예방의학회지 제42권 제6호(2009년 11월)
J Prev Med Public Health 2009;42(6):349-355
DOI: 10.3961/jpmph.2009.42.6.349

Special Issues

Identification and Application of Biomarkers in Molecular and Genomic Epidemiologic Research

Kyoung-Mu Lee[1,2], Sohee Han[2], Woong-Yang Park[3], Daehee Kang[2,4,5]

Clinical Research Institute, Seoul National University Hospital[1]; Department of Preventive Medicine, Seoul National University College of Medicine[2]; Department of Biochemistry, Seoul National University College of Medicine[3]; Cancer Research Institute, Seoul National University College of Medicine[4]; Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science, Seoul National University[5]

Biomarkers are characteristic biological properties that can be detected and measured in a variety of biological matrices in the human body, including the blood and tissue, to give an indication of whether there is a threat of disease, if a disease already exists, or how such a disease may develop in an individual case. Along the continuum from exposure to clinical disease and progression, exposure, internal dose, biologically effective dose, early biological effect, altered structure and/or function, clinical disease, and disease progression can potentially be observed and quantified using biomarkers. While the traditional discovery of biomarkers has been a slow process, the advent of molecular and genomic medicine has resulted in explosive growth in the discovery of new biomarkers. In this review, issues in evaluating biomarkers will be discussed and the biomarkers of environmental exposure, early biologic effect, and susceptibility identified and validated in epidemiological studies will be summarized. The spectrum of genomic approaches currently used to identify and apply biomarkers and strategies to validate genomic biomarkers will also be discussed.

J Prev Med Public Health 2009;42(6):349-355

Key words : Biomarker, Molecular epidemiology, Genomic epidemiology, Validity, Exposure

INTRODUCTION

Biomarkers are characteristic biological properties that can be detected and measured in a variety of biological matrices in the human body, including the blood, urine and tissue, to give an indication of whether there is a threat of disease, if a disease already exists, or how such a disease may develop in an individual case [1]. The definition of a biomarker has been expanded to include pharmacologic responses to therapeutic intervention by the Biomarkers Consortium (http://www.biomarkersconsor-tium.org).

Molecular epidemiology is a branch of epidemiology that uses biomarkers to deal with the contribution of potential genetic and environmental risk factors identified at the molecular level, to the etiology, distribution and control of the disease in human populations. Since the phrase “molecular epidemiology” was formalized in the first book by Schulte and Perera [1], molecular epidemiology has improved our understanding of specific pathways, molecules and susceptibility genes that influence the risk of developing disease.

With the completion of the Human Genome Project [2] and the HapMap Project (www.hapmap.org, 2005), genomic epidemiology (also known as, human genome epidemiology) has been established as “an evolving field of inquiring that uses the systematic application of epidemiologic methods and approaches in population-based studies of the impact of human genetic variation on health and disease” (http://www.cdc.gov/genomics/hugenet) [3]. Molecular epidemiology, in its broader definition, includes the scope of human genome epidemiology, which focuses on genetic variation, one type of biomarker, and its interaction with environmental risk factors in the development of human disease. With the recent advance of various types of high-throughput technology in the area of genomics, epigenomics, and proteomics, the scope of molecular epidemiology and genomic epidemiology has expanded rapidly.

In this review, issues in evaluating biomarkers will be discussed, and biomarkers of environmental exposure, early biologic effect, and susceptibility identified and validated in epidemiological studies will be summarized. The spectrum of genomic approaches to identify and apply biomarkers and strategies to validate genomic biomarkers is then discussed.

CLASSIFICATION OF BIOMARKERS

In molecular epidemiology, biomarkers have usually been classified into three categories: exposure, effect (disease), and susceptibility [3]

Received : 2 October 2009, Accepted : 2 November 2009
Corresponding author : Daehee Kang (28 Yongon-Dong, Chongno-Gu, Seoul , Korea, Tel : 82-2-740-8326, Fax : 82-2-747-4830, E-mail : dhkang@snu.ac.kr)

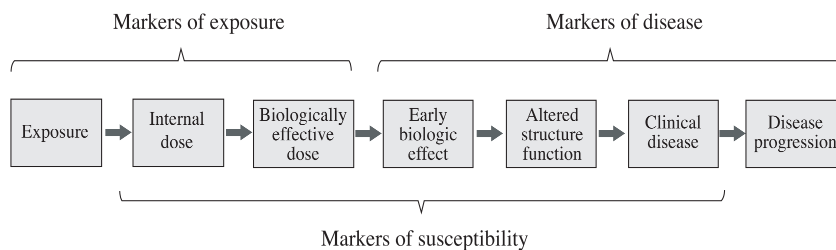

Figure 1. Types of biomarkers: exposure, disease and susceptibility biomarkers along the continuum from exposure to clinical disease and progression (modified from Schulte and Perera [1]).

(Figure 1). Along the continuum from biomarkers of exposure to clinical disease and progression, exposure, internal dose, biologically effective dose, early biologic effect, altered structure and/or function, clinical disease, and disease progression can potentially be observed and quantified.

Biomarkers of exposure include exogenous chemicals or metabolite(s), internal dosimeters of the chemical or metabolite concentrations and biologically effective doses. These markers may integrate multiple portals of entry and fluctuating exposures, and relate the time of exposure to the internal dose [4]. Biomarkers of exposure are important in biological monitoring, because they may be an indicator of internal dose, or the amount of chemical exposure into the body. Significant advances have been made in the development of analytical methods which can detect and/or quantify the presence of many natural or synthetic toxins or their breakdown products (metabolites) in a biological matrix (such as blood or urine). It is note worthy that a adequate understanding of the chemistry and toxicology of the substance under consideration is required to measure biomarkers of exposure accurately.

Biomarkers of effects include the measurable alterations of an organism that can indicate a potential or established health impairment or disease (e.g., blood pressure, cholesterol levels and viral load in HIV). Markers of altered structure and/or function are useful for assessing morphological and/or functional changes following toxicant-cell interactions. Recently, a number of studies were conducted to discover biomarkers for prognosis as well as diagnosis; thus, biomarkers of prognosis may be considered as a separate category.

Biomarkers of susceptibility are indicators of inherent or acquired properties that may lead to an increase in the internal dose of a chemical or an increased level of the response resulting from environmental exposure. Examples of susceptibility biomarkers are polymorphisms in genes involved in carcinogen metabolism or DNA repair [5-6].

EVALUATION OF BIOMARKERS

In order to be useful, biomarkers must be rigorously tested prior to application in risk assessment, disease diagnosis and predictions of prognosis. Given that the combination of a number of factors limits useful applications of biomarkers, we summarized the general criteria that should be taken into account when evaluating biomarkers.

I. Biological Relevance

A useful biomarker needs to have biological relevance. Typically, limited information is available regarding where the markers are located along the multi-step pathway of human disease. Evidence from experimental studies (*in vitro* and *in vivo*) on exposure-effect relationships can sometimes provide certain levels of insight into the evaluation of biomarkers [7].

II. Biological Sample and Assay

An easily attainable sample is one that can be obtained with limited stringency of preparation. Urine is one of readily useable samples in a non-invasive way, and blood can be a good source for biomarker information because it is exposed directly to all organs of the body. Therefore urine and blood samples are collected by almost all ongoing prospective cohort studies.

The detection method for a biomarker must be accurate, rapid and as easy to carry out as possible. Levels of exposures and biomarkers may be below the analytical limits of detection.

The results from different laboratories should not differ significantly from each other. Reliability is often described as the coefficient of variation between repeated measurements of the same samples.

Determining the specificity and sensitivity are critical components of the evaluation process of biomarkers. Specificity refers to the ability of a measurement to identify negative responses effectively in order to minimize the number of false positives, while sensitivity refers to the ability of a measurement to detect positive responses.

A biomarker observed well before the onset of disease may have a low predictive value as a biomarker of effect but may be very useful as a biomarker of exposure, enabling long-term surveillance of an exposed population. In contrast, a biomarker of effect that is expressed long after exposure could be of relatively little use in exposure assessment whereas it may be very useful in calculating risk of disease and predicting the progression.

III. Population and Clinical Validity

Factors related to the population validity of biomarkers include the frequency or prevalence of the biomarkers, greater inter-individual variation than intra-individual variation, potential confounders, and invasiveness of specimen collection [7]. Information on conditions under which samples were collected, stored, and analyzed (e.g., fasting status, season of blood sampling, batch of the assay, differential handling of specimens from cases and control, etc.) is also

Table 1. Selected cancer biomarkers of exposure and biologically effect dose validated in epidemiological studies

Biomarker	Cancer	Study type	Subjects (Studies)	Exposure/Biological matrix	OR/RR/HR/SIR (CI or p-value)	Reference
Environmental exposure						
Phenoxy herbicides	Leukemia	MA	12 cohort studies	Occupational exposure in plants manufacturing pesticides; serum, blood, fat and/or urine	RR = 1.60 (1.02-2.52) ; [6.99 (1.96-24.90) for myeloid leukemia]	Van et al., [8]
	Prostate	MA	16 cohort studies	Occupational exposure for the manufacturing workers; serum, blood, fat and/or urine	RR = 1.80 (1.03-3.13)	Van et al., [9]
Internal dose						
Cotinine	Lung	CC	1,741 Ca/ 1,741 Con	Average log-transformed cotinine levels; serum	$OR_{>378.8\ ng/mL.\ vs.\ 0.1\text{-}5.0}$ = 55.1 (35.7-85.0)	Boffetta et al., [10]
Hepatitis C virus (HCV)	Liver	C	172 within a cohort of 22,073	Blood	$SIR_{HCV\ vs.\ non\text{-}HCV}$ = 1.27 (1.02-1.56)	McDonald et al., [11]
	Non-Hodgkin Lymphoma (NHL)	RC	146,394 HCV / 572,293 non-HCV	Blood	$HR_{HCV\ vs.\ non\text{-}HCV}$ = 1.28 (1.12-1.45)	Giordano et al., [12]
HPV-16	Head and neck (squamous)	NCC	292 Ca / 1,568 Con within a cohort of 900,000	Blood	$OR_{HPV\text{-}16\ vs.\ non\text{-}HPV\text{-}16}$ = 2.2 (1.4-3.4)	Mork et al., [13]
Biologically effective dose						
Aflatoxin B1-albumin adducts	Liver	NCC	230 Ca / 1,052 Con (cohort: n=23,943)	Concentrations of AFB1-albumin adducts (>59.8 fmol/mg [median] vs. ≤59.8); serum	OR= 1.54 (1.01-2.36)	Wu et al., [14]
PAH-DNA adducts	Colorectal adenoma	CC	82 Ca / 111 Con	Mean (±SD) values and quartile category of PAH adduct levels (per 108 nucleotides); leukocyte	$OR_{continuous}$ = 1.5 (1.1-2.2) $OR_{quartile\ 4\ (>1.52)\ vs.\ quartile1\ (\leq 0.71)}$= 2.8 (1.2-6.5); p_{trend} = 0.048	Gunter et al., [15]

CC: case-control, C: cohort, RC: retrospective cohort, PC: prospective cohort, NCC: nested case-control, MA: meta-analysis, PA: pooled analysis, Ca: case, Con: control, RR: relative risk, OR: odds ratio, HR: hazard ratio, SIR: standard incidence ratio, CI: confidence interval.

important [16]. This information can possibly be used for adjustment or stratification in the statistical analysis in order to minimize measurement error.

In clinical settings, biomarkers must to be evaluated in terms of validity in diagnosis and prognosis. Based on the evaluation, appropriate and effective clinical intervention can be pursued through early diagnosis and intervention.

BIOMARKERS VALIDATED IN EPIDEMIOLOGICAL STUDIES

Validation of biomarkers is a lengthy process that takes years of laboratory work and inter-laboratory collaborations. Well-defined biomarkers that are ultimately used in risk appraisal and in the clinics are likely to be in small numbers. Only a few biomarkers related to research on cancer have been validated in prospective cohort studies.

Selected biomarkers of exposure, early biologic effect, and susceptibility evaluated and validated in epidemiological studies are listed in Table 1, Table 2, and Table 3, respectively.

GENOMIC APPROACHES TO IDENTIFY AND APPLY BIOMARKERS

A paradigm shift has recently been realized, whereby single-biomarker analysis is being replaced by multiparametric analyses of genes or proteins [17]. Various genomics-based approaches are being successfully used to identify and apply biomarkers in molecular and genomic epidemiology. Until recently, the discovery of cancer biomarkers has been a slow approach to identify proteins that are dysregulated as a consequence of the disease and shed into body fluids such as serum, urine, or saliva [18]. The recent advancements in genomic and proteomic technologies including gene array technology coupled with advancements in bioinformatics tools, shows great promise of meeting the demand for the discovery of a variety of new biomarkers that are both sensitive and specific.

Genomic approaches could unseal the molecular mechanisms or biological pathways of various physiological conditions [19,20] as well as human diseases [21,22]. Molecular profiling via genomics, proteomics, and metabolomics has opened new windows to study disease etiology and prognosis.

SPECTRUM OF GENOMIC BIOMARKERS

Although DNA microarray technologies were launched initially to analyze mRNA, other powerful tools for global analyses of cellular constituents are now available. These include technologies for global analyses of polymorphisms and proteins [23], and cellular metabolites [24]. The use of proteomic and metabolomic tools in combination with genomic techniques will be important for obtaining a comprehensive picture of physiological and pathological changes in cell, tissue and organs.

I. Single Nucleotide Polymorphism (SNP)

In addition to the study of gene expression/proteins/metabolites, identification of gene sequence variation (polymorphisms) is also available in a microarray format. Such methods are making possible systematic evaluations of the effects of common genetic variants or treatments [25]. Examples of genetic variants

Table 2. Selected cancer biomarkers of early biologic effect validated in epidemiological studies

Biomarker	Cancer	Study type	Subjects (Studies)	Exposure/Biological matrix	OR/RR/HR (CI or p-value)	Reference
Chromosomal aberrations	All cancer	PA	11 studies	Tertile category of chromosomal aberration frequency; lymphocyte	$RR_{\text{All CA, medium vs. low}}$ = 1.3 (1.07-1.60) $RR_{\text{All CA, high vs. low}}$ = 1.4 (1.16-1.72) [3.1 (1.17-8.39) for stomach cancer] $RR_{\text{Chromosome-type aberrations, medium vs. low}}$ = 1.3 (1.05-1.59) $RR_{\text{Chromosome-type aberrations, high vs. low}}$ = 1.4 (1.17-1.71) $RR_{\text{Ring chromosomes, yes vs. no}}$ = 2.2 (1.34-3.68)	Bonassi et al., [26]
Antibodies to Chlamydia (serotype G, I, D)	Cervical (squamous)	NCC	181 Ca / 533 Con within a cohort of 530,000	Serum	$OR_{\text{serotype G}}$ = 6.6 (1.6-27.0) $OR_{\text{serotype I}}$ = 3.8 (1.3-11.0) $OR_{\text{serotype D}}$ = 2.7 (1.3-5.6)	Anttila et al., [27]
IL-6	Head and neck (squamous)	PC	444 within a cohort of 869	Serum	$HR_{\text{recurrence (>12 pg/mL vs.} \leq 12)}$ = 1.32 (1.11-1.58) $HR_{\text{poor survival (>12 pg/mL vs.} \leq 12)}$ = 1.22 (1.02-1.46)	Duffy et al., [28]
IGF-1	Colorectal	NCC MA	1,121 Ca / 1,121 Con 10 studies	Quintile category of IGF-1 concentration; serum	$RR_{\text{highest vs. lowest quintile}}$ = 1.07 (1.01-1.14) $RR_{\text{highest vs. lowest quintile among milk consumer}}$ = 1.43 (1.13-1.93)	Rinaldi et al., [29]
Cell-free DNA	Liver (HCV related)	CC	87 Ca / 100 Con	Serum	HR_{survival} = 3.4 (1.5 -7.6) $HR_{\text{recurrence}}$ = 4.5 (1.3-14.9)	Tokuhisa et al., [30]
HRAS1 minisatellite locus	Ovarian	CC	136 Ca / 108 Con	DNA	OR = 1.70 (1.03-2.80); p_{trend} = 0.04	Weitzel et al., [31]

CC: case-control, C: cohort, RC: retrospective cohort, PC: prospective cohort, NCC: nested case-control, MA: meta-analysis, PA: pooled analysis, Ca: case, Con: control, RR: relative risk, OR: odds ratio, HR: hazard ratio, IL-6: interleukin 6, IGF-1: insulin-like growth factor-1, CRP: C-reactive protein, CI: confidence interval.

Table 3. Selected cancer biomarkers of susceptibility, DNA repair and metabolic polymorphisms validated in meta-/pooled analyses ; Polymorphisms in genes involved in carcinogen metabolism or DNA repair

Biomarker	Cancer	Study type	Studies (Subjects)	OR (CI) / p-value	Reference
CYP1A1	Esophageal	MA	9 studies (754 Ca / 1,563 Con)	$OR_{\text{Ile/Val vs. Ile/Ile}}$ = 1.4 (1.09-1.71) $OR_{\text{Val/Val vs. Ile/Ile}}$ = 2.5 (1.62-3.91)	Yang et al., [32]
GSTM1	Head and neck	MA	31 studies (4,635 Ca / 5,770 Con)	$OR_{\text{null vs. present}}$ = 1.2 (1.06-1.42)	Hashibe et al., [33]
	Nasopharyngeal	MA	8 studies (1,112 Ca / 1,601 Con)	$OR_{\text{null vs. present}}$ = 1.4 (1.21-1.66)	Zhuo et al., [34]
GST	Head and neck	PA	11 studies (2,334 Ca / 2,766 Con)	$OR_{\text{null vs. present}}$ = 1.3 (1.07-1.62) $OR_{\text{Combination of } GSTM1 \text{ null, } GSTT1 \text{ null, and } GSTP1 \text{ Val105 alleles}}$ = 2.1 (1.11-3.81), p_{trend} = 0.04	Hashibe et al., [33]
ERCC	Lung	MA	15 studies (5,004 Ca / 6,478 Con)	$OR_{\text{Gln/Gln vs. Lys/Lys}}$ = 1.3 (1.13 -1.49)	Kiyohara et al., [35]
XPA	Lung	MA	7 studies (1,913 Ca / 1,882 Con)	$OR_{\text{G/G vs. A/A}}$ = 0.8 (0.59-0.95)	Kiyohara et al., [35]
OGG1	Lung	MA	7 studies (3,253 Ca / 3,371 Con)	$OR_{\text{Cys/Cys vs. Ser/Ser}}$ = 1.2 (1.01, 1.53)	Hung et al., [36]
XRCC1	Breast	MA	40 studies (21,467 Ca / 22,766 Con)	$OR_{\text{All populations, Arg/Arg+Arg/Gln vs. Gln/Gln}}$ = 1.1 (1.02, 1.23) [1.6 (1.22-2.09) for Asian population]	Li et al., [37]
	Lung	MA	6 studies (1,702 Ca / 2,010 Con)	$OR_{\text{Asian population, Gln/Gln vs. Arg/Arg}}$ = 1.3 (1.16-1.54)	Kiyohara et al., [35]
	Tobacco-related cancers*	MA	16 studies (4,895 Ca / 5,977 Con)	$OR_{\text{Arg/Trp or Trp/Trp vs. Arg/Arg}}$ = 0.9 (0.77-0.95)	Hung et al., [36]
	Head and neck	PA	3 studies (430 Ca / 695 Con)	$OR_{\text{Whites, GlnGln vs. Arg/Arg}}$ = 0.6 (0.32-0.94)	Huang et al., [38]
MGMT	Head and neck	PA	3 studies (430 Ca / 695 Con)	$OR_{\text{Whites, PhePhe vs. Leu/Leu}}$ = 0.7 (0.51-0.98) $OR_{\text{Whites, Ile/Val+Val/Val vs. Ile/Ile}}$ = 0.7 (0.47-0.92)	Huang et al., [38]

CC: case-control, MA: meta-analysis, PA: pooled analysis, OR: odds ratio, *CYP1A1*: cytochrome P450, family 1, subfamily a, polypeptide 1, *GSTM1*: glutathione S-transferase mu 1, *GST*: glutathione S-transferase, *ERCC*: excision repair cross complementing, *XPA*: xeroderma pigmentosum, complementation group A, *XRCC1*: x-ray repair cross-complementing group 1, *MGMT*: O(6)-methylguanine DNA methyl-transferase, CI: confidence interval.

* included cancers of the lung, upper aerodigestive tract, bladder, stomach, liver, and pancreas, as well as myeloid leukemia.

that affect the sensitivity to drugs and cancer prognosis have long been known. High throughput genotyping of SNPs is now revealing a growing number of disease susceptibility genes [39,40], as well as polymorphisms that determine individual diversity in drug responses [41].

II. Copy Number Variation

The copy number variant (CNV) is the variable number of copies of a particular gene or segment of DNA and involves as much as 12% of human genome [42]. Studies of the human genome have revealed an unsuspected amount of variation between people in the number of copies of genes they have. Recent evidence shows that the gene copy number can be elevated in cancer cells. Because extra copies of DNA segment can cause disease, scientists should consider CNVs together with mutations or SNPs when seeking genetic causes for the response to chemicals.

III. Epigenetic Change of DNA

DNA methylation is a type of chemical modification of DNA that can be inherited and subsequently removed without changing the original DNA sequence. DNA methylation involves the addition of a methyl group to DNA with the specific effect of reducing gene expression; It typically occurs in a CpG dinucleotide context in the 5' regulatory regions of many genes. Alterations of DNA methylation have been recognized as an important component of diseases including cancer [43].

Adductomics is the study of DNA adducts in an entire genome. DNA adducts are compounds bound to DNA, causing damage and mutations. Cellular DNA is always exposed to various DNA damaging agents such as UV light, ionizing radiation and environmental carcinogens. Endogenous reactive oxygen species, reactive nitrogen species and lipid peroxide also cause DNA damage. The science of adductomics seeks to identify all DNA adducts and the target sequence of each adduct.

IV. Transcriptome

Given a number of diseases generally involves changes in a series of gene expression and their interactions, a global analysis of gene expression (i.e., microarray) may provide a more comprehensive view of human diseases [44]. Gene expression is altered either directly or indirectly as a result of environmental exposure (e.g., ionizing radiation) [45]. The promising fields of microarray technology include identification of specific biomarkers for the monitoring of disease progression and prognosis as well as gene-environment interaction [46,47]. As a new genomic biomarker, the genome-wide profiling of the microRNAs may be a more powerful tool to find early detection markers for cancer [48].

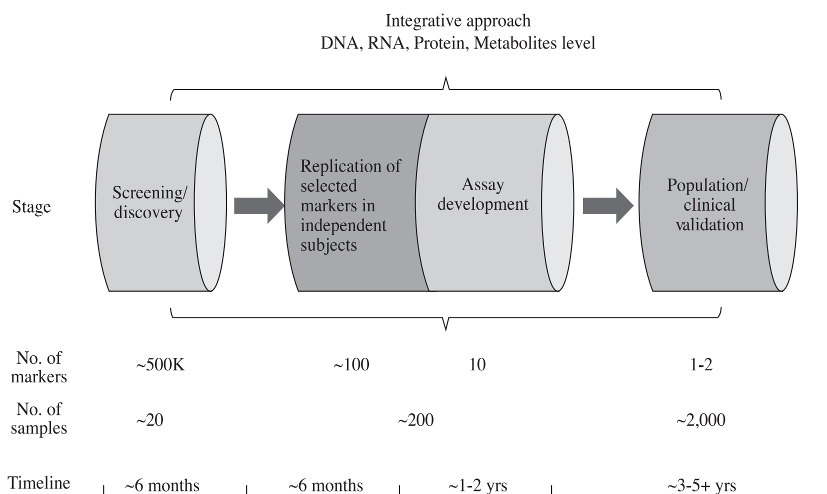

Figure 2. Procedures of genomic approaches to biomarker identification and application [adapted from : Ayers institute biomarker pipeline (http://www.vicc.org/jimayersinstitute/devt/)].

V. Proteome

Proteomics is a global analysis of the full set of proteins (the proteome) of a tissue or cell type encoded by a genome and deals with separation, quantitation and functional characterization of expressed proteins in biological samples [49]. The most commonly used techniques for studying the proteome are two-dimensional (2D) gel electrophoresis and mass spectrometry. Recently, new proteomic technologies such as multi-dimensional protein identification technology [50] and specific protein and antibodies arrays [51] have allowed the genome-wide functional and structural characterization of proteins.

Profiling of global protein expression patterns provides important complementary information to genomics, as RNA is subject to post-transcriptional control to regulate the amount and quality of proteins i.e., phosphorylation, ubiquitination and other process). Thus, proteomic technology may offer better knowledge of the cellular process along with complementary information to genomics.

VI. Metabolome

Metabolome refers to the complete set of small-molecule metabolites including metabolic intermediates and secondary metabolites that are found in biological samples [52]. Metabolomics combines the application of analytical technologies (e.g., mass spectro-metry and nuclear magnetic resonance spectrometry) with statistics and bioinformatics to allow the quantitative measurement of the levels of cellular metabolites [53].

While it is quite difficult to extract cellular mRNA and protein from limited amounts of tissues, metabolites can be easily purified from tissue as well as fluids such as serum, plasma, cerebrospinal fluids and urine.

IDENTIFICATION AND VALIDATION OF GENOMIC BIOMARKERS

Despite the current optimism related to genomic approaches, a number of important limitations to the discovery of novel single biomarkers have been suggested, including study design bias, artifacts related to the collection and storage of samples, and false-positive results [17].

The development of various emerging biomarkers might be considered as a process similar to therapeutic drug development; Figure 2 shows the overall procedures used in the identification and application of genomic biomarkers. In the discovery (screening) stage, up to hundreds of thousands of genomic markers should be identified from the analyses of a small number of very well-characterized

samples such as cancers, pre-cancers or normal tissue specimens. Prior to further evaluation, other information such as gene expression patterns may help to prioritize candidates for further evaluation. Hundreds of candidates are subjected to an iterative process of verification and assay development in which highly sensitive, targeted analyses replicate the results. The final stage of the procedure is an evaluation of biomarker candidates in populations or clinical trials. A few highly promising biomarker candidates may remain for evaluation by this stage.

As a means of increasing the efficiency, 'Integrative Functional Informatics' represents a novel direction; The interfacing and integration of different technologies for data collection and analysis are expected to be pivotal to biomarker identification, characterization, validation and application [54,55].

CONCLUSION

In this review, we briefly summarized issues in evaluating biomarkers and selected cancer biomarkers that were validated in epidemiological studies. We also discussed genomic biomarkers as a new approach to identify and apply biomarkers.

We note that the utility and importance of biomarkers has been well recognized and that biomarker discovery efforts are now common in both academic and industrial settings. Research on the discovery and development of biomarkers is expected to shift toward a more organized and integrative approach with technological advances in molecular and genomic epidemiology.

REFERENCES

1. Schulte PA. A conceptual and historical framework for molecular epidemiology. In: Schulte PA, Perera FP, editors. *Molecular Epidemiology: Principles and Practices*. San Diego: Academic Press; 1993. p. 1-36.
2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* 2001; 409(6822): 860-921.
3. Khoury MJ, Dorman JS. The human genome epidemiology network. *Am J Epidemiol* 1998; 148(1): 1-3.
4. Strickland PT, Qian Z, Friesen MD, Rothman N, Sinha R. Metabolites of 2-amino-1-methyl-6-phenylimidazo(4,5-b)pyridine (PhIP) in human urine after consumption of charbroiled or fried beef. *Mutat Res* 2002; 506-507: 163-173.
5. Lee KM, Kang D, Lee SJ, Park SK, Lee KH, Choi JY, et al. Interactive effect of genetic polymorphism of glutathione S-transferase M1 and smoking on squamous cell lung cancer risk in Korea. *Oncol Rep* 2006; 16(5): 1035-1039.
6. Lee KM, Choi JY, Kang C, Kang CP, Park SK, Cho H, et al. Genetic polymorphisms of selected DNA repair genes, estrogen and progesterone receptor status, and breast cancer risk. *Clin Cancer Res* 2005; 11(12): 4620-4626.
7. Kang D, Lee KH, Lee KM, Kwon HJ, Hong YC, Cho SH, et al. Design issues in cross-sectional biomarkers studies: Urinary biomarkers of PAH exposure and oxidative stress. *Mutat Res* 2005; 592(1-2): 138-146.
8. Van Maele-Fabry G, Duhayon S, Mertens C, Lison D. Risk of leukaemia among pesticide manufacturing workers: A review and meta-analysis of cohort studies. *Environ Res* 2008;106(1):121-137.
9. Van Maele-Fabry G, Libotte V, Willems J, Lison D. Review and meta-analysis of risk estimates for prostate cancer in pesticide manufacturing workers. *Cancer Causes Control* 2006; 17(4): 353-373.
10. Boffetta P, Clark S, Shen M, Gislefoss R, Peto R, Andersen A. Serum cotinine level as predictor of lung cancer risk. *Cancer Epidemiol Biomarkers Prev* 2006; 15(6): 1184-1188.
11. McDonald SA, Hutchinson SJ, Bird SM, Robertson C, Mills PR, Dillon JF, et al. A record-linkage study of the development of hepatocellular carcinoma in persons with hepatitis C infection in Scotland. *Br J Cancer* 2008; 99(5): 805-810.
12. Giordano TP, Henderson L, Landgren O, Chiao EY, Kramer JR, El-Serag H, et al. Risk of non-Hodgkin lymphoma and lympho-proliferative precursor diseases in US veterans with hepatitis C virus. *JAMA* 2007; 297(18): 2010-2017.
13. Mork J, Lie AK, Glattre E, Hallmans G, Jellum E, Koskela P, et al. Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck. *N Engl J Med* 2001; 344(15): 1125-1131.
14. Wu HC, Wang Q, Yang HI, Ahsan H, Tsai WY, Wang LY, et al. Aflatoxin B1 exposure, hepatitis B virus infection, and hepatocellular carcinoma in Taiwan. *Cancer Epidemiol Biomarkers Prev* 2009; 18(3): 846-853.
15. Gunter MJ, Divi RL, Kulldorff M, Vermeulen R, Haverkos KJ, Kuo MM, et al. Leukocyte polycyclic aromatic hydrocarbon-DNA adduct formation and colorectal adenoma. *Carcinogenesis* 2007; 28(7): 1426-1429.
16. Rundle AG, Vineis P, Ahsan H. Design Options for molecular epidemiology research within cohort studies. *Cancer Epidemiol Biomarkers Prev* 2005; 14(8): 1899-1907.
17. Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat Clin Pract Oncol* 2008; 5(10): 588-599.
18. Chatterjee SK, Zetter BR. Cancer biomarkers: Knowing the present and predicting the future. *Future Oncol* 2005; 1(1): 37-50.
19. Lee MS, Jun DH, Hwang CI, Park SS, Kang JJ, Park HS, et al. Selection of neural differentiation-specific genes by comparing profiles of random differentiation. *Stem Cells* 2006; 24(8): 1946-1955.
20. Park WY, Hwang CI, Kang MJ, Seo JY, Chung JH, Kim YS, et al. Gene profile of replicative senescence is different from progeria or elderly donor. *Biochem Biophys Res Commun* 2001; 282(4): 934-939.
21. Kim JH, Ha IS, Hwang CI, Lee YJ, Kim J, Yang SH, et al. Gene expression profiling of anti-GBM glomerulonephritis model: The role of NF-kappaB in immune complex kidney disease. *Kidney Int* 2004; 66(5): 1826-1837.
22. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 1999; 286(5439): 531-537.
23. Yates JR 3rd. Mass spectrometry: From genomics to proteomics. *Trends Genet* 2000; 16(1): 5-8.
24. Holmes E, Nicholson JK, Tranter G. Metabonomic characterization of genetic variations in toxicological and metabolic responses using probabilistic neural networks. *Chem Res Toxicol* 2001; 14(2): 182-191.
25. Yoo KY, Kang D. Current researches on breast cancer epidemiology in Korea. *Breast Cancer* 2003; 10(4): 289-293.
26. Bonassi S, Norppa H, Ceppi M, Stromberg U, Vermeulen R, Znaor A, et al. Chromosomal aberration frequency in lymphocytes predicts the risk of cancer: results from a pooled cohort study of 22,358 subjects in 11 countries. *Carcinogenesis* 2008; 29(6): 1178-1183.

27. Anttila T, Saikku P, Koskela P, Bloigu A, Dillner J, Ikaheimo I, et al. Serotypes of Chlamydia trachomatis and risk for development of cervical squamous cell carcinoma. *JAMA* 2001; 285(1): 47-51.
28. Duffy SA, Taylor JM, Terrell JE, Islam M, Li Y, Fowler KE, et al. Interleukin-6 predicts recurrence and survival among head and neck cancer patients. *Cancer* 2008; 113(4): 750-757.
29. Rinaldi S, Cleveland R, Norat T, Biessy C, Rohrmann S, Linseisen J, et al. Serum levels of IGF-I, IGFBP-3 and colorectal cancer risk: results from the EPIC cohort, plus a meta-analysis of prospective studies. *Int J Cancer* 2009. Epub 2009 Oct 6.
30. Tokuhisa Y, Iizuka N, Sakaida I, Moribe T, Fujita N, Miura T, et al. Circulating cell-free DNA as a predictive marker for distant metastasis of hepatitis C virus-related hepatocellular carcinoma. *Br J Cancer* 2007; 97(10): 1399-1403.
31. Weitzel JN, Ding S, Larson GP, Nelson RA, Goodman A, Grendys EC, et al. The HRAS1 minisatellite locus and risk of ovarian cancer. *Cancer Res* 2000; 60(2): 259-261.
32. Yang CX, Matsuo K, Wang ZM, Tajima K. Phase I/II enzyme gene polymorphisms and esophageal cancer risk: a meta-analysis of the literature. *World J Gastroenterol* 2005; 11(17): 2531-2538.
33. Hashibe M, Brennan P, Strange RC, Bhisey R, Cascorbi I, Lazarus P, et al. Meta- and pooled analyses of GSTM1, GSTT1, GSTP1, and CYP1A1 genotypes and risk of head and neck cancer. *Cancer Epidemiol Biomarkers Prev* 2003;12(12):1509-1517.
34. Zhuo X, Cai L, Xiang Z, Li Q, Zhang X. GSTM1 and GSTT1 polymorphisms and nasopharyngeal cancer risk: An evidence-based meta-analysis. *J Exp Clin Cancer Res* 2009; 28: 46.
35. Kiyohara C, Yoshimasu K. Genetic polymorphisms in the nucleotide excision repair pathway and lung cancer risk: A meta-analysis. *Int J Med Sci* 2007; 4(2): 59-71.
36. Hung RJ, Hall J, Brennan P, Boffetta P. Genetic polymorphisms in the base excision repair pathway and cancer risk: A HuGE review. *Am J Epidemiol* 2005; 162(10): 925-942.
37. Li H, Ha TC, Tai BC. XRCC1 gene polymorphisms and breast cancer risk in different populations: A meta-analysis. *Breast* 2009; 18(3): 183-191.
38. Huang WY, Olshan AF, Schwartz SM, Berndt SI, Chen C, Llaca V, et al. Selected genetic polymorphisms in MGMT, XRCC1, XPD, and XRCC3 and risk of head and neck cancer: A pooled analysis. *Cancer Epidemiol Biomarkers Prev* 2005; 14(7): 1747-1753.
39. Lee DH, Kim EG, Park KS, Jeong SW, Kong ID, Lee JW. Characteristics of P2X7-like receptor activated by adenosine triphosphate in HIT-T15 cells. *Pancreas* 2007; 35(1): 53-62.
40. Lee SA, Lee KM, Park WY, Kim B, Nam J, Yoo KY, et al. Obesity and genetic polymorphism of ERCC2 and ERCC4 as modifiers of risk of breast cancer. *Exp Mol Med* 2005; 37(2): 86-90.
41. Uetrecht J. Idiosyncratic drug reactions: current understanding. *Annu Rev Pharmacol Toxicol* 2007; 47: 513-539.
42. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature* 2006; 444(7118): 444-454.
43. Wilson AS, Power BE, Molloy PL. DNA hypomethylation and human diseases. *Biochim Biophys Acta* 2007; 1775(1): 138-162.
44. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nat Genet* 1999; 21(1 Suppl): 33-37.
45. Park WY, Hwang CI, Im CN, Kang MJ, Woo JH, Kim JH, et al. Identification of radiation-specific responses from gene expression profile. *Oncogene* 2002; 21(55): 8521-8528.
46. Van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415(6871): 530-536.
47. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature* 2000; 406(6797): 747-752.
48. Nelson KM, Weiss GJ. MicroRNAs and cancer: past, present, and potential future. *Mol Cancer Ther* 2008; 7(12): 3655-3660.
49. Anderson NL, Taylor J, Hofmann JP, Esquer-Blasco R, Swift S, Anderson NG. Simultaneous measurement of hundreds of liver proteins: Application in assessment of liver function. *Toxicol Pathol* 1996; 24(1): 72-76.
50. Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001; 19(3): 242-247.
51. Cutler P. Protein arrays: The current state-of-the-art. *Proteomics* 2003; 3(1): 3-18.
52. Collings FB, Vaidya VS. Novel technologies for the discovery and quantitation of biomarkers of toxicity. *Toxicology* 2008; 245(3): 167-174.
53. Robertson DG, Reily MD, Baker JD. Metabonomics in preclinical drug development. *Expert Opin Drug Metab Toxicol* 2005; 1(3): 363-376.
54. Woo HG, Park ES, Lee JS, Lee YH, Ishikawa T, Kim YJ, et al. Identification of potential driver genes in human liver carcinoma by genomewide screening. *Cancer Res* 2009; 69(9): 4059-4066.
55. Bensalah K, Montorsi F, Shariat SF. Challenges of cancer biomarker profiling. *Eur Urol* 2007; 52(6): 1601-1609.