Stein JL. Histone acetylation in vivo at the osteocalcin locus is functionally linked to vitamin D-dependent, bone tissue-specific transcription. J Biol Chem. 2002; 277:20284–20292. [PubMed: 11893738]

184. Xiang N, Zhao R, Song G, Zhong W. Selenite reactivates silenced genes by modifying DNA methylation and histones in prostate cancer cells. Carcinogenesis. 2008; 29:2175–2181. [PubMed: 18676679]

185. Marcu MG, Jung YJ, Lee S, Chung EJ, Lee MJ, Trepel J, Neckers L. Curcumin is an inhibitor of p300 histone acetylatransferase. Med Chem. 2006; 2:169–174. [PubMed: 16787365]

186. Kutay H, Bai S, Datta J, Motiwala T, Pogribny I, Frankel W, Jacob ST, Ghoshal K. Downregulation of miR-122 in the rodent and human hepatocellular carcinomas. J Cell Biochem. 2006; 99:671–678. [PubMed: 16924677]

187. Garzon R, Pichiorri F, Palumbo T, Visentini M, Aqeilan R, Cimmino A, et al. MicroRNA gene expression during retinoic acid-induced differentiation of human acute promyelocytic leukemia. Oncogene. 2007; 26:4148–4157. [PubMed: 17260024]

188. Stone N, Pangilinan F, Molloy AM, Shane B, Scott JM, Ueland PM, Mills JL, Kirke PN, Sethupathy P, Brody LC. Bioinformatic and genetic association analysis of microRNA target sites in one-carbon metabolism genes. PLoS One. 2011; 6:e21851. [PubMed: 21765920]

189. Alvarez-Diaz S, Valle N, Ferrer-Mayorga G, Lombardia L, Herrera M, Dominguez O, Segura MF, Bonilla F, Hernando E, Munoz A. MicroRNA-22 is induced by vitamin D and contributes to its antiproliferative, antimigratory and gene regulatory effects in colon cancer cells. Hum Mol Genet. 2012; 21:2157–2165. [PubMed: 22328083]

190. Maciel-Dominguez A, Swan D, Ford D, Hesketh J. Selenium alters miRNA profile in an intestinal cell line: Evidence that miR-185 regulates expression of GPX2 and SEPSH2. Mol Nutr Food Res. 2013; 57(12):2192–2205.

191. Ali S, Ahmad A, Banerjee S, Padhye S, Dominiak K, Schaffert JM, Wang Z, Philip PA, Sarkar FH. Gemcitabine sensitivity can be induced in pancreatic cancer cells through modulation of miR-200 and miR-21 expression by curcumin or its analogue CDF. Cancer Res. 2010; 70:3606–3617. [PubMed: 20388782]

192. Lee KW, Pausova Z. Cigarette smoking and DNA methylation. Front Genet. 2013; 4:132. [PubMed: 23882278]

193. Takiguchi M, Achanzar WE, Qu W, Li G, Waalkes MP. Effects of cadmium on DNA-(Cytosine-5) methyltransferase activity and DNA methylation status during cadmium-induced cellular transformation. Exp Cell Res. 2003; 286:355–365. [PubMed: 12749863]

194. Ji W, Yang L, Yu L, Yuan J, Hu D, Zhang W, et al. Epigenetic silencing of O6-methylguanine DNA methyltransferase gene in NiS-transformed cells. Carcinogenesis. 2008; 29:1267–1275. [PubMed: 18204074]

195. Bose R, Onishchenko N, Edoff K, Janson Lang AM, Ceccatelli S. Inherited effects of low-dose exposure to methylmercury in neural stem cells. Toxicol Sci. 2012; 130:383–390. [PubMed: 22918959]

196. Bihaqi SW, Huang H, Wu J, Zawia NH. Infant exposure to lead (Pb) and epigenetic modifications in the aging primate brain: implications for Alzheimer's disease. J Alzheimers Dis. 2011; 27:819–833. [PubMed: 21891863]

197. Pilsner JR, Liu X, Ahsan H, Ilievski V, Slavkovich V, Levy D, Factor-Litvak P, Graziano JH, Gamble MV. Genomic methylation of peripheral blood leukocyte DNA: influences of arsenic and folate in Bangladeshi adults. Am J Clin Nutr. 2007; 86:1179–1186. [PubMed: 17921400]

198. Tao L, Ge R, Xie M, Kramer PM, Pereira MA. Effect of trichloroethylene on DNA methylation and expression of early-intermediate protooncogenes in the liver of B6C3F1 mice. J Biochem Mol Toxicol. 1999; 13:231–237. [PubMed: 10402556]

199. De Prins S, Koppen G, Jacobs G, Dons E, Van de Mieroop E, Nelen V, et al. Influence of ambient air pollution on global DNA methylation in healthy adults: a seasonal follow-up. Environ Int. 2013; 59:418–424. [PubMed: 23917442]

200. Bollati V, Baccarelli A, Hou L, Bonzini M, Fustinoni S, Cavallo D, et al. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. Cancer Res. 2007; 67:876–880. [PubMed: 17283117]

201. Dolinoy DC, Huang D, Jirtle RL. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. Proc Natl Acad Sci U S A. 2007; 104:13056–13061. [PubMed: 17670942]

202. Ito K, Lim S, Caramori G, Chung KF, Barnes PJ, Adcock IM. Cigarette smoking reduces histone deacetylase 2 expression, enhances cytokine expression, and inhibits glucocorticoid actions in alveolar macrophages. FASEB J. 2001; 15:1110–1112. [PubMed: 11292684]

203. Ke Q, Li Q, Ellen TP, Sun H, Costa M. Nickel compounds induce phosphorylation of histone H3 at serine 10 by activating JNK-MAPK pathway. Carcinogenesis. 2008; 29:1276–1281. [PubMed: 18375956]

204. Ke Q, Davidson T, Chen H, Kluz T, Costa M. Alterations of histone modifications and transgene silencing by nickel chloride. Carcinogenesis. 2006; 27:1481–1488. [PubMed: 16522665]

205. Gadhia SR, Calabro AR, Barile FA. Trace metals alter DNA repair and histone modification pathways concurrently in mouse embryonic stem cells. Toxicol Lett. 2012; 212:169–179. [PubMed: 22641096]

206. Zhou X, Sun H, Ellen TP, Chen H, Costa M. Arsenite alters global histone H3 methylation. Carcinogenesis. 2008; 29:1831–1836. [PubMed: 18321869]

207. Cantone L, Nordio F, Hou L, Apostoli P, Bonzini M, Tarantini L, et al. Inhalable metal-rich air particles and histone H3K4 dimethylation and H3K9 acetylation in a cross-sectional study of steel workers. Environ Health Perspect. 2011; 119:964–969. [PubMed: 21385672]

208. Ishihama M, Toyooka T, Ibuki Y. Generation of phosphorylated histone H2AX by benzene metabolites. Toxicol in Vitro. 2008; 22:1861–1868. [PubMed: 18835433]

209. Doherty LF, Bromer JG, Zhou Y, Aldad TS, Taylor HS. In utero exposure to diethylstilbestrol (DES) or bisphenol-A (BPA) increases EZH2 expression in the mammary gland: an epigenetic mechanism linking endocrine disruptors to breast cancer. Horm Cancer. 2010; 1:146–155. [PubMed: 21761357]

210. Schembri F, Sridhar S, Perdomo C, Gustafson AM, Zhang X, Ergun A, et al. MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. Proc Natl Acad Sci U S A. 2009; 106:2319–2324. [PubMed: 19168627]

211. Bollati V, Marinelli B, Apostoli P, Bonzini M, Nordio F, Hoxha M, et al. Exposure to metal-rich particulate matter modifies the expression of candidate microRNAs in peripheral blood leukocytes. Environ Health Perspect. 2010; 118:763–768. [PubMed: 20061215]

212. Ji W, Yang L, Yuan J, Zhang M, Qi D, Duan X, et al. MicroRNA-152 targets DNA methyltransferase 1 in NiS-transformed cells via a feedback mechanism. Carcinogene-sis. 2013; 34:446–453.

213. Pallocca G, Fabbri M, Sacco MG, Gribaldo L, Pamies D, Laurenza I, Bal-Price A. miRNA expression profiling in a human stem cell-based model as a tool for developmental neurotoxicity testing. Cell Biol Toxicol. 2013; 29:239–257. [PubMed: 23903816]

214. Martinez-Pacheco M, Hidalgo-Miranda A, Romero-Cordoba S, Valverde M, Rojas E. mRNA and miRNA expression patterns associated to pathways linked to metal mixture health effects. Gene. 2014; 533:508–514. [PubMed: 24080485]

215. Cao Y, Yu SL, Wang Y, Guo GY, Ding Q, An RH. MicroRNA-dependent regulation of PTEN after arsenic trioxide treatment in bladder cancer cell line T24. Tumour Biol. 2011; 32:179–188. [PubMed: 20857258]

216. Zhang L, McHale CM, Rothman N, Li G, Ji Z, Vermeulen R, et al. Systems biology of human benzene exposure. Chem Biol Interact. 2010; 184:86–93. [PubMed: 20026094]

217. Avissar-Whiting M, Veiga KR, Uhl KM, Maccani MA, Gagne LA, Moen EL, Marsit CJ. Bisphenol A exposure leads to specific microRNA alterations in placental cells. Reprod Toxicol. 2010; 29:401–406. [PubMed: 20417706]

218. Chmielecki J, Crago AM, Rosenberg M, O'Connor R, Walker SR, Ambrogio L, et al. Whole-exome sequencing identifies a recurrent NAB2-STAT6 fusion in solitary fibrous tumors. Nat Genet. 2013; 45:131–132. [PubMed: 23313954]

219. Tenedini E, Bernardis I, Artusi V, Artuso L, Roncaglia E, Guglielmelli P, et al. Targeted cancer exome sequencing reveals recurrent mutations in myeloproliferative neoplasms. Leukemia. 2013; doi: 10.1038/leu.2013.302

220. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature Methods. 2013; 10:57–59. [PubMed: 23202435]

221. Lilljebjorn H, Agerstam H, Orsmark-Pietras C, Rissler M, Ehrencrona H, Nilsson L, Richter J, Fioretos T. RNA-seq identifies clinically relevant fusion genes in leukemia including a novel MEF2D/CSF1R fusion responsive to imatinib. Leukemia. 2013; doi: 10.1038/leu.2013.324

222. Smith RM, Webb A, Papp AC, Newman LC, Handelman SK, Suhy A, Mascarenhas R, Oberdick J, Sadee W. Whole transcriptome RNA-Seq allelic expression in human brain. BMC Genomics. 2013; 14:571. [PubMed: 23968248]

223. Rizzo JM, Bard JE, Buck MJ. Standardized collection of MNase-seq experiments enables unbiased dataset comparisons. BMC Molecular Biology. 2012; 13:15. [PubMed: 22559821]

224. Winter DR, Song L, Mukherjee S, Furey TS, Crawford GE. DNase-seq predicts regions of rotational nucleosome stability across diverse human cell types. Genome Res. 2013; 23:1118–1129. [PubMed: 23657885]

Table 1

Examples of Genes-Environment Interaction in Cancer Development

Cancer	Gene	Environment Factor	Reference
Melanoma	XP-V, NER, CS	Ultraviolet Radiation	Van Steeg 1999 [67]; Masutani 1999 [68]
Lung Cancer	CYP1A1, ERCC6	Cigarette Smoking	McLemore 1990 [167]; Lin 2008 [69]
Liver Cancer	TERT, MAPK1, p53	Viral Infection AflatoxinB1	Murakami 2005 [168]; Bressac 1991 [169]
Bladder Cancer	GSTM1 NAT2	PAH Cigarette Smoking	Bell 1993 [170] Marcus 2000 [171]
Colorectal Cancer	MTHFR NAT1, NAT2, CYP1A2	Folate, VitB12, Red Meat	Chen 1996 [172]; Ma 1997 [173] Chen 1998 [174]; Kadlubar 1992 [175]

Table 2

Examples of Interaction between Environmental Exposure and Epigenetic Alteration

	DNA Methylation	Histone Modification	ncRNA Expression
Food Constituents			
Folate	Affected [176]	Affected [182]	Affected [186]
Retinoic Acid	Affected[177]	Affected [177]	Affected [187]
Vitamin B-12	Affected [178]	Affected [100]	Affected [188]
Vitamin D	Affected [179]	Affected [183]	Affected [189]
Selenium	Affected [180]	Affected [184]	Affected [190]
Polyphenols	Affected [181]	Affected [185]	Affected [191]
Pollutants			
Cigarette smoke	Affected [192]	Reduced [202]	Affected [210]
Heavy metals			
Cadmium	Decreased [193]	Reduced Acetylation [194] Increased Phosphorylation [203]	Affected [211]
Nickel	Increased [194]	Decreased Acetylation [204] Increased Demethylation [204]	Affected [212]
Mercury	Globally Decreased [195]	Reduced Methylation [205]	Affected [213]
Lead	Decreased [196]	Reduced [196]	Affected by Pb Mix. [214]
Arsenic	Globally Increased [197]	Globally Affected [206]	Affected [215]
Disinfection byproducts (TCE, DCA, TCA)	Increased [198]	Unknown	Unknown
Air pollution	Globally Decreased [199]	Affected [207]	Affected [211]
Other Chemicals			
Benzene	Affected [200]	Increased Phosphorylation [208]	Affected [216]
Bisphenol A	Globally Decreased [201]	Increased Trimethylation [209]	Affected [217]

Table 3

Summary of Next-Generation Sequencing Platforms

Company	Sequencing reaction	Amplification	Sequencer	Read length (base)	Number of reads (M/run)	Run time (day)	Throughput (Gb/day)	Error type	Comments
Illumina	SBS * with RT‡	Bridge PCR±	HiSeq 1500/1000 high output	36–100	1500–3000	2–8.5	23–27	Substitution	Currently the most widely used platforms with uitra high throughput, without scalability as SOLiD 5500×1
			HiSeq 2500/2000 high output	36–100	3000–6000	2–11	49–55		
			HiSeq 1500 rapid run	36–150	300–600	0.3–1.7	30–54		
			HiSeq 2500 rapid run	36–150	600–1200	0.3–1.7	60–108		
			HiScanSQ	35–100	750–1500	1.5–8.5	15–18		Dual platform: microarray and NGS Fast, long reads, automated workflow, highly scalable
			MiSeq	25–300	22–50	0.17–2.7	3.3–5.5		
			GA-llx	35–150	320–640	2–14	5–6.8		Legacy platform
Life Technologies	SBL †	Emulsion PCR	SOLID	75/2×50	3200	1	20–45	Substitution	ECC-based error detection and correction, scalable runs for each lane on flow cell
			5500×1 W	MP/50×50 PE					
	SBS with H + detection		Ion PGM 314 Chip	200/400	0.4–0.55	0.1–0.15	0.31–0.65	Indel	Label-free chemistry, cheap, run fast, highly scalable, long reads.
			Ion PGM 316 Chip	200/400	2–3	0.12–0.2	2.4–4.9		
			Ion PGM 318 Chip	200/400	4–5	0.18–0.3	3.3–6.6		
			Ion Proton	<200	60–80	0.08–0.17	60–120		
Roche/454	Pyrosequencing	Emulsion PCR	GS Junior	~400	0.07/0.1	0.42	0.035		Long reads and short run time, high

J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. Author manuscript; available in PMC 2017 December 03.

Company	Sequencing reaction	Amplification	Sequencer	Read length (base)	Number of reads (M/run)	Run time (day)	Throughput (Gb/day)	Error type	Comments
									reagent cost and high error rate in homopolymer repeats
			GS FLX	~700	1	0.96	0.7	Indel	
			Titanium XL + GS FLX	~450	1	0.42	1		
Pacific Bioscience	SBS, SMRT[λ]	None	Titanium XLR70 PacBio RS II	~5000	~0.05	<0.1	2.2		Long reads, single-molecular sequencing, less PCR bids
Oxford Nanopore Technologies Ltd.	Strand sequencing	None	MinlON					Deletion	Long reads, single-molecular DNA sequencing. RNA and protein sequencing are under development
			GridlON						

[*] SBS, sequencing by synthesis;

[ǂ] RT, reversible terminators;

[±] PCR, polymarase chain reaction;

[†] SBL, sequencing by ligation;

[λ] SMRT, single molecule real time sequencing.

J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. Author manuscript; available in PMC 2017 December 03.

Table 4

Next-Generation Sequencing-Based Approaches and Their Applications

Sequencing Approaches	Strategy	Feature	Case Study
Whole genome sequencing (WGS)	Sequencing DNA fragments from a whole genome	SNVs, Indels, structural rearrangement, and copy number variations s can be obtained in a single assay with single-base resolution; high cost for data generation, storage and analysis, low coverage.	Whole genome resequencing [140], de novo genome sequencing, and assembly [139]
Targeted resequencing	Sequencing DNA fragments from captured genome regions	Low cost, easy to get deep coverage for exomes, specific genes, or other genomic regions of interest; Need isolation of genomic regions of interest, only targeted regions can be seen	Whole exome sequencing (WES) [218], a targeted panel of selected genes [219], amplicon sequencing [220]
RNA-Seq	Sequencing RNA/cDNA fragments	High dynamic range, enable to identify novel genes and transcripts, gene fusion events; High cost, lack of standards for data analysis	Gene expression quantification [105], transcriptome annotation [126], gene fusion detection [221], allele specific expression [222]
ChIP-Seq	Combination of chromatin immunoprecipitation (ChIP) with NGS	High resolution and sensitivity, but the cost is correlated with sensitivity and resolution	Exploring interactions between protein, DNA and RNA [135]
MethyIC-Seq (BS-Seq)	Sequencing DNA fragments from bisulphite treated genome	Genome-wide survey	Genome-wide methylation characterization [129]
MNase-Seq	Sequencing DNA fragments from micrococcal nuclease digestion	Genome-wide survey	Detection of nucleosome localization [223]
DNase I-Seq	Sequencing DNA fragments from DNase I digestion	Genome-wide survey	Detection of chromatin accessibility [224]