Title
Analysis of the transcriptome in molecular epidemiology studies

Journal
Environmental and Molecular Mutagenesis, 54(7)

ISSN
0893-6692

Authors
McHale, Cliona M
Zhang, Luoping
Thomas, Reuben
et al.

Publication Date
2013-08-01

DOI
10.1002/em.21798

Peer reviewed

Published in final edited form as:

Environ Mol Mutagen. 2013 August ; 54(7): 500–517. doi:10.1002/em.21798.

Analysis of the transcriptome in molecular epidemiology studies

Cliona M. McHale*, **Luoping Zhang**, **Reuben Thomas**, and **Martyn T. smith**

Genes and Environment Laboratory, School of Public Health, Division of Environmental Health Sciences, University of California, Berkeley, CA 94720

Abstract

The human transcriptome is complex, comprising multiple transcript types, mostly in the form of non-coding RNA (ncRNA). The majority of ncRNA is of the long form (lncRNA, ≥200bp), which plays an important role in gene regulation through multiple mechanisms including epigenetics, chromatin modification, control of transcription factor binding, and regulation of alternative splicing. Both mRNA and ncRNA exhibit additional variability in the form of alternative splicing and RNA editing. All aspects of the human transcriptome can potentially be dysregulated by environmental exposures. Next-generation RNA sequencing (RNA-Seq) is the best available methodology to measure this although it has limitations, including experimental bias. The third phase of the MicroArray Quality Control Consortium project (MAQC-III), also called Sequencing Quality Control (SeQC), aims to address these limitations through standardization of experimental and bioinformatic methodologies. A limited number of toxicogenomic studies have been conducted to date using RNA-Seq. This review describes the complexity of the human transcriptome, the application of transcriptomics by RNA-Seq or microarray in molecular epidemiology studies, and limitations of these approaches including the type of cell or tissue analyzed, experimental variation, and confounding. By using good study designs with precise, individual exposure measurements, sufficient power and incorporation of phenotypic anchors, studies in human populations can identify biomarkers of exposure and/or early effect and elucidate mechanisms of action underlying associated diseases, even at low doses. Analysis of datasets at the pathway level can compensate for some of the limitations of RNA-Seq and, as more datasets become available, will increasingly elucidate the exposure-disease continuum.

Keywords

transcriptome; biomarker; long non-coding RNA; RNA-Seq; microarray

*Cliona M. McHale, 237 Hildebrand Hall, Genes and Environment Laboratory, School of Public Health, Division of Environmental Health Sciences, University of California, Berkeley, CA 94720. 510-643-5349 (TEL), 510-642-0427 (FAX), cmchale@berkeley.edu.

Conflict of Interest

Dr. Smith has received consulting and expert testimony fees from lawyers representing both plaintiffs and defendants in cases involving claims related to exposure to benzene.

Statement of Author Contribution

Dr. McHale prepared the manuscript draft with important intellectual and editorial input from Drs. Zhang, Thomas and Smith. All authors approved the final manuscript.

INTRODUCTION

The transcriptome is dynamic, continuously responding to changing physiological and environmental conditions in a cell, tissue, or organism, and its analysis provides the first functional readout between the genome and the expressed phenotype. Transcriptomics, the analysis of the transcriptome, has long been a cornerstone of toxicogenomic studies and has been increasingly applied in human molecular epidemiology. Microarray analysis, a hybridization-based methodology, became the most widely used technology for analysis of known transcriptomes due to its low cost, ease of use and analysis, and optimized framework of quality control (Brazma et al. 2001). However, the rapid evolution of next generation RNA sequencing technology (RNA-Seq) (Wang et al. 2009; Pertea 2012), which directly measures an entire transcriptome encompassing both known and novel components, has expanded our understanding of the scope and complexity of the human transcriptome and the myriad ways in which it can potentially be altered on the exposure-disease continuum. The unfolding knowledge produced by RNA-Seq offers the potential for a deeper understanding of the mechanism of action of chemical exposures as well as new opportunities to identify biomarkers of toxicity and early disease. In this review, we sought to summarize recent developments in our understanding of the human transcriptome and its analysis and to discuss key considerations in the application of transcriptomics in molecular epidemiology studies. We searched the peer-reviewed scientific literature in PubMed through February 2013 using combinations of search terms including transcriptome, transcriptomics, microarray, RNA sequencing, toxicogenomics, disease, molecular epidemiology, exposure, non-coding RNA (ncRNA), long non-coding RNA (lncRNA), small non-coding RNA (sncRNA), and human.

COMPOSITION AND FUNCTION OF THE HUMAN TRANSCRIPTOME

The current concept of a gene – a DNA sequence that is transcribed to a functional product – includes protein-coding genes, of which there are ~22,000 in the human genome (Pertea and Salzberg 2010), as well as non-protein coding genes, bringing the total number of estimated genes to 30,000–40,000 (Pertea 2012). Non-coding RNA is broadly categorized as sncRNA (≤ 200 bp) and lncRNA (>200 bp). This categorization is based on size rather than biological significance; the different ncRNA classes have distinct biogenesis machineries and functions.

It was recently estimated that of the base pairs in the human transcriptome, 62% are in the form of mRNAs, 53% lncRNAs and 0.7% sncRNAs (numbers do not add up to 100 % as some base pairs are part of overlapping transcripts that fall into different categories) (Pertea 2012). As ncRNAs are generally smaller than mRNAs, the vast majority of human transcripts comprise ncRNAs, with lncRNAs and sncRNAs numbering 28,191 and 10,473, respectively, compared to only 8,490 mRNAs (Pertea 2012).

Overall, ncRNAs exhibit complex patterns of expression and regulation and play important roles in transcriptional and post-transcriptional gene regulation via *cis*- and *trans*-acting mechanisms, chromatin modification, control of transcription factor binding, and regulation of alternative splicing (Pertea 2012). sncRNAs include the well-known ribosomal RNA

(rRNA) and transfer RNA (tRNA), as well as micro RNA (miRNA), small interfering RNA (siRNA), small nucleolar RNA (snoRNA), PIWI-interacting RNA (piRNA), and small nuclear RNA (snRNA) (Martens-Uzunova et al. 2013). miRNA and siRNA are post-transcriptional modulators of gene expression that bind to specific mRNA targets; snoRNAs guide the chemical modification of other RNAs; snRNA function in the processing of pre-mRNA; and piRNA form RNA-protein complexes through interactions with PIWI proteins and mediate epigenetic and post-transcriptional gene silencing of retrotransposons and other genetic elements in germ line cells. Dysregulation of ncRNA is involved in cancer and in neurological, developmental, cardiovascular and other diseases (Esteller 2011).

lncRNA in exposure and disease

lncRNAs are dysregulated in various types of cancer and in other diseases (Taft et al. 2010; Esteller 2011; Prensner et al. 2011; Moran et al. 2012) and, mechanistically, may act as tumor suppressor genes e.g. lincp21 (Huarte et al. 2010), as protooncogenes, e.g. GAGE6 (Li et al. 2009) promoters of metastasis, e.g. HOTAIR in breast cancer (Gupta et al. 2010), and as regulators of alternative splicing, e.g. MALAT1 in lung cancer (Tripathi et al. 2010). In a limited number of studies, stress and environmental exposure has been shown to alter expression of lncRNAs, e.g *SatIII* in the heat shock response (Jolly and Lakhotia 2006), several long intergenic non-coding RNA (lincRNAs) regulated by the p53 pathway in the DNA damage response (Huarte et al. 2010), *Psoriasis susceptibility-related RNA Gene Induced by Stress (PRINS)* by ultraviolet-B irradiation and viral infection (Sonkoly et al. 2005), and several lncRNAs by the tobacco carcinogen nicotine-derived nitrosamine ketone in normal human bronchial epithelial cells (Silva et al. 2010).

sncRNA in exposure and disease

Expression of sncRNAs is altered in response to environmental chemical exposures. Using bioinformatics approaches, miRNA target genes were found to be significantly enriched among genes reported to have their expression altered by environmental chemicals (Wu and Song 2011). miRNA expression is altered by multiple environmental factors including arsenic, metal-rich particulate matter, cigarette smoke, dioxins, and benzo[a]pyrene (Choudhuri 2010; Wu and Song 2011; Smirnova et al. 2012), and in various diseases (Rederstorff and Huttenhofer 2010; Esteller 2011; Law et al. 2013).

Arsenic—Results from studies in several human cell lines suggest a role for altered miRNA expression in arsenic toxicity. Exposure of human TK6 lymphoblastoid cells to sodium arsenite altered the expression of 5 miRNAs (up-regulation of hsa-miR-22, hsa-miR-34a, hsa-miR-221, and hsa-miR-222 and down-regulation of hsa-miR-210)(Marsit et al. 2006). Long-term exposure of TP53-knockdown human bronchial epithelial cells to low doses of sodium arsenite induced down-regulation of miR-200 family members and malignant transformation (Wang et al. 2011). Exposure of human umbilical vein endothelial cells to sodium arsenite induced up-regulation of 5 miRNAs and down-regulation of 52 miRNAs, suggesting a role in arsenic-induced vascular injury (Li et al. 2012). Exposure of T24 human bladder carcinoma cells to arsenic trioxide, an anticancer agent, was found to down-regulate expression of oncogenic miR-19a and to up-regulate expression of PTEN, a target of miR-19a (Cao et al. 2011). Exposure of the acute promyelocytic leukemia cell line,

NB4, to pharmacological concentrations, of arsenic trioxide induced apoptosis and up-regulation of the expression of 48 miRNAs with tumor and metastatic suppressor function (Ghaffari et al. 2012).

Particulate matter—Steel workers at a production plant in Italy (n=63), who were exposed to particulate matter (PM) and metallic PM components including arsenic, exhibited significantly increased expression of blood leukocyte miR-222 and miR-21 expression after 3 days of work, compared with baseline expression levels (Bollati et al. 2010). Exposure of human bronchial epithelial cells grown at an air–liquid interface, to diesel-generated PM, induced altered expression (>1.5-fold) of 197 miRNAs (130 up-regulated, 67 down-regulated) (Jardim et al. 2009).

Cigarette smoke—Cigarette smoke exposure has been shown to alter miRNA expression in human subjects and human cell lines. In 20 healthy subjects (10 active smokers compared with 10 never smokers), Schembri et al. found that expression of 28 miRNAs was altered (80% downregulated) in the bronchial airway epithelial cells of smokers compared with non-smokers (Schembri et al. 2009). Follow up studies on miRNA-218 (down-regulated 4-fold in the smokers), showed that down-regulation of miR-218 induced a number of smoking-related genes in airway epithelium, revealing a potential mechanism of smoking-induced disease risk (Schembri et al. 2009). In a separate study by Mascaux et al., down-regulation of miRNA expression, including miRNA-218, was detected in the biopsied bronchial epithelium of smokers with metaplasia or dysplasia compared with normal epithelium of nonsmokers (Mascaux et al. 2009).

Additional RNA diversity

As well as being comprised of multiple RNA classes, the mammalian transcriptome exhibits additional diversity. More than 90% of multi-exon protein-coding genes (Kampa et al. 2004; Wang et al. 2008) and 30% of ncRNA genes (Ravasi et al. 2006; Cabili et al. 2011) undergo alternative splicing, contributing to cellular and functional diversity. RNA editing, a process by which single nucleotide changes occur after RNA has been transcribed, affects both protein-coding and ncRNA genes (Athanasiadis et al. 2004; Sie and Kuchka 2011; Peng et al. 2012) and may generate even more transcriptome diversity than alternative splicing (Barak et al. 2009).

Further complexity in the human transcriptome comes from inherited inter-individual variability in gene expression and sequence. The majority (~90%) of disease- and trait-associated single nucleotide polymorphisms (SNPs) identified by genome-wide association studies are intronic or intragenic (Freedman et al. 2011). These inherited loci, expression quantitative trait loci (eQTL), may act in *cis* or *trans* and account for gene expression variation in the population (Morley et al. 2004; Goring et al. 2007; Idaghdour et al. 2010; Powell et al. 2012).

All aspects of the transcriptome can potentially be dysregulated on the exposure-disease continuum and have variously been assessed globally by two main technologies, microarrays and RNA-Seq.

ANALYSIS OF THE HUMAN TRANSCRIPTOME

Microarray

Through probe-based hybridization, gene expression microarrays analyze the expression of known transcripts at the resolution of genes or, in the case of exon or tiling arrays, of exons and known splicing isoforms (Kirby et al. 2007). The main advantages of microarrays are their affordability and low computational complexity. However, this hybridization-based method suffers from background issues leading to a noisy output signal and cross-hybridization issues leading to false positive signals. Further, microarrays are semi-quantitative, have a limited dynamic range due to signal saturation and lack of sensitivity to detect low abundant transcripts. Quality control practices designed to overcome biases and experimental variability and to validate the technology and data analysis in risk assessment have been developed by the MicroArray Quality Control (MAQC) Consortium and have been widely accepted by the scientific community (Shi et al. 2006; Shi et al. 2010). Despite these improvements, microarrays still suffer from limitations including an inability to comprehensively detect novel transcripts or splice variants, which can be detected by RNA-Seq.

The contribution of the microarray era to toxicogenomics is considerable as evidenced by repositories of data stored at free, publically accessible databases: Gene Expression Omnibus (GEO) (Barrett et al. 2009; Barrett et al. 2011; Barrett et al. 2013), CEBS (Waters et al. 2008b), ArrayExpress (Brazma et al. 2003; Parkinson et al. 2005; Parkinson et al. 2007; Rustici et al. 2013), Japanese Toxicogenomics Project (TGP) (Uehara et al. 2010) and DrugMatrix (Ganter et al. 2006). Microarray analysis has been applied to discovery of biomarkers of exposure and early effect, mechanism of action and risk assessment (Cui and Paules 2010; McHale et al. 2010; Currie 2012). We identified genes and pathways significantly altered by benzene exposure in the peripheral blood mononuclear cells of 125 workers exposed to a range of benzene levels (< 1 ppm to > 10 ppm) (McHale et al. 2011). Among the most significant pathways were acute myeloid leukemia (AML) and immune response. A 16-gene expression signature, with the genes having roles in immune response, inflammatory response, cell adhesion, cell–matrix adhesion, and blood coagulation, was associated with all levels of benzene exposure.

Next-generation RNA Sequencing

RNA-Seq allows the entire transcriptome to be analyzed in a high-throughput and quantitative manner (Wang et al. 2009). In Illumina's sequencing by synthesis (SBS) approach, during each sequencing cycle, fluorescently-labeled reversible terminators are incorporated, imaged and cleaved to allow incorporation of the next base. Sequencing is conducted simultaneously on millions of different anchored template molecules in parallel. Amplification of template molecules generates a detectable signal. Unlike hybridization-based methods, RNA-Seq has no saturation bias and relatively low background noise, a much larger dynamic range than microarrays and the ability to detect low expressed genes. Since it is not limited to the detection of known transcripts, the most significant advantage associated with RNA-Seq is its ability to assess diverse aspects of the transcriptome. RNA-Seq has led to the discovery of new species of ncRNAs (MacLean et al. 2009; Zhou et al.

2010) such as piRNA and has increased understanding of the expression and regulation of lncRNAs (Atkinson et al. 2012). It also has unprecedented base pair resolution, allowing the precise identification of exon and intron boundaries. Co-analysis of genotype data and RNA-Seq data in 60–70 individuals led to the discovery of new eQTLs (Montgomery et al. 2010; Pickrell et al. 2010). RNA-Seq has revealed several classes of alterations in cancer cells, revealing pathogenic mechanisms, including fusion transcripts, alternative splicing, allelic imbalance, single nucleotide variations, and RNA editing (Morin et al. 2008; Chepelev 2012; Pertea 2012; Costa et al. 2013; Mutz et al. 2013). RNA-Seq has even been applied to identify known and novel microorganisms in high throughput sequencing data generated from human tissue, using a program called PathSeq, which performs computational subtraction to identify non-human nucleic acids (Kostic et al. 2011). Thus, RNA-Seq has the ability and to measure the complete transcriptomic response to stress and stimuli and to clarify the interrelationship among all aspects of gene expression regulation.

Despite its significant advantages over microarrays, RNA-Seq suffers from limitations (Fang and Cui 2011; Schwartz et al. 2011; Hansen et al. 2012; Costa et al. 2013). Standard methods of preparation of RNA for RNA-Seq employ selection of polyadenylated (polyA) RNAs or rRNA depletion (ribo-depletion). Some studies have suggested that ribo-depletion may be superior for producing reliable coding and non-coding gene expression data (Cui et al. 2010; Huang et al. 2011). Experimental bias can arise at multiple steps in sample preparation for RNA-Seq, including fragmentation, cDNA amplification efficiency, and PCR amplification. Another source of bias is heterogeneity in coverage across the length of a transcript (Bullard et al. 2010; Hansen et al. 2010). Third-generation sequencing approaches are being developed that directly sequence RNA without the need for cDNA or PCR amplification, but they have high error rates and are costly (Schadt et al. 2010). Challenges in data generation, storage and interpretation also exist, as discussed previously (Costa et al. 2010; Pertea 2012; Costa et al. 2013; Mutz et al. 2013; Riedmaier and Pfaffl 2013). Up to 600 gigabases (Gb) can be generated in a single run, equivalent to 200-fold coverage of the human genome (Pertea 2012). Data analysis consists of aligning the short read sequences to a reference genome or transcriptome or *de novo* assembly, counting the number of mapped reads, calculating transcript expression level and differential gene expression using normalized gene expression scores and statistical tests. Alignment can be challenging in the case of alternative transcripts with shared exons, strand-specific sequences, and low-abundance transcripts (Pertea 2012). For differential expression, non-parametric algorithms are less dependent on sequencing depth and may achieve more robust results (Tarazona et al. 2011). The third phase of the MAQC project, MAQC-III, called Sequence Quality Control (SeQC), seeks to assess the technical performance of RNA-Seq platforms through the generation of benchmark data sets with reference samples (http://www.fda.gov/MicroArrayQC/). Tools to assess sequencing performance and library quality which are critical to the interpretation of RNA-seq data, are being developed such as RNA-SeQC, a program which provides key measures of data quality (DeLuca et al. 2012). RNA-SeQC metrics include yield, alignment and duplication rates; GC bias, rRNA content, regions of alignment (exon, intron and intragenic), continuity of coverage, 3′/5′bias and count of detectable transcripts.

RNA-Seq has been employed to analyze the transcriptome of multiple diseases (Kavanagh et al. 2012; Raghavachari et al. 2012; Costa et al. 2013; Jakhesara et al. 2013; Mills et al. 2013). Fewer toxicogenomic and molecular epidemiology studies have been published on studies utilizing RNA-Seq (Beane et al. 2011; Su et al. 2011; Hackett et al. 2012). Studies comparing RNA-Seq and microarray data have illustrated strengths and weaknesses of both technologies enabling their optimization.

Comparison of RNA-Seq and Microarray Data

Several studies have addressed the overlap of RNA-Seq and microarray data in mammalian cells. Liu et al. profiled gene expression in chimpanzees and rhesus macaques using high-density Affymetrix Human Exon Junction Array and Illumina RNA-Seq (Liu et al. 2011). They identified 40% more differentially expressed genes by RNA-Seq, reflecting the greater dynamic range. They observed a systematic increase in the RNA-Seq error rate for low-expressed genes, which they further confirmed in the comparison of two MAQC human reference RNA samples. This increased error may be due to insufficient coverage; 10–20 million reads were generated per sample in the study. Marioni et al. (Marioni et al. 2008) reported high coefficients of variance (CV) at low read counts. As RNA-Seq is a sampling method, stochastic events/Poisson error counting are a source of error in the quantification of rare transcripts.

Raghavachari et al. compared the whole blood transcriptomes of 6 sickle cell disease (SCD) patients and 3 controls by Illumina RNA-Seq and Affy Human Exon 1.0 ST microarray (Raghavachari et al. 2012). As with Liu et al., they reported a greater sensitivity and dynamic range for RNA-Seq. They reported a higher technical variability among RNA-Seq replicates, indicated by CV independent of expression level, perhaps due to an inadequate sequencing depth of 10 million reads. Nonetheless, both platforms revealed similar biology of SCD. In addition, RNA-Seq identified 16 alternatively spliced genes, as well as novel GEX from an unannotated genomic region, a novel exon in the ALAS2 gene, and mutations in transcripts.

Two studies examined the alteration of global gene expression in human airway epithelium by cigarette smoking, using both RNA-Seq and microarrays. Hackett et al. performed RNA-Seq to quantify the human small airway epithelium transcriptome (SAE) of 5 nonsmokers and 6 healthy smokers and compared the data to Affymetrix Human Genome U133 Plus 2.0 microarray data generated from 12 healthy smokers and 12 non-smokers (Hackett et al. 2012). Beane compared Illumina RNA-Seq data and Affymetrix Exon 1.0 ST and HGU133A 2.0 microarray data in pooled bronchial airway epithelial cell samples from healthy never smoker (n=3) and current smoker volunteers (n=3) and smokers with (n=8) and without lung cancer (n=5) (Beane et al. 2011). In both studies, the two methods were well correlated at the fold change level. RNA-Seq detected many additional smoking (Beane et al. 2011; Hackett et al. 2012) and cancer-related transcripts (Beane et al. 2011). Cigarette smoke creates a field of injury in the airway epithelium of the respiratory tract (Brody 2012) and these new data have expanded the understanding of the biology of that effect induced by smoking. In addition, Hackett et al. found that smoking had no effect on SAE gene splicing, a known feature of SAE in lung cancer (Xi et al. 2008; Misquitta-Ali et al. 2011) and Beane

et al. identified differentially expressed ncRNAs (lincRNAs, pseudogenes, and processed transcripts) which may have important gene regulatory functions in lung carcinogenesis (Beane et al. 2011). Beane et al. also suggested that to completely characterize the transcriptome, library preparation protocols that measure the expression of non-polyadenylated RNAs are needed, using longer read lengths and PE sequencing, both of which yield a higher percentage of mapped reads (greater than 30M).

At 30–50M reads, Su et al. found that Illumina RNA-Seq was more sensitive that Affymetrix Rat Genome 230 2.0 microarrays at detecting aristolochic acid-induced transcriptional changes in low-expressed genes in the kidneys of exposed rats (n=4) compared with controls (n=4) (Su et al. 2011). RNA-Seq detected ~50% more differentially expressed genes and 300% more if multiple testing was applied to the statistical analysis. Both platforms revealed the underlying biology but RNA-Seq was more sensitive. Van Delft examined the effects of BaP in HepG2 cells after 12 and 24 hrs by RNA-Seq and Affy HGU133 Plus 2.0 GeneChip array (van Delft et al. 2012). RNA-Seq detected 20% more genes and 3-fold more differentially expressed genes as well as providing more insight into biology and mechanisms, through the identification of more significant pathways and processes. In addition, RNA-Seq revealed novel isoforms, including novel exon-skipping events in 735 genes and splice variants with altered expression in 839 genes.

In a pilot case study described in detail in an accompanying paper in this issue (Thomas 2013), we analyzed by RNA-Seq the transcriptomes of 10 workers highly-exposed (>5 ppm) to the leukemogen benzene and 10 unexposed control study subjects matched by age, sex, and smoking status that had been previously analyzed by microarray (McHale et al. 2011). We compared the data obtained by both methods. Overall correlation between RNA-Seq and microarray intensities was 0.6, comparable to published studies (Marioni et al. 2008; Bradford et al. 2010) and suggested that RNA-Seq was better able to detect low intensity gene expression. In these 20 subjects, we identified 146 statistically significant differentially expressed genes (including 29 ncRNAs) by RNA-Seq compared with 1 gene by microarray. There was overlap among the genes and pathways identified in the RNA-seq pilot study and those identified in 125 subjects by microarray. We also identified differential splicing as a potential mechanism of benzene toxicity which should be further investigated as splicing diversity is involved in hematopoietic function and differentiation, (Tondeur et al. 2010) and alternative splicing is a known leukemogenesis pathway (Maciejewski and Padgett 2012).

CONSIDERATIONS IN TRANSCRIPTOME STUDY DESIGN, ANALYSIS AND INTERPRETATION

Multiple factors influence the usefulness of transcriptomics in molecular epidemiology studies, including the type of tissue or cell type analyzed, study design, and analysis and interpretation of the results, as summarized in Figure 1 and described in detail below.

Choice of target tissue/cell type to analyze

In human studies, transcriptomic effects are typically analyzed in readily available tissues such as blood. Such studies have revealed biomarkers of exposure and mechanisms of

toxicity in whole blood or peripheral blood leukocytes in populations exposed to benzene (Forrest et al. 2005; McHale et al. 2011), arsenic (Argos et al. 2006; Andrew et al. 2008), perfluorooctanoic acid (Rylander et al. 2011), acetaminophen (APAP) (Fannin et al. 2010; Jetten et al. 2012), metal fumes (Wang et al. 2005), cadmium (Dakeshita et al. 2009), diesel exhaust (Peretz et al. 2007; Pettit et al. 2012), dioxin (McHale et al. 2007) cigarette smoke (Spira et al. 2004a; Spira et al. 2004b; Spira et al. 2004c; Charlesworth et al. 2010; Beineke et al. 2012; Bosse et al. 2012), polychlorinated hydrocarbons (Dutta et al. 2012; Mitra et al. 2012), and metal-rich particulate matter (miRNA) (Bollati et al. 2010). Biomarkers and mechanisms of disease have also been revealed through analysis of peripheral blood transcriptomes in a range of diseases including colorectal cancer (Han et al. 2008), autism spectrum disorder (Kong et al. 2012), sickle cell disease (Raghavachari et al. 2012), hypertension and Type 2 Diabetes (Stoynev et al. 2013), coronary artery disease (Nuhrenberg et al. 2013), myocardial infarction (Kiliszek et al. 2012), aggressive/advanced prostate cancer (Liong et al. 2012), Alzheimer's disease (Lunnon et al. 2013) and Graves' disease (Liu et al. 2012).

It is unclear how good a surrogate the blood transcriptome is for target tissues. The peripheral blood transcriptome was shown to overlap considerably with those of nine other tissue types in healthy individuals (Liew et al. 2006). A study mapping expressed genes to gene ontologies recently showed that the white blood cell transcriptome was a good surrogate for a generalized multi-organ transcriptome constructed using profiles from healthy and diseased individuals (Kohane and Valtchinov 2012). Similar patterns of gene expression in RNA-stabilized whole blood and lung were reported for early stage lung adenocarcinoma (Rotunno et al. 2011) and non-small cell lung cancer (NSCLC) cases compared with controls (Showe et al. 2009; Zander et al. 2011). However, fewer probes (~300), by an order of magnitude, were found to be affected by smoking in blood lymphocytes (Charlesworth et al. 2010) than in non-tumor lung tissue of lung cancer patients (>300) (Bosse et al. 2012). Though the number of differentially expressed probes identified was likely influenced by the statistical methodologies used, this suggests that effects in blood may not fully recapitulate those in the target tissues. In the NSCLC studies, the cancer-associated genes were enriched in immune function (Showe et al. 2009; Zander et al. 2011), and peripheral blood mononuclear cell (PBMC) gene signatures of immune response were also shown to predict NSCLC outcome (Kossenkov et al. 2012; Showe et al. 2012).

Immune response is frequently impacted in toxicogenomic studies. Studies examining gene expression changes in the blood transcriptome during liver injury induced by APAP have identified mechanisms and potential predictive biomarkers of hepatotoxicity in rats and humans (Cui and Paules 2010; Fannin et al. 2010). Similar expression changes and biological processes, including immune response, in a subset of genes in liver and blood of APAP exposed rats were reported by analysis of toxicogenomics data from prior studies using an Extracting Patterns and Identifying Co-expressed Genes (EPIG) approach (Zhang et al. 2012). Using available rat gene expression data sets, Huang demonstrated that blood gene expression profiles could predict liver necrosis induced by exposure to a wide variety of hepatotoxicants and validated their findings in independent data sets (Huang et al. 2010). Pathways impacted by the predicted genes included immune (Toll-like receptor signaling) and inflammatory response. We have reported effects of benzene on immune response,

including Toll-like receptor signaling and B- and T-cell receptor signaling, in exposed individuals (McHale et al. 2011). Together, these data suggest that the blood transcriptome may identify biomarkers relevant to the organ of interest and capture processes such as immune response, but may not capture the entirety of the mechanistic response in the target tissue. Even for immune response, circulating immune cell subsets may not fully reflect the entire immune response at the tissue level.

Another challenge associated with analysis of the blood transcriptome is that it is typically analyzed in whole blood or PBMC, a mixed cell population in which proportions of distinct cell types vary by individual. Whole blood or PBMC are typically analyzed because it is not always known *a priori* which blood population to analyze and available samples or sample preparation and storage techniques may preclude such analyses. Further, positive selection approaches such as incubation of PBMCs with anti-CD19 or anti-CD20- to purify B-cell populations may activate cell-surface receptors and alter gene expression. However, blood cell populations may be altered by certain hematotoxic exposures, e.g. benzene (Lan et al. 2004) and trichloroethylene (Lan et al. 2010) and changes in gene expression may simply reflect this.

Statistical analysis can be used to adjust for blood cell counts in analysis of transcriptomics if these data are available. Exposure to benzene in air could affect the mean gene expression measured in PBMC via two causal pathways – either directly or via the ensuing hematotoxicity. In a recent analysis we attempted to estimate the direct effects (Petersen et al. 2006) of benzene exposure on changes in mean gene expression. We estimated these effects non-parametrically using the SuperLearner (Sinisi et al. 2007) an approach which allowed the data to guide the choice of models of mean gene expression as functions of benzene exposure, counts of different types of PBMC and other potential confounders like gender and smoking status (in submission). Deconvolution approaches have been applied to identify cell-type specific effects from whole blood transcriptomes (Abbas et al. 2009; Shen-Orr et al. 2010; Bolen et al. 2011). Gene expression profiles of blood cell subsets are increasingly available (Martinez 2009; Watkins et al. 2009; Shen-Orr et al. 2010; Tondeur et al. 2010; Beyer et al. 2012), facilitating such analyses. Similar approaches have been taken in the assessment of epigenetic effects in blood samples (Houseman et al. 2012; Liu et al. 2013). Such deconvolution approaches, however, may not be able to reveal information on minor immune subsets or on cells in various stages of differentiation and activation in each heterogeneous lineage. Single cell sequencing may ultimately be necessary to obtain that level of detail. In the case of leukemogens, the relevant target may be the hematopoietic stem cell (McHale et al. 2012).

Accessible, disease-relevant tissues other than blood are useful surrogates for investigating the effects of smoking and possibly other lung toxicants in lung cancer. Cigarette smoke creates a field of injury in the airway epithelium of the respiratory tract (Brody 2012). Several groups have demonstrated similarly altered gene expression induced by cigarette smoke in the cytologically normal small and large airway epithelium (Spira et al. 2004b; Beane et al. 2007; Zhang et al. 2008; Tilley et al. 2009; Beane et al. 2011; Gower et al. 2011) and in nasal and buccal epithelium (Sridhar et al. 2008; Boyle et al. 2010; Zhang et al. 2010). Tan et al. found concordant expression levels of antioxidant and xenobiotic genes and

p16 in laser-captured alveolar macrophages and distal airway epithelial cells of 62 smokers without cancer (Tan et al. 2009). Further, airway gene expression in COPD was shown to reflect molecular processes occurring in more distal diseased lung tissue (Gower et al. 2011). Expression signatures in these accessible tissues have prognostic capability; cytologically normal epithelial cells collected at bronchoscopy from smokers with suspected lung cancer revealed an epithelial cell GEX-based biomarker with diagnostic accuracy of 83% (Spira et al. 2007).

Another disease relevant tissue analyzed in molecular epidemiology studies is exfoliated bladder cells in the investigation of bladder cancer (Rosser et al. 2009; Urquidi et al. 2012).

Experimental variation

The number of transcriptomic endpoints generated by an RNA-Seq study is very large, with the current number of genes estimated at 30–40,000 plus alternative isoforms and transcript variants (Pertea 2012). Toxicogenomic studies therefore need to be designed with sufficient power (relatively large sample sizes) to detect effects of the exposure under consideration. Depending on the goal of the study, as discussed above, RNA-Seq studies need to have sufficient depth of coverage to accurately measure expression and sequence of rare variants (Marioni et al. 2008; Liu et al. 2012). Inadequate study design with respect to these factors can increase the probability of false positive findings (Ioannidis 2005; Robles et al. 2012). In microarray analysis, we (McHale et al. 2011) and others (Kitchen et al. 2011; Schurmann et al. 2012) found variation due to factors such as RNA extraction, labeling, hybridization, chip assignment that require statistical adjustment. In our study, analysis with a mixed-effects model minimized potential confounding and experimental variability.

Factors related to sample preparation and storage can influence the quality of the data generated and limit inter-study comparisons. Millions of human biosamples are currently stored in biobanks and have the potential to yield valuable transcriptomic data. Hebels et al. investigated the effect of handling and prolonged storage on the suitability of fresh and biobanked blood samples and isolated components for transcriptomic analysis (Hebels et al. 2013). They found that adequate amounts of microarray-quality RNA with RNA Integrity Number (RIN)>6.0 (average RIN=7. 2, similar to fresh samples) could be isolated from ~85% of the biobank samples tested, even after 13–17 years of storage. Differences in gene expression profiles were mainly associated with longer bench times prior to sample processing, followed by choice of anticoagulant (mainly EDTA vs. heparin) and, to a much lesser extent, storage temperature. They also found that transcriptomics quality RNA could be isolated from buffy coat samples frozen in the absence of RNA preservative, by thawing these samples in the presence of RNAlater, provided the buffy coats had been deep frozen within 8hr of blood collection. In another study, even a 4h processing delay after phlebotomy led to altered expression of genes involved in inflammatory, immunologic, and cancer pathways (Barnes et al. 2010). The effect of different isolation techniques has also been examined. Compared to PBMC, PAXgene RNA-stabilized samples showed a lower number of expressed genes, lower gene expression values, and higher variability, probably due to the differing cell populations in each sample type and the presence of globin RNA in the PAXgene samples (Min et al. 2010). Globin reduction was shown to improve data

quality from microarray analysis on Illumina BeadChips (Tian et al. 2009) but may not be necessary for RNA-Seq analyses (Raghavachari et al. 2012). Debey-Pascher et al. reported substantially differing expression profiles in fresh and cryopreserved PBMC and found that expression profiles in cryopreserved PBMC samples were significantly altered with increasing storage period, whereas profiles from PAXgene RNA-stabilized remained unaltered (Debey-Pascher et al. 2011). Weber et al. 2010 found that superior RNA yield and integrity values were obtained from blood samples stabilized with RNALater than with PAXgene, though both produced RNA of acceptable quality and detected similar expression levels of specific genes by qRT-PCR (Weber et al. 2010). Poor overlap in expression profiles from PBMC and PAXgene RNA-stabilized WB was observed in several studies (Debey et al. 2004; Debey et al. 2006; Zander et al. 2011). All of these factors can limit the usefulness of cross-study comparisons using publicly available datasets.

Confounding factors

An individual's transcriptome reflects characteristics of the individual including genotype, specifically eQTL; the microbiome or enterotype; irreversible alterations in gene expression acquired *in utero* and throughout life; and dynamic transcriptional responses to the exposure of interest and all other confounding exposures at the time of sampling. Molecular epidemiological studies assessing toxicogenomic endpoints typically account for age, gender, smoking, some dietary features, infection, alcohol intake, medication use, confounding exposures, and mixed cell populations. It may be difficult to account for all aspects of dietary effect on the human blood transcriptome as so many components have effects including macronutrients and micronutrients (Pagmantidis et al. 2008; Ryu et al. 2011; de Mello et al. 2012; Drew 2012; Sagaya et al. 2012; Vedin et al. 2012). Through regulation of host gene expression, the human gut microbiome performs functions critical to host physiology including processing and biotransformation of xenobiotics, regulation of human metabolism, and shaping the development of the immune system (Nicholson and Wilson 2003; Clemente et al. 2012; Nicholson et al. 2012). There are three major forms of host enterotype with a huge degree of inter-individual variation (Arumugam et al. 2011). Environmental stressors that disturb the balance between commensal microbes and their human hosts may alter host physiology and underlie many disease states (Spor et al. 2011; Maurice et al. 2013). Stress (Kawai et al. 2007) and exercise (Connolly et al. 2004; Zieker et al. 2005; Carlson et al. 2011) also impact the transcriptome.

Studies of gene expression in different ethnic groups in Morocco and Fiji suggested that over a third of the human transcriptome is influenced by environmental geography, with much lesser influences of age, gender, and genetic factors (Idaghdour et al. 2008; Idaghdour et al. 2010; Nath et al. 2012). Aspects of immune function were found to be strongly affected by regional factors, potentially influencing susceptibility to respiratory and inflammatory disease, in the Morocco study.

Distal environmental conditions, such as *in utero* or early childhood exposures, can influence an individual's response to a later exposure (Szyf 2007) and disease risk (Votavova et al. 2011; Martinez et al. 2012). Fetal exposure to carcinogens was shown to have gender specific effects on gene expression in the cord blood of newborns (Hochstenbach et al. 2012)

and maternal smoking was reported to cause significant changes in the transcriptome of placental and fetal cells and to deregulate pathways associated with autoimmune diseases in the newborns of smokers (Votavova et al. 2011). These effects could be mediated through long-term effects on gene expression and cumulative damage such as genetic or epigenetic mutations could increase the risk of disease even at low exposures, particularly those diseases occurring later in life. Indeed, a greater effect of environmental tobacco smoke (ETS) was found among smokers compared to never-smokers in a large prospective study of respiratory cancer and chronic obstructive pulmonary disease (Vineis et al. 2005). Several studies have shown that changes in gene expression induced by cigarette smoke remain altered many years after smoking cessation (Spira et al. 2004c; Beane et al. 2007; Bosse et al. 2012).

Given the huge potential for confounding, human toxicogenomic studies need to be carefully designed and analyzed in order to identify true causal associations with the exposure of interest. For example, this can be done with appropriate choices of cases and controls in case-control studies matched on a given set of confounders and possibly accounting for other confounders in statistical analyses like those involving mixed models (Laird and Ware 1982). Mixed models were developed to quantify variation of an outcome like gene expression from different sources (genotype, diet). This would result in a lower residual (unexplained) variance and thus provide greater power in detecting exposure effects.

Interpretation of subtle changes in expression and low-dose effects

In our analysis of global gene expression in 125 benzene-exposed subjects and controls, we reported the subtle (small fold-change, many far less than 2-fold) alteration of expression of ~2000 genes, dose-dependent effects on gene expression and biochemical pathways, and an apparently supra-linear response in the expression of a 16-gene signature (McHale et al. 2011). Using non-parametric approaches to statistically model the dose-response of AML pathway gene expression in our benzene-exposed population including exposed and control individuals, with air benzene exposure levels in the latter estimated from unmetabolized urinary benzene levels, we found that the AML pathway and pathway representative genes exhibited similar supralinear responses and responses at benzene levels as low as 100 ppb in air (unpublished data). It is unclear what the implication of these pathway alterations at low benzene exposure levels is for AML risk.

In a recent study examining the effects of short term, low dose APAP in human volunteers, transcriptomics (mRNA and miRNA) outperformed clinical chemistry tests, revealing novel response pathways to APAP and detecting dose-specific immune-modulating effects that suggested the occurrence of possible pre-toxic effects of therapeutic APAP doses (Jetten et al. 2012). Both of these studies analysed the transcriptome by microarray; RNA-Seq has the potential to detect more subtle changes in gene expression. The implications of such subtle changes in expression, and the distinction of adaptive responses from adverse effects at low doses (Jennings 2013), will present challenges in the application of transcriptomics in molecular epidemiology studies.

Pathway analysis, systems biology and the exposure-disease continuum

As described earlier, most recently published transcriptomic data is publicly available through the GEO (Barrett et al. 2009; Barrett et al. 2011; Barrett et al. 2013), CEBS (Waters et al. 2008b), ArrayExpress (Brazma et al. 2003; Parkinson et al. 2005; Parkinson et al. 2007; Rustici et al. 2013), Japanese Toxicogenomics Project (TGP) (Uehara et al. 2010) and DrugMatrix (Ganter et al. 2006) databases. Gene Expression Transcriptomic and other toxicogenomic datasets are available through the Comparative Toxicogenomics Database (CTD) (Mattingly et al. 2006) and Chemical Effects in Biological Systems (CEBS) (Waters et al. 2008a). In July 2012, CTD contained manually curated data on 599182 chemical-gene interactions, 176627 chemical-disease, and 23395 gene-disease relationships, internal integration of which leads to >10.1 million inferred gene-disease relationships and 913 622 inferred chemical-disease relationships (Davis et al. 2013). Integration with GO, KEGG and reactome provides 15.6 million toxicogenomic relationships for analysis, 3.6-fold and 10.6-fold since CTD's 2011 and 2009 reports, respectively. The data sets in these databases from multiple toxicogenomic studies provide opportunities for meta-analyses of exposure effects and for comparing and relating effects of different exposures.

Integration of transcriptomic and other toxicogenomic data in a single study can provide enhanced information on pathways by increasing the number of data points in a pathway. Integration of different omic datasets can also inform the mechanism of dysregulation of gene expression by an exposure as gene expression is regulated by multiple interrelated proceses including epigenetics and non-coding RNA. Comparison of the differential transcriptomic profiles of lung alveolar macrophages in smokers and non-smokers with corresponding DNA methylation (Philibert et al. 2012) and miRNA (Graff et al. 2012) profiles showed that changes in expression are caused by both of these mechanisms. Hawkins *et al.* recently reviewed issues related to the use of integrated analysis of RNA-Seq and other genomic datasets to understand mechanisms of gene regulation, homeostasis and responses to the environment (Hawkins et al. 2010). Analysis of both miRNA and lncRNA deregulation in lung cancer has improved the understanding of lung cancer biology (Enfield et al. 2012). Analysis at the pathway level can compensate for some of the limitations in cross-study comparisons resulting from experimental variation and confounding discussed earlier. Despite a lack of overlap in differential expression of individual genes identified at the transcriptome level between COPD studies, a striking overlap in biologic themes was reported (Zeskind et al. 2008).

Gene expression changes exist on an exposure-disease continuum as illustrated by the finding of a group of smokers with a "COPD-like" small airway epithelium transcriptome (Tilley et al. 2011). Integrated analyses of multiple toxicogenomic datasets can identify more robust biomarkers and clarify mechanisms of toxicity and disease. A network-based gene-environment-disease approach utilized data from the CTD and the Genetic Association database (Becker et al. 2004) and identified key regulatory pathways that integrate genetic and environmental modulators of disease (Gohlke et al. 2009). Recently, we sought to identify pathways common to chemical leukemogens and to determine whether leukemogens could be distinguished from non-leukemogenic carcinogens, using bioinformatics approaches (Thomas et al. 2012). We assessed enrichment of all 250 human

biochemical pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database in data on gene/protein targets available in the CTD for 29 leukemogens and 11 non-leukemogenic carcinogens. The top pathways targeted by the leukemogens included metabolism of xenobiotics by cytochrome P450, glutathione metabolism, neurotrophin signaling pathway, apoptosis, MAPK signaling, Toll-like receptor signaling and various cancer pathways. The 29 leukemogens formed 18 distinct clusters comprising 1 to 3 chemicals that did not correlate with known mechanism of action or with structural similarity. Using two-class random forests to estimate leukemogen and non-leukemogen patterns, we estimated a 76% chance of distinguishing a random leukemogen/non-leukemogen pair from each other based on their known toxicogenomic targets.

Toxicity pathways and potential biomarkers of phthalate toxicity were identified from a recent analysis of CTD data which revealed a total of 445 interactions between five frequently curated phthalates and 249 unique genes/proteins were found (Singh and Li 2011). We previously described our systems biology approach to analyzing various toxicogenomic datasets generated from our occupational study of benzene exposure (Zhang et al, 2010).

Currently, the majority of toxicogenomic datasets are generated from studies examining the effects of single chemicals or complex mixtures such as cigarette smoke or air pollution. In reality, the exposure component of the exposure-disease continuum is much more complex, with people exposed to many different chemicals/stresses throughout their lives, cumulatively influencing their disease risk. The totality of all exposures from conception onward, including xenobiotics and internally generated toxicants, was defined as the exposome by Christopher Wild in 2005 (Wild 2005). Moving away from the single exposure association approach, exposomics seeks to measure the internal chemical environment of an individual using omic technologies, including transcriptomics, and to relate snapshots and dynamic changes in the omic data over time to preclinical and clinical perturbations and disease onset in prospective studies (Rappaport and Smith 2010; Rappaport 2012). We previously suggested an exposomic approach to identify key biomarkers and exposures associated with AML, for which only about ~20% of environmental causes (such as benzene) are known (Smith et al. 2011). Such a study could compare exposome patterns between *de novo* leukemia cases and controls in longitudinal studies, and ultimately identify leukemia-inducing exposures.

Conclusion

With appropriate study design, RNA-Seq can measure all aspects of the complex and diverse human transcriptome in human toxicogenomic studies. However, factors including cell/tissue choice, sample quality and preparation, experimental variation, confounding, and interpretation of subtle changes and low-dose effects in terms of adaptive and adverse effects limit the true potential of RNA-Seq and other omics methodologies in individual studies and cross-study comparisons. Pathway analysis can compensate for some of these limitations and will, together with systems biology approaches to analyze traditional toxicogenomic and exposome studies, increasingly elucidate the exposure-disease continuum.

Acknowledgments

Our studies on benzene described here were supported by National Institutes of Health grant (P42ES004705 to M.T.S) and by EPA contract number EP-11-001398.

References

Abbas AR, Wolslegel K, Seshasayee D, Modrusan Z, Clark HF. Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. PLoS One. 2009; 4(7):e6098. [PubMed: 19568420]

Andrew AS, Jewell DA, Mason RA, Whitfield ML, Moore JH, Karagas MR. Drinking-water arsenic exposure modulates gene expression in human lymphocytes from a U.S. population. Environ Health Perspect. 2008; 116(4):524–531. [PubMed: 18414638]

Argos M, Kibriya MG, Parvez F, Jasmine F, Rakibuz-Zaman M, Ahsan H. Gene expression profiles in peripheral lymphocytes by arsenic exposure and skin lesion status in a Bangladeshi population. Cancer Epidemiol Biomarkers Prev. 2006; 15(7):1367–1375. [PubMed: 16835338]

Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Dore J, Antolin M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, Dervyn R, Foerstner KU, Friss C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-Vlieg J, Jamet A, Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E, Merieux A, Melo Minardi R, M'Rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y, Zeller G, Weissenbach J, Ehrlich SD, Bork P. Enterotypes of the human gut microbiome. Nature. 2011; 473(7346):174–180. [PubMed: 21508958]

Athanasiadis A, Rich A, Maas S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. 2004; 2(12):e391. [PubMed: 15534692]

Atkinson SR, Marguerat S, Bahler J. Exploring long non-coding RNAs through sequencing. Semin Cell Dev Biol. 2012; 23(2):200–205. [PubMed: 22202731]

Barak M, Levanon EY, Eisenberg E, Paz N, Rechavi G, Church GM, Mehr R. Evidence for large diversity in the human transcriptome created by Alu RNA editing. Nucleic Acids Res. 2009; 37(20): 6905–6915. [PubMed: 19740767]

Barnes MG, Grom AA, Griffin TA, Colbert RA, Thompson SD. Gene Expression Profiles from Peripheral Blood Mononuclear Cells Are Sensitive to Short Processing Delays. Biopreserv Biobank. 2010; 8(3):153–162. [PubMed: 21743826]

Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muertter RN, Holko M, Ayanbule O, Yefanov A, Soboleva A. NCBI GEO: archive for functional genomics data sets–10 years on. Nucleic Acids Res. 2011; 39:D1005–1010. (Database issue). [PubMed: 21097893]

Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muertter RN, Edgar R. NCBI GEO: archive for high-throughput functional genomic data. Nucleic Acids Res. 2009; 37:D885–890. (Database issue). [PubMed: 18940857]

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013; 41:D991–995. (Database issue). [PubMed: 23193258]

Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. Genome Biol. 2007; 8(9):R201. [PubMed: 17894889]

Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, Luo L, Zhang XH, Xiao J, Alekseyev YO, Wang S, Levy S, Massion PP, Lenburg M, Spira A. Characterizing the impact of

smoking and lung cancer on the airway transcriptome using RNA-Seq. Cancer Prev Res (Phila). 2011; 4(6):803–817. [PubMed: 21636547]

Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. Nat Genet. 2004; 36(5):431–432. [PubMed: 15118671]

Beineke P, Fitch K, Tao H, Elashoff MR, Rosenberg S, Kraus WE, Wingrove JA, Investigators P. A whole blood gene expression-based signature for smoking status. BMC Med Genomics. 2012; 5:58. [PubMed: 23210427]

Beyer M, Mallmann MR, Xue J, Staratschek-Jox A, Vorholt D, Krebs W, Sommer D, Sander J, Mertens C, Nino-Castro A, Schmidt SV, Schultze JL. High-resolution transcriptome of human macrophages. PLoS One. 2012; 7(9):e45466. [PubMed: 23029029]

Bolen CR, Uduman M, Kleinstein SH. Cell subset prediction for blood genomic studies. BMC Bioinformatics. 2011; 12:258. [PubMed: 21702940]

Bollati V, Marinelli B, Apostoli P, Bonzini M, Nordio F, Hoxha M, Pegoraro V, Motta V, Tarantini L, Cantone L, Schwartz J, Bertazzi PA, Baccarelli A. Exposure to metal-rich particulate matter modifies the expression of candidate microRNAs in peripheral blood leukocytes. Environ Health Perspect. 2010; 118(6):763–768. [PubMed: 20061215]

Bosse Y, Postma DS, Sin DD, Lamontagne M, Couture C, Gaudreault N, Joubert P, Wong V, Elliott M, van den Berge M, Brandsma CA, Tribouley C, Malkov V, Tsou JA, Opiteck GJ, Hogg JC, Sandford AJ, Timens W, Pare PD, Laviolette M. Molecular signature of smoking in human lung tissues. Cancer Res. 2012; 72(15):3753–3763. [PubMed: 22659451]

Boyle JO, Gumus ZH, Kacker A, Choksi VL, Bocker JM, Zhou XK, Yantiss RK, Hughes DB, Du B, Judson BL, Subbaramaiah K, Dannenberg AJ. Effects of cigarette smoke on the human oral mucosal transcriptome. Cancer Prev Res (Phila). 2010; 3(3):266–278. [PubMed: 20179299]

Bradford JR, Hey Y, Yates T, Li Y, Pepper SD, Miller CJ. A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. BMC Genomics. 2010; 11:282. [PubMed: 20444259]

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet. 2001; 29(4):365–371. [PubMed: 11726920]

Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, Holloway E, Kapushesky M, Kemmeren P, Lara GG, Oezcimen A, Rocca-Serra P, Sansone SA. ArrayExpress–a public repository for microarray gene expression data at the EBI. Nucleic Acids Res. 2003; 31(1): 68–71. [PubMed: 12519949]

Brody JS. Transcriptome alterations induced by cigarette smoke. Int J Cancer. 2012; 131(12):2754–2762. [PubMed: 22961494]

Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics. 2010; 11:94. [PubMed: 20167110]

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011; 25(18):1915–1927. [PubMed: 21890647]

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(1):179–188. [PubMed: 20857258]

Carlson LA, Tighe SW, Kenefick RW, Dragon J, Westcott NW, Leclair RJ. Changes in transcriptional output of human peripheral blood mononuclear cells following resistance exercise. Eur J Appl Physiol. 2011; 111(12):2919–2929. [PubMed: 21437602]

Charlesworth JC, Curran JE, Johnson MP, Goring HH, Dyer TD, Diego VP, Kent JW Jr, Mahaney MC, Almasy L, MacCluer JW, Moses EK, Blangero J. Transcriptomic epidemiology of smoking: the effect of smoking on gene expression in lymphocytes. BMC Med Genomics. 2010; 3:29. [PubMed: 20633249]

Chepelev I. Detection of RNA editing events in human cells using high-throughput sequencing. Methods Mol Biol. 2012; 815:91–102. [PubMed: 22130986]

Choudhuri S. Small noncoding RNAs: biogenesis, function, and emerging significance in toxicology. J Biochem Mol Toxicol. 2010; 24(3):195–216. [PubMed: 20143452]

Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. Cell. 2012; 148(6):1258–1270. [PubMed: 22424233]

Connolly PH, Caiozzo VJ, Zaldivar F, Nemet D, Larson J, Hung SP, Heck JD, Hatfield GW, Cooper DM. Effects of exercise on gene expression in human peripheral blood mononuclear cells. J Appl Physiol. 2004; 97(4):1461–1469. [PubMed: 15194674]

Costa V, Angelini C, De Feis I, Ciccodicola A. Uncovering the complexity of transcriptomes with RNA-Seq. J Biomed Biotechnol. 2010; 2010:853916. [PubMed: 20625424]

Costa V, Aprile M, Esposito R, Ciccodicola A. RNA-Seq and human complex diseases: recent accomplishments and future perspectives. Eur J Hum Genet. 2013; 21(2):134–142. [PubMed: 22739340]

Cui P, Lin Q, Ding F, Xin C, Gong W, Zhang L, Geng J, Zhang B, Yu X, Yang J, Hu S, Yu J. A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. Genomics. 2010; 96(5):259–265. [PubMed: 20688152]

Cui Y, Paules RS. Use of transcriptomics in understanding mechanisms of drug-induced toxicity. Pharmacogenomics. 2010; 11(4):573–585. [PubMed: 20350139]

Currie RA. Toxicogenomics: the challenges and opportunities to identify biomarkers, signatures and thresholds to support mode-of-action. Mutat Res. 2012; 746(2):97–103. [PubMed: 22445948]

Dakeshita S, Kawai T, Uemura H, Hiyoshi M, Oguma E, Horiguchi H, Kayama F, Aoshima K, Shirahama S, Rokutan K, Arisawa K. Gene expression signatures in peripheral blood cells from Japanese women exposed to environmental cadmium. Toxicology. 2009; 257(1–2):25–32. [PubMed: 19118595]

Davis AP, Murphy CG, Johnson R, Lay JM, Lennon-Hopkins K, Saraceni-Richards C, Sciaky D, King BL, Rosenstein MC, Wiegers TC, Mattingly CJ. The Comparative Toxicogenomics Database: update 2013. Nucleic Acids Res. 2013; 41:D1104–1114. (Database issue). [PubMed: 23093600]

de Mello VD, Kolehmanien M, Schwab U, Pulkkinen L, Uusitupa M. Gene expression of peripheral blood mononuclear cells as a tool in dietary intervention studies: What do we know so far? Mol Nutr Food Res. 2012; 56(7):1160–1172. [PubMed: 22610960]

Debey-Pascher S, Hofmann A, Kreusch F, Schuler G, Schuler-Thurner B, Schultze JL, Staratschek-Jox A. RNA-stabilized whole blood samples but not peripheral blood mononuclear cells can be stored for prolonged time periods prior to transcriptome analysis. J Mol Diagn. 2011; 13(4):452–460. [PubMed: 21704280]

Debey S, Schoenbeck U, Hellmich M, Gathof BS, Pillai R, Zander T, Schultze JL. Comparison of different isolation techniques prior gene expression profiling of blood derived cells: impact on physiological responses, on overall expression and the role of different cell types. Pharmacogenomics J. 2004; 4(3):193–207. [PubMed: 15037859]

Debey S, Zander T, Brors B, Popov A, Eils R, Schultze JL. A highly standardized, robust, and cost-effective method for genome-wide transcriptome analysis of peripheral blood applicable to large-scale clinical trials. Genomics. 2006; 87(5):653–664. [PubMed: 16387473]

DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, Reich M, Winckler W, Getz G. RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics. 2012; 28(11):1530–1532. [PubMed: 22539670]

Drew JE. Cellular defense system gene expression profiling of human whole blood: opportunities to predict health benefits in response to diet. Adv Nutr. 2012; 3(4):499–505. [PubMed: 22797985]

Dutta SK, Mitra PS, Ghosh S, Zang S, Sonneborn D, Hertz-Picciotto I, Trnovec T, Palkovicova L, Sovcikova E, Ghimbovschi S, Hoffman EP. Differential gene expression and a functional analysis of PCB-exposed children: understanding disease and disorder development. Environ Int. 2012; 40:143–154. [PubMed: 21855147]

Enfield KS, Pikor LA, Martinez VD, Lam WL. Mechanistic Roles of Noncoding RNAs in Lung Cancer Biology and Their Clinical Implications. Genet Res Int. 2012; 2012:737416. [PubMed: 22852089]

Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011; 12(12):861–874. [PubMed: 22094949]

Fang Z, Cui X. Design and validation issues in RNA-seq experiments. Brief Bioinform. 2011; 12(3): 280–287. [PubMed: 21498551]

Fannin RD, Russo M, O'Connell TM, Gerrish K, Winnike JH, Macdonald J, Newton J, Malik S, Sieber SO, Parker J, Shah R, Zhou T, Watkins PB, Paules RS. Acetaminophen dosing of humans results in blood transcriptome and metabolome changes consistent with impaired oxidative phosphorylation. Hepatology. 2010; 51(1):227–236. [PubMed: 19918972]

Forrest MS, Lan Q, Hubbard AE, Zhang L, Vermeulen R, Zhao X, Li G, Wu YY, Shen M, Yin S, Chanock SJ, Rothman N, Smith MT. Discovery of novel biomarkers by microarray analysis of peripheral blood mononuclear cell gene expression in benzene-exposed workers. Environ Health Perspect. 2005; 113(6):801–807. [PubMed: 15929907]

Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, Casey G, De Biasi M, Carlson C, Duggan D, James M, Liu P, Tichelaar JW, Vikis HG, You M, Mills IG. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet. 2011; 43(6):513–518. [PubMed: 21614091]

Ganter B, Snyder RD, Halbert DN, Lee MD. Toxicogenomics in drug discovery and development: mechanistic analysis of compound/class-dependent effects using the DrugMatrix database. Pharmacogenomics. 2006; 7(7):1025–1044. [PubMed: 17054413]

Ghaffari SH, Bashash D, Dizaji MZ, Ghavamzadeh A, Alimoghaddam K. Alteration in miRNA gene expression pattern in acute promyelocytic leukemia cell induced by arsenic trioxide: a possible mechanism to explain arsenic multi-target action. Tumour Biol. 2012; 33(1):157–172. [PubMed: 22072212]

Gohlke JM, Thomas R, Zhang Y, Rosenstein MC, Davis AP, Murphy C, Becker KG, Mattingly CJ, Portier CJ. Genetic and environmental pathways to complex diseases. BMC Syst Biol. 2009; 3:46. [PubMed: 19416532]

Goring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, Jowett JB, Abraham LJ, Rainwater DL, Comuzzie AG, Mahaney MC, Almasy L, MacCluer JW, Kissebah AH, Collier GR, Moses EK, Blangero J. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nat Genet. 2007; 39(10):1208–1216. [PubMed: 17873875]

Gower AC, Steiling K, Brothers JF 2nd, Lenburg ME, Spira A. Transcriptomic studies of the airway field of injury associated with smoking-related lung disease. Proc Am Thorac Soc. 2011; 8(2): 173–179. [PubMed: 21543797]

Graff JW, Powers LS, Dickson AM, Kim J, Reisetter AC, Hassan IH, Kremens K, Gross TJ, Wilson ME, Monick MM. Cigarette smoking decreases global microRNA expression in human alveolar macrophages. PLoS One. 2012; 7(8):e44066. [PubMed: 22952876]

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010; 464(7291):1071–1076. [PubMed: 20393566]

Hackett NR, Butler MW, Shaykhiev R, Salit J, Omberg L, Rodriguez-Flores JL, Mezey JG, Strulovici-Barel Y, Wang G, Didon L, Crystal RG. RNA-Seq quantification of the human small airway epithelium transcriptome. BMC Genomics. 2012; 13:82. [PubMed: 22375630]

Han M, Liew CT, Zhang HW, Chao S, Zheng R, Yip KT, Song ZY, Li HM, Geng XP, Zhu LX, Lin JJ, Marshall KW, Liew CC. Novel blood-based, five-gene biomarker set for the detection of colorectal cancer. Clin Cancer Res. 2008; 14(2):455–460. [PubMed: 18203981]

Hansen KD, Brenner SE, Dudoit S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res. 2010; 38(12):e131. [PubMed: 20395217]

Hansen KD, Irizarry RA, Wu Z. Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics. 2012; 13(2):204–216. [PubMed: 22285995]

Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. Nat Rev Genet. 2010; 11(7):476–486. [PubMed: 20531367]

Hebels DG, Georgiadis P, Keun HC, Athersuch TJ, Vineis P, Vermeulen R, Portengen L, Bergdahl IA, Hallmans G, Palli D, Bendinelli B, Krogh V, Tumino R, Sacerdote C, Panico S, Kleinjans JC, de

Kok TM, Smith MT, Kyrtopoulos SA. Performance in Omics Analyses of Blood Samples in Long-Term Storage: Opportunities for the Exploitation of Existing Biobanks in Environmental Health Research. Environ Health Perspect. 2013

Hochstenbach K, van Leeuwen DM, Gmuender H, Gottschalk RW, Lovik M, Granum B, Nygaard U, Namork E, Kirsch-Volders M, Decordier I, Vande Loock K, Besselink H, Tornqvist M, von Stedingk H, Rydberg P, Kleinjans JC, van Loveren H, van Delft JH. Global gene expression analysis in cord blood reveals gender-specific differences in response to carcinogenic exposure in utero. Cancer Epidemiol Biomarkers Prev. 2012; 21(10):1756–1767. [PubMed: 22879202]

Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012; 13:86. [PubMed: 22568884]

Huang J, Shi W, Zhang J, Chou JW, Paules RS, Gerrish K, Li J, Luo J, Wolfinger RD, Bao W, Chu TM, Nikolsky Y, Nikolskaya T, Dosymbekov D, Tsyganova MO, Shi L, Fan X, Corton JC, Chen M, Cheng Y, Tong W, Fang H, Bushel PR. Genomic indicators in the blood predict drug-induced liver injury. Pharmacogenomics J. 2010; 10(4):267–277. [PubMed: 20676066]

Huang R, Jaritz M, Guenzl P, Vlatkovic I, Sommer A, Tamir IM, Marks H, Klampfl T, Kralovics R, Stunnenberg HG, Barlow DP, Pauler FM. An RNA-Seq strategy to detect the complete coding and non-coding transcriptome including full-length imprinted macro ncRNAs. PLoS One. 2011; 6(11):e27288. [PubMed: 22102886]

Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T, Rinn JL. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010; 142(3):409–419. [PubMed: 20673990]

Idaghdour Y, Czika W, Shianna KV, Lee SH, Visscher PM, Martin HC, Miclaus K, Jadallah SJ, Goldstein DB, Wolfinger RD, Gibson G. Geographical genomics of human leukocyte gene expression variation in southern Morocco. Nat Genet. 2010; 42(1):62–67. [PubMed: 19966804]

Idaghdour Y, Storey JD, Jadallah SJ, Gibson G. A genome-wide gene expression signature of environmental geography in leukocytes of Moroccan Amazighs. PLoS Genet. 2008; 4(4):e1000052. [PubMed: 18404217]

Ioannidis JP. Why most published research findings are false. PLoS Med. 2005; 2(8):e124. [PubMed: 16060722]

Jakhesara SJ, Koringa PG, Bhatt VD, Shah TM, Vangipuram S, Shah S, Joshi CG. RNA-Seq reveals differentially expressed isoforms and novel splice variants in buccal mucosal cancer. Gene. 2013; 516(1):24–32. [PubMed: 23266631]

Jardim MJ, Fry RC, Jaspers I, Dailey L, Diaz-Sanchez D. Disruption of microRNA expression in human airway cells by diesel exhaust particles is linked to tumorigenesis-associated pathways. Environ Health Perspect. 2009; 117(11):1745–1751. [PubMed: 20049127]

Jennings P. Stress response pathways, toxicity pathways and adverse outcome pathways. Arch Toxicol. 2013; 87(1):13–14. [PubMed: 23149676]

Jetten MJ, Gaj S, Ruiz-Aracama A, de Kok TM, van Delft JH, Lommen A, van Someren EP, Jennen DG, Claessen SM, Peijnenburg AA, Stierum RH, Kleinjans JC. 'Omics analysis of low dose acetaminophen intake demonstrates novel response pathways in humans. Toxicol Appl Pharmacol. 2012; 259(3):320–328. [PubMed: 22285215]

Jolly C, Lakhotia SC. Human sat III and Drosophila hsr omega transcripts: a common paradigm for regulation of nuclear RNA processing in stressed cells. Nucleic Acids Res. 2006; 34(19):5508–5514. [PubMed: 17020918]

Kampa D, Cheng J, Kapranov P, Yamanaka M, Brubaker S, Cawley S, Drenkow J, Piccolboni A, Bekiranov S, Helt G, Tammana H, Gingeras TR. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res. 2004; 14(3):331–342. [PubMed: 14993201]

Kavanagh T, Mills JD, Kim WS, Halliday GM, Janitz M. Pathway Analysis of the Human Brain Transcriptome in Disease. J Mol Neurosci. 2012

Kawai T, Morita K, Masuda K, Nishida K, Shikishima M, Ohta M, Saito T, Rokutan K. Gene expression signature in peripheral blood cells from medical students exposed to chronic psychological stress. Biol Psychol. 2007; 76(3):147–155. [PubMed: 17766027]

Kiliszek M, Burzynska B, Michalak M, Gora M, Winkler A, Maciejak A, Leszczynska A, Gajda E, Kochanowski J, Opolski G. Altered gene expression pattern in peripheral blood mononuclear cells in patients with acute myocardial infarction. PLoS One. 2012; 7(11):e50054. [PubMed: 23185530]

Kirby J, Heath PR, Shaw PJ, Hamdy FC. Gene expression assays. Adv Clin Chem. 2007; 44:247–292. [PubMed: 17682345]

Kitchen RR, Sabine VS, Simen AA, Dixon JM, Bartlett JM, Sims AH. Relative impact of key sources of systematic noise in Affymetrix and Illumina gene-expression microarray experiments. BMC Genomics. 2011; 12:589. [PubMed: 22133085]

Kohane IS, Valtchinov VI. Quantifying the white blood cell transcriptome as an accessible window to the multiorgan transcriptome. Bioinformatics. 2012; 28(4):538–545. [PubMed: 22219206]

Kong SW, Collins CD, Shimizu-Motohashi Y, Holm IA, Campbell MG, Lee IH, Brewster SJ, Hanson E, Harris HK, Lowe KR, Saada A, Mora A, Madison K, Hundley R, Egan J, McCarthy J, Eran A, Galdzicki M, Rappaport L, Kunkel LM, Kohane IS. Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. PLoS One. 2012; 7(12):e49475. [PubMed: 23227143]

Kossenkov AV, Dawany N, Evans TL, Kucharczuk JC, Albelda SM, Showe LC, Showe MK, Vachani A. Peripheral immune cell gene expression predicts survival of patients with non-small cell lung cancer. PLoS One. 2012; 7(3):e34392. [PubMed: 22479623]

Kostic AD, Ojesina AI, Pedamallu CS, Jung J, Verhaak RG, Getz G, Meyerson M. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. Nat Biotechnol. 2011; 29(5): 393–396. [PubMed: 21552235]

Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982; 38(4):963–974. [PubMed: 7168798]

Lan Q, Zhang L, Li G, Vermeulen R, Weinberg RS, Dosemeci M, Rappaport SM, Shen M, Alter BP, Wu Y, Kopp W, Waidyanatha S, Rabkin C, Guo W, Chanock S, Hayes RB, Linet M, Kim S, Yin S, Rothman N, Smith MT. Hematotoxicity in workers exposed to low levels of benzene. Science. 2004; 306(5702):1774–1776. [PubMed: 15576619]

Lan Q, Zhang L, Tang X, Shen M, Smith MT, Qiu C, Ge Y, Ji Z, Xiong J, He J, Reiss B, Hao Z, Liu S, Xie Y, Guo W, Purdue MP, Galvan N, Xin KX, Hu W, Beane Freeman LE, Blair AE, Li L, Rothman N, Vermeulen R, Huang H. Occupational exposure to trichloroethylene is associated with a decline in lymphocyte subsets and soluble CD27 and CD30 markers. Carcinogenesis. 2010; 31(9):1592–1596. [PubMed: 20530238]

Law PT, Qin H, Ching AK, Lai KP, Co NN, He M, Lung RW, Chan AW, Chan TF, Wong N. Deep sequencing of small RNA transcriptome reveals novel non-coding RNAs in hepatocellular carcinoma. J Hepatol. 2013

Li L, Feng T, Lian Y, Zhang G, Garen A, Song X. Role of human noncoding RNAs in the control of tumorigenesis. Proc Natl Acad Sci U S A. 2009; 106(31):12956–12961. [PubMed: 19625619]

Li X, Shi Y, Wei Y, Ma X, Li Y, Li R. Altered expression profiles of microRNAs upon arsenic exposure of human umbilical vein endothelial cells. Environ Toxicol Pharmacol. 2012; 34(2):381–387. [PubMed: 22728250]

Liew CC, Ma J, Tang HC, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. J Lab Clin Med. 2006; 147(3):126–132. [PubMed: 16503242]

Liong ML, Lim CR, Yang H, Chao S, Bong CW, Leong WS, Das PK, Loh CS, Lau BE, Yu CG, Ooi EJ, Nam RK, Allen PD, Steele GS, Wassmann K, Richie JP, Liew CC. Blood-based biomarkers of aggressive prostate cancer. PLoS One. 2012; 7(9):e45802. [PubMed: 23071848]

Liu R, Ma X, Xu L, Wang D, Jiang X, Zhu W, Cui B, Ning G, Lin D, Wang S. Differential microRNA expression in peripheral blood mononuclear cells from Graves' disease patients. J Clin Endocrinol Metab. 2012; 97(6):E968–972. [PubMed: 22456620]

Liu S, Lin L, Jiang P, Wang D, Xing Y. A comparison of RNA-Seq and high-density exon array for detecting differential gene expression between closely related species. Nucleic Acids Res. 2011; 39(2):578–588. [PubMed: 20864445]

Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, Shchetynsky K, Scheynius A, Kere J, Alfredsson L, Klareskog L, Ekstrom TJ, Feinberg AP. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol. 2013; 31(2):142–147. [PubMed: 23334450]

Lunnon K, Sattlecker M, Furney SJ, Coppola G, Simmons A, Proitsi P, Lupton MK, Lourdusamy A, Johnston C, Soininen H, Kloszewska I, Mecocci P, Tsolaki M, Vellas B, Geschwind D, Lovestone S, Dobson R, Hodges A, dNeuroMed C. A blood gene expression marker of early Alzheimer's disease. J Alzheimers Dis. 2013; 33(3):737–753. [PubMed: 23042217]

Maciejewski JP, Padgett RA. Defects in spliceosomal machinery: a new pathway of leukaemogenesis. Br J Haematol. 2012; 158(2):165–173. [PubMed: 22594801]

MacLean D, Jones JD, Studholme DJ. Application of 'next-generation' sequencing technologies to microbial genetics. Nat Rev Microbiol. 2009; 7(4):287–296. [PubMed: 19287448]

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008; 18(9):1509–1517. [PubMed: 18550803]

Marsit CJ, Eddy K, Kelsey KT. MicroRNA responses to cellular stress. Cancer Res. 2006; 66(22): 10843–10848. [PubMed: 17108120]

Martens-Uzunova ES, Olvedy M, Jenster G. Beyond microRNA – novel RNAs derived from small non-coding RNA and their implication in cancer. Cancer Lett. 2013

Martinez FO. The transcriptome of human monocyte subsets begins to emerge. J Biol. 2009; 8(11):99. [PubMed: 20067595]

Martinez JA, Cordero P, Campion J, Milagro FI. Interplay of early-life nutritional programming on obesity, inflammation and epigenetic outcomes. Proc Nutr Soc. 2012; 71(2):276–283. [PubMed: 22390978]

Mascaux C, Laes JF, Anthoine G, Haller A, Ninane V, Burny A, Sculier JP. Evolution of microRNA expression during human bronchial squamous carcinogenesis. Eur Respir J. 2009; 33(2):352–359. [PubMed: 19010987]

Mattingly CJ, Rosenstein MC, Colby GT, Forrest JN Jr, Boyer JL. The Comparative Toxicogenomics Database (CTD): a resource for comparative toxicological studies. J Exp Zoolog A Comp Exp Biol. 2006; 305(9):689–692.

Maurice CF, Haiser HJ, Turnbaugh PJ. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. Cell. 2013; 152(1–2):39–50. [PubMed: 23332745]

McHale CM, Zhang L, Hubbard AE, Smith MT. Toxicogenomic profiling of chemically exposed humans in risk assessment. Mutat Res. 2010; 705(3):172–183. [PubMed: 20382258]

McHale CM, Zhang L, Hubbard AE, Zhao X, Baccarelli A, Pesatori AC, Smith MT, Landi MT. Microarray analysis of gene expression in peripheral blood mononuclear cells from dioxin-exposed human subjects. Toxicology. 2007; 229(1–2):101–113. [PubMed: 17101203]

McHale CM, Zhang L, Lan Q, Vermeulen R, Li G, Hubbard AE, Porter KE, Thomas R, Portier CJ, Shen M, Rappaport SM, Yin S, Smith MT, Rothman N. Global gene expression profiling of a population exposed to a range of benzene levels. Environ Health Perspect. 2011; 119(5):628–634. [PubMed: 21147609]

McHale CM, Zhang L, Smith MT. Current understanding of the mechanism of benzene-induced leukemia in humans: implications for risk assessment. Carcinogenesis. 2012; 33(2):240–252. [PubMed: 22166497]

Mills JD, Nalpathamkalam T, Jacobs HI, Janitz C, Merico D, Hu P, Janitz M. RNA-Seq analysis of the parietal cortex in Alzheimer's disease reveals alternatively spliced isoforms related to lipid metabolism. Neurosci Lett. 2013; 536:90–95. [PubMed: 23305720]

Min JL, Barrett A, Watts T, Pettersson FH, Lockstone HE, Lindgren CM, Taylor JM, Allen M, Zondervan KT, McCarthy MI. Variability of gene expression profiles in human blood and lymphoblastoid cell lines. BMC Genomics. 2010; 11:96. [PubMed: 20141636]

Misquitta-Ali CM, Cheng E, O'Hanlon D, Liu N, McGlade CJ, Tsao MS, Blencowe BJ. Global profiling and molecular characterization of alternative splicing events misregulated in lung cancer. Mol Cell Biol. 2011; 31(1):138–150. [PubMed: 21041478]

Mitra PS, Ghosh S, Zang S, Sonneborn D, Hertz-Picciotto I, Trnovec T, Palkovicova L, Sovcikova E, Ghimbovschi S, Hoffman EP, Dutta SK. Analysis of the toxicogenomic effects of exposure to persistent organic pollutants (POPs) in Slovakian girls: correlations between gene expression and disease risk. Environ Int. 2012; 39(1):188–199. [PubMed: 22208759]

Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET. Transcriptome genetics using second generation sequencing in a Caucasian population. Nature. 2010; 464(7289):773–777. [PubMed: 20220756]

Moran VA, Perera RJ, Khalil AM. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. Nucleic Acids Res. 2012; 40(14):6391–6400. [PubMed: 22492512]

Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ, Marra MA. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. 2008; 18(4): 610–621. [PubMed: 18285502]

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. Genetic analysis of genome-wide variation in human gene expression. Nature. 2004; 430(7001):743–747. [PubMed: 15269782]

Mutz KO, Heilkenbrinker A, Lonne M, Walter JG, Stahl F. Transcriptome analysis using next-generation sequencing. Curr Opin Biotechnol. 2013; 24(1):22–30. [PubMed: 23020966]

Nath AP, Arafat D, Gibson G. Using blood informative transcripts in geographical genomics: impact of lifestyle on gene expression in fijians. Front Genet. 2012; 3:243. [PubMed: 23162571]

Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, Pettersson S. Host-gut microbiota metabolic interactions. Science. 2012; 336(6086):1262–1267. [PubMed: 22674330]

Nicholson JK, Wilson ID. Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. Nat Rev Drug Discov. 2003; 2(8):668–676. [PubMed: 12904817]

Nuhrenberg TG, Langwieser N, Binder H, Kurz T, Stratz C, Kienzle RP, Trenk D, Zohlnhofer-Momm D, Neumann FJ. Transcriptome analysis in patients with progressive coronary artery disease: identification of differential gene expression in peripheral blood. J Cardiovasc Transl Res. 2013; 6(1):81–93. [PubMed: 23188564]

Pagmantidis V, Meplan C, van Schothorst EM, Keijer J, Hesketh JE. Supplementation of healthy volunteers with nutritionally relevant amounts of selenium increases the expression of lymphocyte protein biosynthesis genes. Am J Clin Nutr. 2008; 87(1):181–189. [PubMed: 18175754]

Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, Holloway E, Kolesnykov N, Lilja P, Lukk M, Mani R, Rayner T, Sharma A, William E, Sarkans U, Brazma A. ArrayExpress–a public database of microarray experiments and gene expression profiles. Nucleic Acids Res. 2007; 35:D747–750. (Database issue). [PubMed: 17132828]

Parkinson H, Sarkans U, Shojatalab M, Abeygunawardena N, Contrino S, Coulson R, Farne A, Lara GG, Holloway E, Kapushesky M, Lilja P, Mukherjee G, Oezcimen A, Rayner T, Rocca-Serra P, Sharma A, Sansone S, Brazma A. ArrayExpress–a public repository for microarray gene expression data at the EBI. Nucleic Acids Res. 2005; 33:D553–555. (Database issue). [PubMed: 15608260]

Peng Z, Cheng Y, Tan BC, Kang L, Tian Z, Zhu Y, Zhang W, Liang Y, Hu X, Tan X, Guo J, Dong Z, Liang Y, Bao L, Wang J. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat Biotechnol. 2012; 30(3):253–260. [PubMed: 22327324]

Peretz A, Peck EC, Bammler TK, Beyer RP, Sullivan JH, Trenga CA, Srinouanprachnah S, Farin FM, Kaufman JD. Diesel exhaust inhalation and assessment of peripheral blood mononuclear cell gene transcription effects: an exploratory study of healthy human volunteers. Inhal Toxicol. 2007; 19(14):1107–1119. [PubMed: 17987463]

Pertea M. The Human Transcriptome: An Unfinished Story. Genes (Basel). 2012; 3(3):344–360. [PubMed: 22916334]

Pertea M, Salzberg SL. Between a chicken and a grape: estimating the number of human genes. Genome Biol. 2010; 11(5):206. [PubMed: 20441615]

Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. Epidemiology. 2006; 17(3):276–284. [PubMed: 16617276]

Pettit AP, Brooks A, Laumbach R, Fiedler N, Wang Q, Strickland PO, Madura K, Zhang J, Kipen HM. Alteration of peripheral blood monocyte gene expression in humans following diesel exhaust inhalation. Inhal Toxicol. 2012; 24(3):172–181. [PubMed: 22369193]

Philibert RA, Sears RA, Powers LS, Nash E, Bair T, Gerke AK, Hassan I, Thomas CP, Gross TJ, Monick MM. Coordinated DNA methylation and gene expression changes in smoker alveolar macrophages: specific effects on VEGF receptor 1 expression. J Leukoc Biol. 2012; 92(3):621–631. [PubMed: 22427682]

Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature. 2010; 464(7289):768–772. [PubMed: 20220758]

Powell JE, Henders AK, McRae AF, Caracella A, Smith S, Wright MJ, Whitfield JB, Dermitzakis ET, Martin NG, Visscher PM, Montgomery GW. The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. PLoS One. 2012; 7(4):e35430. [PubMed: 22563384]

Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, Laxman B, Asangani IA, Grasso CS, Kominsky HD, Cao X, Jing X, Wang X, Siddiqui J, Wei JT, Robinson D, Iyer HK, Palanisamy N, Maher CA, Chinnaiyan AM. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nat Biotechnol. 2011; 29(8):742–749. [PubMed: 21804560]

Raghavachari N, Barb J, Yang Y, Liu P, Woodhouse K, Levy D, O'Donnell CJ, Munson PJ, Kato GJ. A systematic comparison and evaluation of high density exon arrays and RNA-seq technology used to unravel the peripheral blood transcriptome of sickle cell disease. BMC Med Genomics. 2012; 5:28. [PubMed: 22747986]

Rappaport SM. Discovering environmental causes of disease. J Epidemiol Community Health. 2012; 66(2):99–102. [PubMed: 22199396]

Rappaport SM, Smith MT. Epidemiology. Environment and disease risks. Science. 2010; 330(6003): 460–461. [PubMed: 20966241]

Ravasi T, Suzuki H, Pang KC, Katayama S, Furuno M, Okunishi R, Fukuda S, Ru K, Frith MC, Gongora MM, Grimmond SM, Hume DA, Hayashizaki Y, Mattick JS. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Res. 2006; 16(1):11–19. [PubMed: 16344565]

Rederstorff M, Huttenhofer A. Small non-coding RNAs in disease development and host-pathogen interactions. Curr Opin Mol Ther. 2010; 12(6):684–694. [PubMed: 21154160]

Riedmaier I, Pfaffl MW. Transcriptional biomarkers–high throughput screening, quantitative verification, and bioinformatical validation methods. Methods. 2013; 59(1):3–9. [PubMed: 22967906]

Robles JA, Qureshi SE, Stephen SJ, Wilson SR, Burden CJ, Taylor JM. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. BMC Genomics. 2012; 13:484. [PubMed: 22985019]

Rosser CJ, Liu L, Sun Y, Villicana P, McCullers M, Porvasnik S, Young PR, Parker AS, Goodison S. Bladder cancer-associated gene expression signatures identified by profiling of exfoliated urothelia. Cancer Epidemiol Biomarkers Prev. 2009; 18(2):444–453. [PubMed: 19190164]

Rotunno M, Hu N, Su H, Wang C, Goldstein AM, Bergen AW, Consonni D, Pesatori AC, Bertazzi PA, Wacholder S, Shih J, Caporaso NE, Taylor PR, Landi MT. A gene expression signature from peripheral whole blood for stage I lung adenocarcinoma. Cancer Prev Res (Phila). 2011; 4(10): 1599–1608. [PubMed: 21742797]

Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Ison J, Keays M, Kurbatova N, Malone J, Mani R, Mupo A, Pedro Pereira R, Pilicheva E, Rung J, Sharma A, Tang YA, Ternent T, Tikhonov A, Welter D, Williams E, Brazma A, Parkinson H, Sarkans U. ArrayExpress update–trends in database growth and links to data analysis tools. Nucleic Acids Res. 2013; 41:D987–990. (Database issue). [PubMed: 23193272]

Rylander C, Dumeaux V, Olsen KS, Waaseth M, Sandanger TM, Lund E. Using blood gene signatures for assessing effects of exposure to perfluoroalkyl acids (PFAAs) in humans: the NOWAC postgenome study. Int J Mol Epidemiol Genet. 2011; 2(3):207–216. [PubMed: 21915359]

Ryu MS, Langkamp-Henken B, Chang SM, Shankar MN, Cousins RJ. Genomic analysis, cytokine expression, and microRNA profiling reveal biomarkers of human dietary zinc depletion and homeostasis. Proc Natl Acad Sci U S A. 2011; 108(52):20970–20975. [PubMed: 22171008]

Sagaya FM, Hurrell RF, Vergeres G. Postprandial blood cell transcriptomics in response to the ingestion of dairy products by healthy individuals. J Nutr Biochem. 2012; 23(12):1701–1715. [PubMed: 22569349]

Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. Hum Mol Genet. 2010; 19(R2):R227–240. [PubMed: 20858600]

Schembri F, Sridhar S, Perdomo C, Gustafson AM, Zhang X, Ergun A, Lu J, Liu G, Zhang X, Bowers J, Vaziri C, Ott K, Sensinger K, Collins JJ, Brody JS, Getts R, Lenburg ME, Spira A. MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. Proc Natl Acad Sci U S A. 2009; 106(7):2319–2324. [PubMed: 19168627]

Schurmann C, Heim K, Schillert A, Blankenberg S, Carstensen M, Dorr M, Endlich K, Felix SB, Gieger C, Grallert H, Herder C, Hoffmann W, Homuth G, Illig T, Kruppa J, Meitinger T, Muller C, Nauck M, Peters A, Rettig R, Roden M, Strauch K, Volker U, Volzke H, Wahl S, Wallaschofski H, Wild PS, Zeller T, Teumer A, Prokisch H, Ziegler A. Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. PLoS One. 2012; 7(12):e50938. [PubMed: 23236413]

Schwartz S, Oren R, Ast G. Detection and removal of biases in the analysis of next-generation sequencing reads. PLoS One. 2011; 6(1):e16685. [PubMed: 21304912]

Shen-Orr SS, Tibshirani R, Khatri P, Bodian DL, Staedtler F, Perry NM, Hastie T, Sarwal MM, Davis MM, Butte AJ. Cell type-specific gene expression differences in complex tissues. Nat Methods. 2010; 7(4):287–289. [PubMed: 20208531]

Shi L, Campbell G, Jones WD, Campagne F, Wen Z, Walker SJ, Su Z, Chu TM, Goodsaid FM, Pusztai L, Shaughnessy JD Jr, Oberthuer A, Thomas RS, Paules RS, Fielden M, Barlogie B, Chen W, Du P, Fischer M, Furlanello C, Gallas BD, Ge X, Megherbi DB, Symmans WF, Wang MD, Zhang J, Bitter H, Brors B, Bushel PR, Bylesjo M, Chen M, Cheng J, Cheng J, Chou J, Davison TS, Delorenzi M, Deng Y, Devanarayan V, Dix DJ, Dopazo J, Dorff KC, Elloumi F, Fan J, Fan S, Fan X, Fang H, Gonzaludo N, Hess KR, Hong H, Huan J, Irizarry RA, Judson R, Juraeva D, Lababidi S, Lambert CG, Li L, Li Y, Li Z, Lin SM, Liu G, Lobenhofer EK, Luo J, Luo W, McCall MN, Nikolsky Y, Pennello GA, Perkins RG, Philip R, Popovici V, Price ND, Qian F, Scherer A, Shi T, Shi W, Sung J, Thierry-Mieg D, Thierry-Mieg J, Thodima V, Trygg J, Vishnuvajjala L, Wang SJ, Wu J, Wu Y, Xie Q, Yousef WA, Zhang L, Zhang X, Zhong S, Zhou Y, Zhu S, Arasappan D, Bao W, Lucas AB, Berthold F, Brennan RJ, Buness A, Catalano JG, Chang C, Chen R, Cheng Y, Cui J, Czika W, Demichelis F, Deng X, Dosymbekov D, Eils R, Feng Y, Fostel J, Fulmer-Smentek S, Fuscoe JC, Gatto L, Ge W, Goldstein DR, Guo L, Halbert DN, Han J, Harris SC, Hatzis C, Herman D, Huang J, Jensen RV, Jiang R, Johnson CD, Jurman G, Kahlert Y, Khuder SA, Kohl M, Li J, Li L, Li M, Li QZ, Li S, Li Z, Liu J, Liu Y, Liu Z, Meng L, Madera M, Martinez-Murillo F, Medina I, Meehan J, Miclaus K, Moffitt RA, Montaner D, Mukherjee P, Mulligan GJ, Neville P, Nikolskaya T, Ning B, Page GP, Parker J, Parry RM, Peng X, Peterson RL, Phan JH, Quanz B, Ren Y, Riccadonna S, Roter AH, Samuelson FW, Schumacher MM, Shambaugh JD, Shi Q, Shippy R, Si S, Smalter A, Sotiriou C, Soukup M, Staedtler F, Steiner G, Stokes TH, Sun Q, Tan PY, Tang R, Tezak Z, Thorn B, Tsyganova M, Turpaz Y, Vega SC, Visintainer R, von Frese J, Wang C, Wang E, Wang J, Wang W, Westermann F, Willey JC, Woods M, Wu S, Xiao N, Xu J, Xu L, Yang L, Zeng X, Zhang J, Zhang L, Zhang M, Zhao C, Puri RK, Scherf U, Tong W, Wolfinger RD, Consortium M. The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. Nat Biotechnol. 2010; 28(8):827–838. [PubMed: 20676074]

Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY, Luo Y, Sun YA, Willey JC, Setterquist RA, Fischer GM, Tong W, Dragan YP, Dix DJ, Frueh FW, Goodsaid FM, Herman D, Jensen RV, Johnson CD, Lobenhofer EK, Puri RK, Schrf U, Thierry-Mieg J, Wang C, Wilson M, Wolber PK, Zhang L, Amur S, Bao W, Barbacioru CC, Lucas AB, Bertholet V, Boysen C, Bromley B, Brown D, Brunner A, Canales R,

Cao XM, Cebula TA, Chen JJ, Cheng J, Chu TM, Chudin E, Corson J, Corton JC, Croner LJ, Davies C, Davison TS, Delenstarr G, Deng X, Dorris D, Eklund AC, Fan XH, Fang H, Fulmer-Smentek S, Fuscoe JC, Gallagher K, Ge W, Guo L, Guo X, Hager J, Haje PK, Han J, Han T, Harbottle HC, Harris SC, Hatchwell E, Hauser CA, Hester S, Hong H, Hurban P, Jackson SA, Ji H, Knight CR, Kuo WP, LeClerc JE, Levy S, Li QZ, Liu C, Liu Y, Lombardi MJ, Ma Y, Magnuson SR, Maqsodi B, McDaniel T, Mei N, Myklebost O, Ning B, Novoradovskaya N, Orr MS, Osborn TW, Papallo A, Patterson TA, Perkins RG, Peters EH, Peterson R, Philips KL, Pine PS, Pusztai L, Qian F, Ren H, Rosen M, Rosenzweig BA, Samaha RR, Schena M, Schroth GP, Shchegrova S, Smith DD, Staedtler F, Su Z, Sun H, Szallasi Z, Tezak Z, Thierry-Mieg D, Thompson KL, Tikhonova I, Turpaz Y, Vallanat B, Van C, Walker SJ, Wang SJ, Wang Y, Wolfinger R, Wong A, Wu J, Xiao C, Xie Q, Xu J, Yang W, Zhang L, Zhong S, Zong Y, Slikker W Jr. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol. 2006; 24(9):1151–1161. [PubMed: 16964229]

Showe MK, Kossenkov AV, Showe LC. The peripheral immune response and lung cancer prognosis. Oncoimmunology. 2012; 1(8):1414–1416. [PubMed: 23243612]

Showe MK, Vachani A, Kossenkov AV, Yousef M, Nichols C, Nikonova EV, Chang C, Kucharczuk J, Tran B, Wakeam E, Yie TA, Speicher D, Rom WN, Albelda S, Showe LC. Gene expression profiles in peripheral blood mononuclear cells can distinguish patients with non-small cell lung cancer from patients with nonmalignant lung disease. Cancer Res. 2009; 69(24):9202–9210. [PubMed: 19951989]

Sie CP, Kuchka M. RNA editing adds flavor to complexity. Biochemistry (Mosc). 2011; 76(8):869–881. [PubMed: 22022960]

Silva JM, Perez DS, Pritchett JR, Halling ML, Tang H, Smith DI. Identification of long stress-induced non-coding transcripts that have altered expression in cancer. Genomics. 2010; 95(6):355–362. [PubMed: 20214974]

Singh S, Li SS. Phthalates: toxicogenomics and inferred human diseases. Genomics. 2011; 97(3):148–157. [PubMed: 21156202]

Sinisi SE, Polley EC, Petersen ML, Rhee SY, van der Laan MJ. Super learning: an application to the prediction of HIV-1 drug resistance. Stat Appl Genet Mol Biol. 2007; 6:Article7. [PubMed: 17402922]

Smirnova L, Sittka A, Luch A. On the role of low-dose effects and epigenetics in toxicology. EXS. 2012; 101:499–550. [PubMed: 22945581]

Smith MT, Zhang L, McHale CM, Skibola CF, Rappaport SM. Benzene, the exposome and future investigations of leukemia etiology. Chem Biol Interact. 2011; 192(1–2):155–159. [PubMed: 21333640]

Sonkoly E, Bata-Csorgo Z, Pivarcsi A, Polyanka H, Kenderessy-Szabo A, Molnar G, Szentpali K, Bari L, Megyeri K, Mandi Y, Dobozy A, Kemeny L, Szell M. Identification and characterization of a novel, psoriasis susceptibility-related noncoding RNA gene, PRINS. J Biol Chem. 2005; 280(25): 24159–24167. [PubMed: 15855153]

Spira A, Beane J, Pinto-Plata V, Kadar A, Liu G, Shah V, Celli B, Brody JS. Gene expression profiling of human lung tissue from smokers with severe emphysema. Am J Respir Cell Mol Biol. 2004a; 31(6):601–610. [PubMed: 15374838]

Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, Palma J, Brody JS. Effects of cigarette smoke on the human airway epithelial cell transcriptome. Proc Natl Acad Sci U S A. 2004b; 101(27): 10143–10148. [PubMed: 15210990]

Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, Gilman S, Dumas YM, Calner P, Sebastiani P, Sridhar S, Beamis J, Lamb C, Anderson T, Gerry N, Keane J, Lenburg ME, Brody JS. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. Nat Med. 2007; 13(3):361–366. [PubMed: 17334370]

Spira A, Schembri F, Beane J, Shah V, Liu G, Brody JS. Impact of cigarette smoke on the normal airway transcriptome. Chest. 2004c; 125(5 Suppl):115S.

Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. Nat Rev Microbiol. 2011; 9(4):279–290. [PubMed: 21407244]

Sridhar S, Schembri F, Zeskind J, Shah V, Gustafson AM, Steiling K, Liu G, Dumas YM, Zhang X, Brody JS, Lenburg ME, Spira A. Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. BMC Genomics. 2008; 9:259. [PubMed: 18513428]

Stoynev N, Dimova I, Rukova B, Hadjidekova S, Nikolova D, Toncheva D, Tankova T. Gene expression in peripheral blood of patients with hypertension and patients with type 2 diabetes. J Cardiovasc Med (Hagerstown). 2013

Su Z, Li Z, Chen T, Li QZ, Fang H, Ding D, Ge W, Ning B, Hong H, Perkins RG, Tong W, Shi L. 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(9):1486–1493. [PubMed: 21834575]

Szyf M. The dynamic epigenome and its implications in toxicology. Toxicol Sci. 2007; 100(1):7–23. [PubMed: 17675334]

Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS. Non-coding RNAs: regulators of disease. J Pathol. 2010; 220(2):126–139. [PubMed: 19882673]

Tan XL, Wang T, Xiong S, Kumar SV, Han W, Spivack SD. Smoking-Related Gene Expression in Laser Capture-Microdissected Human Lung. Clin Cancer Res. 2009; 15(24):7562–7570. [PubMed: 19996203]

Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. Genome Res. 2011; 21(12):2213–2223. [PubMed: 21903743]

Thomas R, Cliona M McHale, Luoping Zhang, Qing Lan, Alan E Hubbard, Roel Vermeulen, Guilan Li, Stephen M Rappaport, Songnian Yin, Martyn T Smith, Nathaniel Rothman. Global gene expression response of a population exposed to benzene: a pilot study exploring the use of RNA sequencing technology. Environmental and Molecular Mutagenesis. 2013 IN SUBMISSION.

Thomas R, Phuong J, McHale CM, Zhang L. Using bioinformatic approaches to identify pathways targeted by human leukemogens. Int J Environ Res Public Health. 2012; 9(7):2479–2503. [PubMed: 22851955]

Tian Z, Palmer N, Schmid P, Yao H, Galdzicki M, Berger B, Wu E, Kohane IS. A practical platform for blood biomarker study by using global gene expression profiling of peripheral whole blood. PLoS One. 2009; 4(4):e5157. [PubMed: 19381341]

Tilley AE, Harvey BG, Heguy A, Hackett NR, Wang R, O'Connor TP, Crystal RG. Down-regulation of the notch pathway in human airway epithelium in association with smoking and chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2009; 179(6):457–466. [PubMed: 19106307]

Tilley AE, O'Connor TP, Hackett NR, Strulovici-Barel Y, Salit J, Amoroso N, Zhou XK, Raman T, Omberg L, Clark A, Mezey J, Crystal RG. Biologic phenotyping of the human small airway epithelial response to cigarette smoking. PLoS One. 2011; 6(7):e22798. [PubMed: 21829517]

Tondeur S, Pangault C, Le Carrour T, Lannay Y, Benmahdi R, Cubizolle A, Assou S, Pantesco V, Klein B, Hamamah S, Schved JF, Fest T, De Vos J. Expression map of the human exome in CD34+ cells and blood cells: increased alternative splicing in cell motility and immune response genes. PLoS One. 2010; 5(2):e8990. [PubMed: 20126548]

Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG, Prasanth KV. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell. 2010; 39(6):925–938. [PubMed: 20797886]

Uehara T, Ono A, Maruyama T, Kato I, Yamada H, Ohno Y, Urushidani T. The Japanese toxicogenomics project: application of toxicogenomics. Mol Nutr Food Res. 2010; 54(2):218–227. [PubMed: 20041446]

Urquidi V, Goodison S, Cai Y, Sun Y, Rosser CJ. A candidate molecular biomarker panel for the detection of bladder cancer. Cancer Epidemiol Biomarkers Prev. 2012; 21(12):2149–2158. [PubMed: 23097579]

van Delft J, Gaj S, Lienhard M, Albrecht MW, Kirpiy A, Brauers K, Claessen S, Lizarraga D, Lehrach H, Herwig R, Kleinjans J. RNA-Seq provides new insights in the transcriptome responses

induced by the carcinogen benzo[a]pyrene. Toxicol Sci. 2012; 130(2):427–439. [PubMed: 22889811]

Vedin I, Cederholm T, Freund-Levi Y, Basun H, Garlind A, Irving GF, Eriksdotter-Jonhagen M, Wahlund LO, Dahlman I, Palmblad J. Effects of DHA-rich n-3 fatty acid supplementation on gene expression in blood mononuclear leukocytes: the OmegAD study. PLoS One. 2012; 7(4):e35425. [PubMed: 22545106]

Vineis P, Airoldi L, Veglia F, Olgiati L, Pastorelli R, Autrup H, Dunning A, Garte S, Gormally E, Hainaut P, Malaveille C, Matullo G, Peluso M, Overvad K, Tjonneland A, Clavel-Chapelon F, Boeing H, Krogh V, Palli D, Panico S, Tumino R, Bueno-De-Mesquita B, Peeters P, Berglund G, Hallmans G, Saracci R, Riboli E. Environmental tobacco smoke and risk of respiratory cancer and chronic obstructive pulmonary disease in former smokers and never smokers in the EPIC prospective study. BMJ. 2005; 330(7486):277. [PubMed: 15681570]

Votavova H, Dostalova Merkerova M, Fejglova K, Vasikova A, Krejcik Z, Pastorkova A, Tabashidze N, Topinka J, Veleminsky M Jr, Sram RJ, Brdicka R. Transcriptome alterations in maternal and fetal cells induced by tobacco smoke. Placenta. 2011; 32(10):763–770. [PubMed: 21803418]

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008; 456(7221): 470–476. [PubMed: 18978772]

Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009; 10(1):57–63. [PubMed: 19015660]

Wang Z, Neuburg D, Li C, Su L, Kim JY, Chen JC, Christiani DC. Global gene expression profiling in whole-blood samples from individuals exposed to metal fumes. Environ Health Perspect. 2005; 113(2):233–241. [PubMed: 15687063]

Wang Z, Zhao Y, Smith E, Goodall GJ, Drew PA, Brabletz T, Yang C. Reversal and prevention of arsenic-induced human bronchial epithelial cell malignant transformation by microRNA-200b. Toxicol Sci. 2011; 121(1):110–122. [PubMed: 21292642]

Waters M, Stasiewicz S, Merrick BA, Tomer K, Bushel P, Paules R, Stegman N, Nehls G, Yost KJ, Johnson CH, Gustafson SF, Xirasagar S, Xiao N, Huang CC, Boyer P, Chan DD, Pan Q, Gong H, Taylor J, Choi D, Rashid A, Ahmed A, Howle R, Selkirk J, Tennant R, Fostel J. CEBS–Chemical Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. Nucleic Acids Res. 2008a; 36:D892–900. (Database issue). [PubMed: 17962311]

Waters M, Stasiewicz S, Merrick BA, Tomer K, Bushel P, Paules R, Stegman N, Nehls G, Yost KJ, Johnson CH, Gustafson SF, Xirasagar S, Xiao N, Huang CC, Boyer P, Chan DD, Pan Q, Gong H, Taylor J, Choi D, Rashid A, Ahmed A, Howle R, Selkirk J, Tennant R, Fostel J. CEBS Chemical Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. Nucleic Acids Res. 2008b; 36:D892–900. (Database issue). [PubMed: 17962311]

Watkins NA, Gusnanto A, de Bono B, De S, Miranda-Saavedra D, Hardie DL, Angenent WG, Attwood AP, Ellis PD, Erber W, Foad NS, Garner SF, Isacke CM, Jolley J, Koch K, Macaulay IC, Morley SL, Rendon A, Rice KM, Taylor N, Thijssen-Timmer DC, Tijssen MR, van der Schoot CE, Wernisch L, Winzer T, Dudbridge F, Buckley CD, Langford CF, Teichmann S, Gottgens B, Ouwehand WH, Bloodomics C. A HaemAtlas: characterizing gene expression in differentiated human blood cells. Blood. 2009; 113(19):e1–9. [PubMed: 19228925]

Weber DG, Casjens S, Rozynek P, Lehnert M, Zilch-Schoneweis S, Bryk O, Taeger D, Gomolka M, Kreuzer M, Otten H, Pesch B, Johnen G, Bruning T. Assessment of mRNA and microRNA Stabilization in Peripheral Human Blood for Multicenter Studies and Biobanks. Biomark Insights. 2010; 5:95–102. [PubMed: 20981139]

Wild CP. Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev. 2005; 14(8):1847–1850. [PubMed: 16103423]

Wu X, Song Y. Preferential regulation of miRNA targets by environmental chemicals in the human genome. BMC Genomics. 2011; 12:244. [PubMed: 21592377]

Xi L, Feber A, Gupta V, Wu M, Bergemann AD, Landreneau RJ, Litle VR, Pennathur A, Luketich JD, Godfrey TE. Whole genome exon arrays identify differential expression of alternatively spliced,

cancer-related genes in lung cancer. Nucleic Acids Res. 2008; 36(20):6535–6547. [PubMed: 18927117]

Zander T, Hofmann A, Staratschek-Jox A, Classen S, Debey-Pascher S, Maisel D, Ansen S, Hahn M, Beyer M, Thomas RK, Gathof B, Mauch C, Delank KS, Engel-Riedel W, Wichmann HE, Stoelben E, Schultze JL, Wolf J. Blood-based gene expression signatures in non-small cell lung cancer. Clin Cancer Res. 2011; 17(10):3360–3367. [PubMed: 21558400]

Zeskind JE, Lenburg ME, Spira A. Translating the COPD transcriptome: insights into pathogenesis and tools for clinical management. Proc Am Thorac Soc. 2008; 5(8):834–841. [PubMed: 19017738]

Zhang L, Bushel PR, Chou J, Zhou T, Watkins PB. Identification of Identical Transcript Changes in Liver and Whole Blood during Acetaminophen Toxicity. Front Genet. 2012; 3:162. [PubMed: 22973295]

Zhang L, Lee JJ, Tang H, Fan YH, Xiao L, Ren H, Kurie J, Morice RC, Hong WK, Mao L. Impact of smoking cessation on global gene expression in the bronchial epithelium of chronic smokers. Cancer Prev Res (Phila). 2008; 1(2):112–118. [PubMed: 19138944]

Zhang X, Sebastiani P, Liu G, Schembri F, Zhang X, Dumas YM, Langer EM, Alekseyev Y, O'Connor GT, Brooks DR, Lenburg ME, Spira A. Similarities and differences between smoking-related gene expression in nasal and bronchial epithelium. Physiol Genomics. 2010; 41(1):1–8. [PubMed: 19952278]

Zhou X, Ren L, Meng Q, Li Y, Yu Y, Yu J. The next-generation sequencing technology and application. Protein Cell. 2010; 1(6):520–536. [PubMed: 21204006]

Zieker D, Fehrenbach E, Dietzsch J, Fliegner J, Waidmann M, Nieselt K, Gebicke-Haerter P, Spanagel R, Simon P, Niess AM, Northoff H. cDNA microarray analysis reveals novel candidate genes expressed in human peripheral blood following exhaustive exercise. Physiol Genomics. 2005; 23(3):287–294. [PubMed: 16118270]

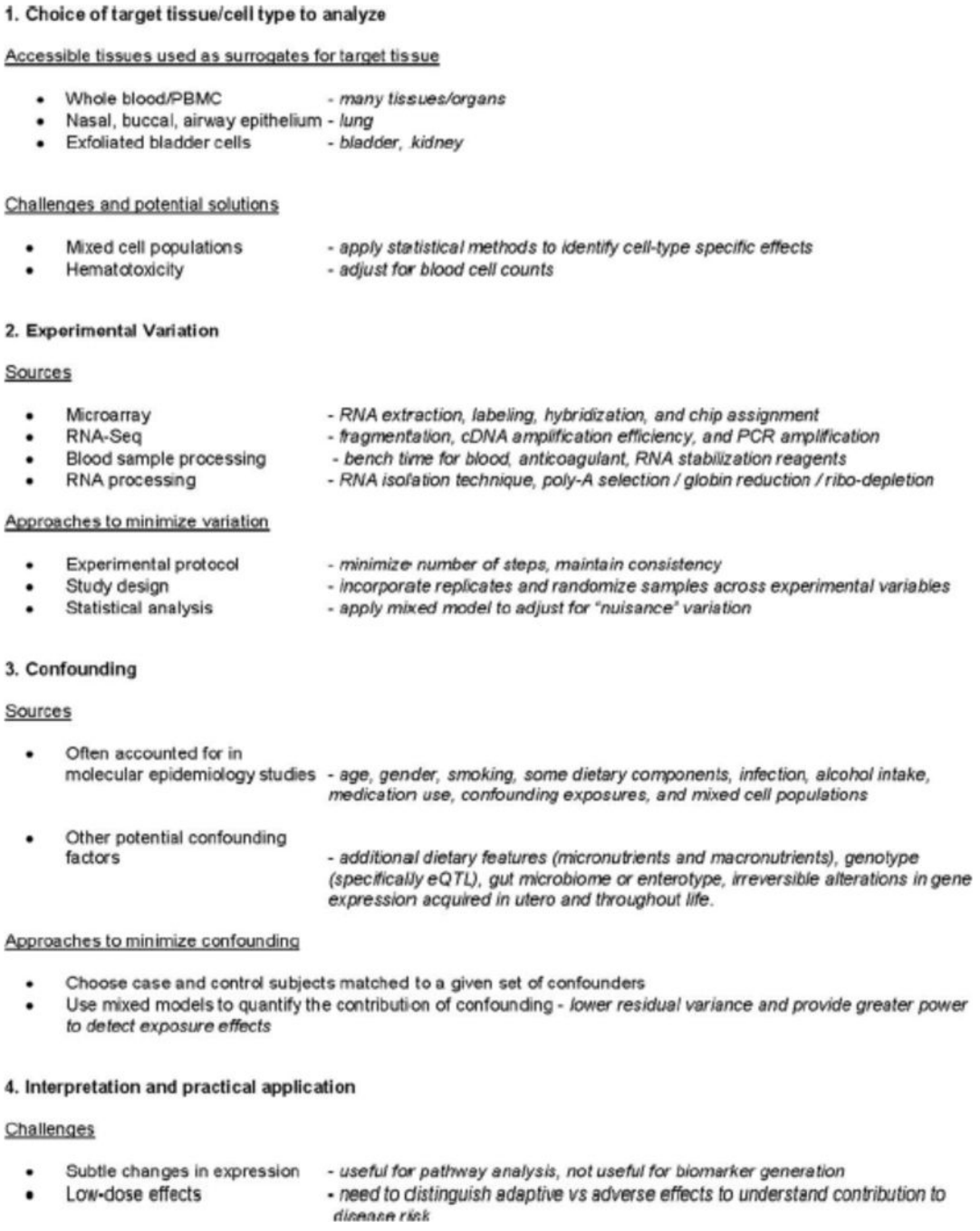

Fig. 1.

Consideration in transcriptome study design, analysis, and interpretation in molecular epidemiology studies.