Circulating nucleic acids in blood as biomarkers

Reidun Øvstebø, Peter Kierulf and Kari Bente Foss Haug

R&D Section, Department of Clinical Chemistry, Ullevål University Hospital, Oslo

Correspondence: Reidun Øvstebø, The R& D Section, Department of Clinical Chemistry, Ullevål University Hospital, 0407 Oslo, Norway
Telephone: 0047 22119490   Telefax 0047 22118189   E-mail: reidun.ovstebo@medisin.uio.no

ABSTRACT

This short review on a rapidly expanding domain in biomarkers focuses on the value of markers derived from either circulating intracellular DNA and RNA (leukocytes) or from free DNA and RNA in plasma or serum. In circulating intracellular DNA biomarkers, importance has been pointed to reside in the ever increasing number of SNPs directly related to disease such as hemochromatosis or associated with genetic make up that leads to different drug-susceptibility. Quantitative gene expression profiling, increasingly using global expression platforms, is gaining momentum in various disease states such as cancer, inflammation, cardiovascular disease and diabetes. Circulating free nucleic acids in plasma or serum gain in importance as biomarkers particularly in cancer and foeto-maternal understanding. The surprising recent findings of circulating free mRNA carries the potential of examining normal and diseased plasma for global gene expression profiling – opening avenues to new biomarkers. When appropriate, this review gives reference to methodological considerations and refers the readers to important literature in the fields.

SAMMENDRAG

I denne korte oversiktsartikkelen redegjøres det for et biomarkørfelt som utvikler seg hurtig. Gjennom en blodprøve kan man få kjennskap til forandringer i sirkulerende leucocytter, intracellulære nukleinsyrer (DNA og RNA) og fritt DNA og RNA fra plasma eller serum. Single Nucleotide Polymorphisms (SNPs) i DNA har allerede bekreftet sine muligheter som biomarkører (f.eks. Hemokromatose, Faktor V Leiden, Cytochrom P450 (CYP'er)). Stadig flere SNP'er vinner innpass i klinisk sammenheng. Siden sirkulerende hvite blodlegemer kan sies kontinuerlig å overvåke kroppens organer og vev, og dette avspeiles i disse blodcellers genekspresjon (RNA), knyttes det i dag forventninger til sykdomsspesifikke genekspresjons-profiler. Både ved visse kreftformer, betennelsestilstander og hjertekar-sykdom viser hvite blodlegemer mer eller mindre tydelig sykdomsspesifikke genekspresjonsprofiler. Denne type sykdomsspesifikke genekspresjonsmarkører vil bli økende viktig fremover. Ved slike markører vil man kunne ha nytte av kvantitativ måling av enkeltmarkører, og også globale genekspresjonsprofiler på mikroarray-plattformer. Sirkulerende fritt DNA og kanskje særlig RNA i plasma åpner for nye sykdomsmarkører i første rekke ved forskjellige kreftformer og ved foeto-maternelle problemstillinger. Oversikten gir også en henvisning til metodologiske referanser i disse feltene.

INTRODUCTION

In general, biomarkers aim at disclosing or monitoring disease. So also with nucleic acids, which comprise DNA as well as RNA. Their main working place is within the cell, whether in fixed tissues (liver, heart etc.) or in circulating blood (white blood cells, reticulocytes). In the present review we shall focus on "circulating nucleic acids", with special emphasis on blood. We shall approach this in a dual way by focusing on recent knowledge that may be extracted from circulating white blood cells as regards both to DNA (mutations, hypermethylation, for review see (1)) and RNA (gene expression). In this latter context we shall consider circulating white blood cells as "ambulatory surveyors of disease". Our second approach will be related to circulatory extracellular DNA and RNA, i.e. obtainable from whole blood, either cell bound or free

in plasma. In both domains, cellular as well as extra-cellular, we shall emphasis pre-analytical as well as analytical challenges when appropriate. At the very end of this review we shall speculate as to how these important fields may develop.

In many ways circulating blood is unique in the human body. It "travels" several miles every day and night, and directly or indirectly perfuses every cell of the body, thus having a unique position to "understand what is going on in our bodies". Circulating blood is a notably dynamic environment involving the turnover of approximately 1 trillion blood cells daily, including 200 billion red blood cells and 70 billion neutrophilic leukocytes (2). Nucleated leukocytes, which include lymphocytes, monocytes and granulocytes, are the most transcriptionally active cells in the blood and thereby targets for both DNA analysis and gene expression profiling.
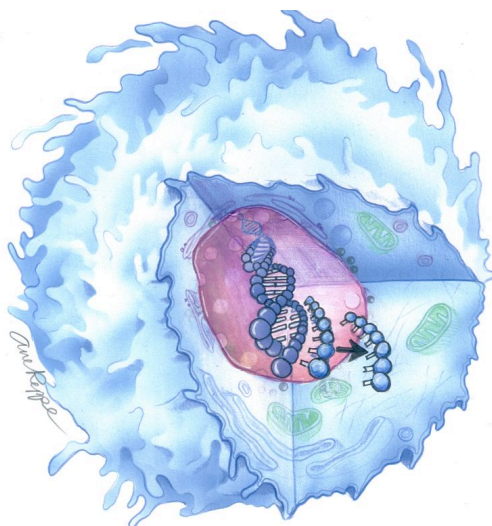
Nucleated, transcriptionally acive cell. Illustrasjon: Ane Reppe.

The apparent ease of sampling blood makes this circulatory tissue attractive in trying to understand disease. For years clinical pathologists and clinical chemists have drawn whole blood for cell count, hemoglobin measurements and processed whole blood into plasma or serum to quantify proteins, hormones and metabolites. Procedural care has been established in order to obtain reliable results. Examining for circulating nucleic acids may pose specific, new problems as may biobanking of these specimens. When appropriate we shall address these issues.

CELLULAR DNA

DNA is the biochemical substance that specifies all the different parts in living organisms and defines an individual. All inherited information is specified by a simple four-letter alphabet, the nucleotides A, G, C and T, and remarkably, it is the order of these billions of four nucleotides that constitutes the genome and encodes all of the instructions necessary to create a complete organism.

The most well known features of the genome are the genes, which encode protein products. The Human Genome Project manifested roughly 33,000 different genes, comprising only a tiny fraction (~1-3%) of the entire human genome (3). Yet it is thought that most of the functional relevance of an organism is encoded within the genes and the various regulatory regions surrounding the genes, although information on non-coding regions are becoming increasingly interesting.

DNA variations

Almost every one of the 100 trillions (10^{14}) of cells in the body has its own copy of the genome, so also with the nucleated white blood cells in circulating blood. Each cell has two such copies, one maternal and one paternal, thus having two versions of every gene and

two recipes for the same protein. Even subtle differences in this recipe may lead to alternative proteins – generating the variance responsible for making all human beings biologically unique.

Surprisingly, 99% of the genome is stipulated to be identical among different human people, but nevertheless, the presence of only 1% variance accounts for the huge phenotypic differences among individuals. Variations in DNA have arisen during evolution as a consequence of mutations. As dividing cells are produced, each new cell needs to replicate and make its own copy of the genome, a process that is not error free. Errors in DNA replication or other damages, although continually proofread, bring the misincorporations into existence and a mutation – or a variation – is born. Novel gene variants may not work as well as the original version and usually die out within a few generations. On the contrary, new variants may be better than or functionally equivalent to the original version. These good or neutral variants may increase in frequency in the population due to chance, or because they confer better odds for survival and successful reproduction for the ones who have them. DNA variations may have functional consequences by altering a protein or its rate of expression or may have no functional consequences at all and constitute merely as a positional marker in the genome.

DNA variations are often referred to as polymorphisms. The term polymorphism is used to indicate that a particular DNA position has more than one form (up to three) in the population, although each individual will only have one or a maximum of two forms. Classically, a distinction has been made between variations and polymorphisms. A locus is only called polymorphic if the most common variant occurs in less than 95% of the population. Consequently all other variants will then occur with a total frequency of 5% or more (4). It is also common to distinguish between a polymorphism and a mutation. A mutation refers to a rare variant that is the primary cause of a clinical phenotype or a disease, whereas polymorphism is used to denote a variant that is present in the population in a relatively high frequency. Polymorphisms are by themselves more seldom sufficient to cause a disease, but may contribute to susceptibility to a disease or to variation in functional properties of a protein.

Single Nucleotide Polymorphisms, termed SNPs, are the most abundant and the simplest form of DNA variations in which a single nucleotide is replaced by another. SNPs are classified into coding and non-coding polymorphisms according to their position in or around genes. Coding SNPs are further classified into synonymous and nonsynonymous variations. Synonymous codons change into another codon coding for the same amino acid leading to no change in the structure of the protein (silent change). Nonsynonymous SNPs change the codon to one specifying for a different amino acid and may therefore change the protein structure. It has been estimated that the human genome contains

one sequence variant in every 200 to 1000 base pairs and that each gene on average has 126 SNPs (5).

Other DNA variants include short and long insertions/deletions ("indels"), whereas one, a few or many nucleotides may be added or removed in the genome. These variants may lead to amino acid addition/subtraction in the protein or to total disruption of the protein message depending on the sequence involved. Incorporation of transposable elements may as well lead to radical changes at the protein level.

Microsatellite repeats is a more complex type of DNA variation consisting of short di-, tri-, tetra- or penta-nucleotides repeated several to hundreds of times along the DNA. Microsatellite repeats occur in all individuals, but the length of the repeat and the position of the sequence in the genome is of vital importance for the extent of damage.

DNA methylation status is another well studied gene regulation mechanism, especially in carcinogenesis. Hypermethylation of CpG-rich islands in the promoter regions of tumor supressor genes leads to gene silencing and is one of the most common molecular transformations in a cancer cell. Hypomethylation has also been recognized as a cause of cancer (6).

In most human cells telomerases shorten during aging, suggesting that telomere length could be a biomarker of aging and age-related morbidity (7).

DNA variants as biomarkers

The introduction of DNA variations in genetic analysis has already contributed in medical diagnostics, because many of these molecular markers can function as biomarkers when associated with development or predisposition to common diseases as well as individual variations in drug response. One criterion to prioritize a polymorphism for study, is whether it has a functional consequence.

Among variations being analyzed, SNPs seem to be the most common type of genetic variation and about 10 millions of SNPs (5) are now documented in several dedicated databases free of charge (http://www.ncbi.nlm.nih.gov/entrez/query.fgci?db=snp, SNP Browser™ v3.1, Applied Biosystems). The scientific focus is now drifting towards how these SNPs co-occur, how they interact to alter the function of a gene, and which SNPs can eventually be used as biomarkers. However, important to notice is that many reported DNA variations have only been identified using a few individuals and have not been strictly defined as polymorphisms. Investigators may also take into consideration that the allele frequency may vary dramatically among ethnic groups.

A group of defined SNPs has already for several years been utilized as biomarkers for diagnosis of simple, single gene defects. Detection of SNPs in for example the Coagulation Factor V (Leiden) and the Prothrombin genes have been established as routine analysis in coagulation defects (8,9), as well as documented SNPs in the candidate genes for Hypercholes-terolemi, Hemochromatosis and Lactosis intolerance (10,11).

The term haplotype may be useful when several SNPs are identified surrounding the gene of interest. Haplotypes refers to a combination of polymorphisms in close proximity that are coinherited in many individuals in a population. Designing haplotypes from a distinct region may provide a more complete description of the different SNPs involved in the disease studied, and databases for searching and generating haplotypes are established (http://www.nhgri.nih.gov/About_NHGRI/Der/haplotype/index.html). Since the number of SNPs and haplotypes reported have increased extensively the last years, the approach to seek for good SNP biomarkers has been more challenging. As complexity of the project precedes, professional assistance from dedicated technical staff (bioinformatics) is strongly recommended.

Determining the genetic bases of the common "multifactorial" diseases, however, represents a major challenge. The genetics of these diseases are complicated by the interplay by many genes in combination with the environment. Because SNPs are densely distributed across the whole genome, they are ideal markers for large scale genome-wide association studies to discover common complex diseases, such as cancer, hypertension, diabetes, obesity, and psychiatric disorders.

Since cancer is a DNA disease characterized by uncontrolled cell proliferation due to accumulation of genetic alterations, genetic instability has also been recognized as a central biomarker in many forms of cancer. Most colorectal cancers are identified through chromosomal instability, either allelic losses of chromosomes or instability of microsatellites. These alterations may contribute to inactivation of tumor suppressor genes and accumulation of mutations in important genes regulating the cell cycle and apoptosis, respectively (12). Aberrant methylation can be used as a marker to detect cancer cells (13). A number of studies have provided evidence that specific methylation changes can alter the response to different therapeutic agents in cancer, and therefore be useful as biomarkers (14).

The focus of SNP analysis is now changing from the identification of new SNPs to their typing in populations. For SNPs to be potential biomarkers, the polymorphisms have to be mapped accurately, their frequency in various populations determined, and automated high-throughput assay techniques developed. One problem researchers face when designing human genetic studies with SNPs, is the difficult task of selecting the most suitable set of DNA variants for the goal at hand and in a cost-effective manner.

Pharmacogenomics

It is commonly accepted that no drug works well for all patients. Some of the differences in how patients respond to a drug are due to personal characteristics

such as age, size and gender, the nature of their disease and what other drugs are being used. Despite all these factors, it is claimed that half of all variation in drug response is attributable to the genetic differences among patients.

Some of the specific genetic factors involved in drug response have been known for a long time and belong to a class of proteins called "drug metabolizing enzymes". The Cytochrom P-450 system (CYP) is the most important and characterized enzyme system involved in drug metabolism (15), but other classes of enzymes and several drug receptors are contributing to pharmacokinetic variations as well (16,17). Variations in the genes that encode these CYP-molecules (e.g. CYP1A2, CYP2C9, CYP2C19, CYP2D6 and CYP3A4) influence how quickly these enzymes process and eliminate the drug and their metabolites from the blood. Drugs metabolized too quickly, may not reach a high enough concentration to cure the targeted disease or relieve the symptoms. Drugs metabolized too slowly, may accumulate and reach a toxic level in the body. The availability of the human DNA sequence, its variation between individuals and the functional understanding of genetic determinants between individuals may enable a safer dosage towards a more effective and personalized drugs.

The ease of using DNA from circulating white blood cells for hunting biomarkers is obvious. DNA is the ultimate stable molecule upon storage and no preanalytical precautions need to be fulfilled. A few microliters of EDTA whole blood is sufficient for up to a hundred genotypings. The challenge for tomorrow will be to exploit and quality control all the reported information from the databases.

CELLULAR RNA

General considerations

Understanding the function of genes and other parts of the genome is known as functional genomics. The Human Genome Project was just the first step in understanding humans at the molecular level. Though the project is complete, many questions still remain unanswered, including the function of most of the estimated 33,000 human genes. Most genes and various regulatory regions surrounding the genes contain information for making specific proteins. mRNAs, the intermediate molecules between DNA and proteins, are the target molecules for gene expression analysis, the process by which proteins are made from the instructions encoded in the DNA molecule.

Cellular RNA in disease

Recent reports have revealed that peripheral leukocytes that communicate with every tissue and organ in the body, have the potential to be used diagnostically as surrogates for direct sampling of sites of different disease processes. Detection of disease-specific prognostic markers from blood cells in leukemia patients has proven its usefulness in diagnosis. Quantification of the kinetics in minimal residual disease, such as chronic myeloid leukemia or acute lymphoblastic leukemia, by RT-PCR has resulted in new clinical staging and response to treatment (18,19). Application of microarray methodology on to carcinoma tissues (breast, colon or prostate) has resulted in the discovery of response changes in a variety of genes. Some of these transcripts may potentially in the future serve as biomarkers to detect occult tumor cells in blood originating from solid cancers (18,20,21). Ma et al. suggest that peripheral blood cells undergo profound molecular changes during atherogenesis showing 108 genes differently expressed in coronary artery disease peripheral blood cells as compared to normal (22). Microarray analysis performed on total RNA in blood cells from patients with schizophrenia or bipolar disorders have shown that each disease state exhibited a uniquely expressed genome signature as compared to normal (23). Whole blood is supposed to become the most important clinical specimen also to obtain surrogate markers for diseases that are not primarily associated with peripheral blood (24-26).

Thus, blood derived RNA appears to represent a novel alternative to tissue biopsies as a source for mRNA gene expression profiling. However, given the heterogenous blood cell population and the challenge of isolation of high-quality total-RNA, there are many factors to be accounted for during sampling, analyzing and evaluation of the results.

Preanalytical precautions

Intracellular RNA will be rapidly degraded ex vivo by specific and nonspecific endonucleases if not stabilized immediately upon sampling. This will lead to changes in the gene expression profiles. Standardization of this preanalytical step may now more easily be done by using integrated and standardized systems for collection and stabilization of whole blood specimens such as the PAXgene (Qiagen) or the Tempus (Applied Biosystems) systems. These systems have been validated to maintain RNA in a satisfactory way both at room temperature for 5 days and when kept frozen. One technical obstacle using these systems however, is that RNA from reticulocytes will be isolated at the same time. Reticulocytes are transcriptionally inactive, but contain huge amounts of globin mRNAs. The number of reticulocytes are normally 10 fold that of leukocytes and they will therefore contribute up to 70% of the total RNA from blood (26). These circumstances have been shown to decrease the sensitivity of detection of other transcripts in some of the detection methods, but methods for reduction of globin mRNAs not affecting further analysis, are now available (26).

In an epidemiological setting robotized isolation of total RNA is a prerequisite. Conventional, manual iso-

lations methods are demanding and will be cumbersome with a high number of samples. Lability of RNA molecules and the numerous possibilities to pollute samples during isolation procedures, forces the isolated total-RNA to be quality checked for the deleterious effect of RNases before subjecting the RNA to further analysis. This may now be tested for by running the samples in a dedicated electrophoretic system (27).

Quantification of specific mRNA transcripts

Most methods for quantification of specific mRNA transcripts require enzymatic steps in advance of detection, such as reverse transcription, to synthesize cDNA from mRNA or labeling of cRNA. Real time PCR or microarrays are most often used as analytical platforms for the quantification. Gene expression profiles from blood cells can be obtained either by analyzing the global gene expression by microarray technology or by quantification sets of specific transcripts by quantitative RT-PCR.

Global gene expression profiles may give information on a large number of transcripts. Depending on array type, filtering and study design, genes overexpressed or suppressed may be registered active at a given time. The level of expression of certain genes may signify a particular disease state. Genes thus consistently overexpressed or suppressed in a certain clinical context may be considered as biomarkers. Many diseases are polygenic and triggered by genetic, environmental and physiological factors. Software clusters of relevant genes may be performed to distinguish among different disease patterns.

Quantitative RT-PCR is normally used for detection of known specific transcripts. Different analytical platforms offer facilities for quantification of one to 384 transcripts at a time. This makes quantification of gene expression profiles from clusters of genes possible within a short time and the future will probably provide "cardio-card", "diabetic-card", "hypertension-card". The combination of microarray analysis, that identify panels of genes relevant for a disease state, and real-time PCR based assays for quantification of the panels in clinical settings, may be valuable tools in the future.

As for today, qRT-PCR assays are most established for detection of viral load and therapy monitoring (HIV, SARS) (28-31) and for diagnosis and detection of disease-specific prognostics markers in patients with leukaemia (32,33).

CIRCULATING EXTRACELLULAR DNA AND RNA

General considerations

The discovery of cell-free nucleic acids in plasma was first reported by Mandel and Metais in 1948 (34), but was initially not widely recognized. By the term circulating extracellular nucleic acids, we imply DNA and RNA that exist in blood plasma or serum as free nucleic acid. Thus, extracellular nucleic acids have escaped their natural environment and have gained access to plasma where they may exist in solution or may be particle-bound. These forms of circulating extracellular nucleic acids therefore should be considered apart from circulating cellular nucleic acids as regards biomarker functionality.

Methodological considerations

Since circulating extracellular nucleic acids may exist in solution in plasma or serum in various particular forms or bound to blood cells, special analytical considerations have to be taken into considerations in order to obtain comparable results. As of today most studies have been performed on plasma processed from EDTA whole blood or serum from clotted blood (35). We have found no studies that have systematically compared the effects of various anticoagulants. A few comparative studies using both plasma and serum have been done (36), the results usually giving higher and more labile serum than plasma values (37), most likely due to the release of cellular constituents upon inclusion of blood cells into the clot. For quantitative estimates we would recommend the use of plasma obtained by centrifugation (1600xg, 10 min, 4°C, plasma pipetted off and recentrifuged 16000xg, 10 min, 4°C (35,36)) of EDTA (5 mmol/L) whole blood. EDTA blood, prior to processing may be stored at +4°C for up to 24 hours prior to processing. Only a few studies on extracellular circulating nucleic, cell-bound (38) or filterable (39), have been reported. Since circulating nucleic acid investigations frequently implies its use in epidemiological studies, where sample number often may be quite large, automated extraction from plasma has been advocated (35).

Circulating extracellular DNA and RNA in disease

General comments
At present there are two major fields where research – on the brink of routine in some places – push on. The first is focused on understanding fetal genetic make up and wellbeing by examining for fetal DNA and RNA in maternal plasma. The other is directed at early or supplementary cancer diagnostics. Most studies in these fields have extracted nucleic acids from plasma (35,37).

Fetal DNA and RNA in maternal plasma
Conventional methods of obtaining fetal tissues for genetic analysis, including amniocentesis and chorionic villus sampling, are invasive and constitute a finite risk to the fetus. This picture changed upon the discovery by Lo et al. in 1997 (40) that fetal DNA (on the Y-chromosome) was present in maternal plasma and serum.

Fetal DNA in maternal plasma has been shown to be useful for the prenatal diagnosis of certain neurological diseases, fetal chromosomal aneuploidies, sex-

linked disorders and fetal rhesus D (RhD) status (35). Fetal RhD genotyping from maternal plasma has become an adopted protocol in routine prenatal diagnosis in several centers (http://www.bloodnet.nbs.nhs.uk/ibgrl/Reference%20Services/RefSer_genotyping.htm). Quantitative aberrations of fetal DNA in maternal plasma have been reported in various disease conditions, i.e. pre-eclampsia, fetal-maternal haemorrhage and polyhydramnion (35), primarily based on the detection of Y-chromosomal sequences in maternal plasma, thus limiting their applications to the 50% of pregnancies involving male fetuses.

Using RT-PCR, fetal RNA of placental origin has recently been detected in maternal plasma (37). An important extension of this placental avenue has lately been shown by Tsui et al. (41), who used systematic microarray based identification of placental mRNA in maternal plasma by subtractive gene expression analysis.

Circulating DNA and RNA in cancer testing
Application of circulating DNA in plasma in cancer testing depends on the accumulation of genetic and epigenetic changes, such as 1) point mutations 2) chromosomal rearrangements 3) microsatellite instability and 4) hypermethylation (35). Circulating N-ras and K-ras gene mutations have been observed in circulating DNA in various cancer forms and persistence of mutated circulating K-ras sequences has been related to recurrence or progressive disease (42). Microsatellite instability, particularly loss of heterozygosity (LOS), has been observed both in the tumor itself and in the corresponding circulating DNA. Also these changes have been correlated with disease progression or recurrence. Real time methylation specific polymerase chain reaction (RT-MSP-PCR) allows quantitative estimates of promoter hypermethylated circulating DNA. Finally, particular viral sequences has been reported in circulation in patients suffering from some Epstein-Barr virus (EBV) associated cancers (nasopharyngeal carcinoma, EBV-associated Hodgkin's disease) as has circulating human papilloma virus (HPV) DNA sequences in plasma of cervical cancer patients associated with metastasis (43).

Circulating mRNA
The finding of mRNA in plasma from patients with malignant melanoma (44) came as a surprise. Apparently cell-derived circulating mRNA is protected from degradation in plasma (45), possibly because of "apoptotic packing". Telomerase mRNAs (hTERT-Telomerase Reverse Transcriptase) in plasma has been found in several cancer forms (1,46). Lung cancer disease was detected in 100% of patients using Her2/neu and hnRNP-B1 serum mRNA as markers (47). Several mRNA markers in plasma have been demonstrated in breast cancer, and of lately *erbB2* mRNA in plasma has been associated with circulating tumor cells and negative estrogen and progesterone receptor status (48).

WHAT HOLDS FOR THE FUTURE?

Scientists frequently display low predictive powers in forecasting future scientific forthcomings. We shall therefore be careful in trying to outline the future developments. We feel sure, however, that the generation of databases containing large number of SNPs, the characterization of haplotypes and patterns of linkage disequilibrium throughout the genome will provide an opportunity for the better understanding of susceptibility to disease, prognosis of disease and responses to drugs. Only the careful use of these strategies and a clear understanding of their statistical limits will allow novel genetic variants for many of the common diseases to be determined as biomarkers. Furthermore, we would predict various types of cards to profile genes predictive of cardiovascular, diabetic and cancer diseases. As to what extent the public themselves, the life insurance companies or the medical professions will advocate such use may be both an ethical and a political question. We also foresee the increasing use of extracted DNA information from non-coding sequences and from epigenetic changes for the understanding of common diseases.

The application of quantitative estimate of mRNAs, whether obtained on global expression platforms or by low density arrays, to understand disease signatures in circulating leukocytes will obviously increase in years to come. Disease up or down regulated genes that code for proteins that potentially may be released into the blood stream, thus giving rise to potential protein biomarkers, may become in focus. As bioinformatics software tools become available and more simplified, more disease specific information will be mined. To what extent endogenous siRNAs will add to the messengers remains elusive. Obviously, the highly sophisticated and complicated, intracellular RNA languages will need the cooperativity of several professions to advance the understanding of systems biology.

As for plasma DNA and mRNA, new useful molecular markers of cancer will appear. Their value in monitoring or predicting disease, however, will depend on increased basal understanding of how these circulating nucleic acids are formed, and how they enter (and leave) the plasma and other biological compartment as well as well designed clinical studies with well delineated clinical endpoints.

METHODS

Isolation of nucleic acids from biological fluids

DNA is a stable molecule which easily can be isolated from biological fluids such as whole blood, cerebrospinal fluid and tissues. RNA, however, may already on sampling be degraded by RNases if not immediately stabilized. With proper stabilization, DNA and RNA molecules can be isolated both manually or robotized from a cell lysate.

Available methods for isolation of nucleic acids may be based on organic extraction or adherence either to magnetic beads or columns. The method of choice should document integrity and isolated molecules should be representative for the content of DNA or RNA in the cell lysates.

Reverse Transcription

Methods used for quantification of mRNAs can not use mRNA molecules directly in their synthesis. For quantification of specific transcripts a process called reverse transcription has to synthesize cDNA molecules with mRNA molecules as templates. This provides a cDNA library synthesized from all the mRNA molecules (the transcriptome) in the lysate where furthermore PCR primers can pick the specific target in the PCR reaction (49). Global gene expression profiles on microarray platforms are dependent on mRNA molecules labeled with either fluorogenic or biotinylated molecules. This involves several steps before a sample can be hybridized to an array.

Real-time PCR

The Polymerase Chain Reaction is a laboratory technique that can amplify one molecule of cDNA or DNA and produce measurable amounts of identical cDNA/DNA amplicons by performing repetitive enzymatic cycles. PCR primers (sequence specific oligos) make it possible to pick out the specific sequence of a mRNA target. Real-time PCR enables monitoring of accumulated amplicons after each cycle due to labeling of primers, probes or amplicons with fluorogenic molecules (50,51). This approach has made it possible to quantitate specific transcripts using either calibration curve methodology (52) or relative quantification (ΔΔCt-method) (53).

Real-time PCR equipment offers the opportunity to perform melting point analysis. Labelled, sequence specific probes will have melting temperatures dependent on the match between the probe and the sample. This approach is utilized for genotyping of SNPs (54).

Microarrays

Different microarray platforms are available today using different types of arrays; spotted cDNA arrays or oligonucleotide arrays for gene expression analysis and (55) and mapping arrays for SNPs, linkage analysis, whole genome association, population genetics, and chromosomal copy number changes (56). However, the basic principle which relies on nucleic acid hybridization between labeled targets and large arrayed sets of gene fragments on solid support, is the same for most microarrays (57,58). One of the challenges is to decide an analytical strategy for identification of new genes or disease mechanisms. One criterion often used is a twofold difference in expression levels. Appropriate statistical analysis is needed to analyze such huge amounts of data. As all array technologies of today are expensive and the data volume enormous, the researchers approaching these technology platforms face several challenges. Experimental design should be carefully planned and ideally biostatistics expertise included from the start. Such expertise will subsequently be needed in mining the data. Considerations in appropriate biobanking of collected specimens will be necessary. In short preanalytical, analytical and postanalytical insight into the biology involved, will be mandatory to obtain reliable results.

REFERENCES

1. Johnson PJ, Lo YM. Plasma nucleic acids in the diagnosis and management of malignant disease. *Clin Chem* 2002; **48**: 1186-93.
2. Ogawa M. Differentiation and proliferation of hematopoietic stem cells. *Blood* 1993; **81**: 2844-53.
3. Peltonen L, McKusick VA. Genomics and medicine. Dissecting human disease in the postgenomic era. *Science* 2001; **291**: 1224-9.
4. Chakravarti A. Population genetics – making sense out of sequence. *Nat Genet* 1999; **21**: 56-60.
5. Crawford DC, Akey DT, Nickerson DA. The patterns of natural variation in human genes. *Annu Rev Genomics Hum Genet* 2005; **6**: 287-312.
6. Esteller M. Aberrant DNA methylation as a cancer-inducing mechanism. *Annu Rev Pharmacol Toxicol* 2005; **45**: 629-56.
7. von Zglinicki T, Martin-Ruiz CM. Telomeres as biomarkers for ageing and age-related diseases. *Curr Mol Med* 2005; **5**: 197-203.
8. Bertina RM, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, et al. Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature* 1994; **369**: 64-7.
9. Poort SR, Rosendaal FR, Reitsma PH, Bertina RM. A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. *Blood* 1996; **88**: 3698-703.
10. Feder JN, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy DA, Basava A, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet* 1996; **13**: 399-408.

11. Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I. Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 2002; **30**: 233-7.

12. Allen WL, Johnston PG. Role of genomic markers in colorectal cancer treatment. *J Clin Oncol* 2005; **23**: 4545-52.

13. Miyamoto K, Ushijima T. Diagnostic and therapeutic applications of epigenetics. *Jpn J Clin Oncol* 2005; **35**: 293-301.

14. Maier S, Dahlstroem C, Haefliger C, Plum A, Piepenbrock C. Identifying DNA methylation biomarkers of cancer drug response. *Am J Pharmacogenomics* 2005; **5**: 223-32.

15. Ingelman-Sundberg M. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends Pharmacol Sci* 2004; **25**: 193-200.

16. Ingelman-Sundberg M. Pharmacogenetics: an opportunity for a safer and more efficient pharmacotherapy. *J Intern Med* 2001; **250**: 186-200.

17. Wadelius M, Chen LY, Downes K, Ghori J, Hunt S, Eriksson N, et al. Common VKORC1 and GGCX polymorphisms associated with warfarin dose. *Pharmacogenomics J* 2005; **5**: 262-70.

18. Bustin SA, Mueller R. Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. *Clin Sci (Lond)* 2005; **109**: 365-79.

19. Zebisch A, Linkesch W, Sill H. Bedside RNA stabilizing kit systems for gene expression analysis of acute leukemias: influence of non-neoplastic white blood cells. *Leukemia* 2005; **19**: 685.

20. DePrimo SE, Wong LM, Khatry DB, Nicholas SL, Manning WC, Smolich BD, et al. Expression profiling of blood samples from an SU5416 Phase III metastatic colorectal cancer clinical trial: a novel strategy for biomarker identification. *BMC Cancer* 2003; **3**: 3.

21 Zieglschmid V, Hollmann C, Bocher O. Detection of disseminated tumor cells in peripheral blood. *Crit Rev Clin Lab Sci* 2005; **42**: 155-96.

22. Ma J, Liew CC. Gene profiling identifies secreted protein transcripts from peripheral blood cells in coronary artery disease. *J Mol Cell Cardiol* 2003; **35**: 993-8.

23. Feezor RJ, Baker HV, Mindrinos M, Hayden D, Tannahill CL, Brownstein BH, et al. Whole blood and leukocyte RNA isolation for gene expression analyses. *Physiol Genomics* 2004; **19**: 247-54.

24. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO. Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci USA* 2003; **100**: 1896-901.

25. Pahl A. Gene expression profiling using RNA extracted from whole blood: technologies and clinical applications. *Expert Rev Mol Diagn* 2005; **5**: 43-52.

26. Fan H, Hegde PS. The transcriptome in blood: challenges and solutions for robust expression profiling. *Curr Mol Med* 2005; **5**: 3-10.

27. Imbeaud S, Graudens E, Boulanger V, Barlet X, Zaborski P, Eveno E, et al. Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Res* 2005; **33**: e56.

28. Schutten M, Niesters HG. Clinical utility of viral quantification as a tool for disease monitoring. *Expert Rev Mol Diagn* 2001; **1**: 153-62.

29. Mulder J, McKinney N, Christopherson C, Sninsky J, Greenfield L, Kwok S. Rapid and simple PCR assay for quantitation of human immunodeficiency virus type 1 RNA in plasma: application to acute retroviral infection. *J Clin Microbiol* 1994; **32**: 292-300.

30. Sun R, Ku J, Jayakar H, Kuo JC, Brambilla D, Herman S, et al. Ultrasensitive reverse transcription-PCR assay for quantitation of human immunodeficiency virus type 1 RNA in plasma. *J Clin Microbiol* 1998; **36**: 2964-9.

31. Rohde G, Borg I, Arinir U, Bufe A, Dorsten C, Schultze-Werninghaus G, Bauer TT. Evaluation of a real-time polymerase-chain reaction for severe acute respiratory syndrome (SARS) associated coronavirus in patients with hospitalised exacerbation of COPD. *Eur J Med Res* 2004; **9**: 505-9.

32. Hochhaus A, Reiter A, Skladny H, Reichert A, Saussele S, Hehlmann R. Molecular monitoring of residual disease in chronic myelogenous leukemia patients after therapy. *Recent Results Cancer Res* 1998; **144**: 36-45.

33. Jones CD, Yeung C, Zehnder JL. Comprehensive validation of a real-time quantitative bcr-abl assay for clinical laboratory use. *Am J Clin Pathol* 2003; **120**: 42-8.

34. Mandel P, Metais P. Les acides nucleiques du plasma sanguin chez l`homme. *Acad Sci Paris* 1948; **142**: 241-3.

35. Tong YK, Lo YM. Diagnostic developments involving cell-free (circulating) nucleic acids. *Clin Chim Acta* 2006; **363**: 187-96.

36. Tsui NB, Ng EK, Lo YM. Stability of endogenous and added RNA in blood specimens, serum, and plasma. *Clin Chem* 2002; **48**: 1647-53.

37. Ng EK, Tsui NB, Lau TK, Leung TN, Chiu RW, Panesar NS, et al. mRNA of placental origin is readily detectable in maternal plasma. *Proc Natl Acad Sci USA* 2003; **100**: 4748-53.

38. Laktionov PP, Tamkovich SN, Rykova EY, Bryzgunova OE, Starikov AV, Kuznetsova NP, Vlassov VV. Cell-surface-bound nucleic acids: Free and cell-surface-bound nucleic acids in blood of healthy donors and breast cancer patients. *Ann N Y Acad Sci* 2004; **1022**: 221-7.
39. Ng EK, Tsui NB, Lam NY, Chiu RW, Yu SC, Wong SC, et al. Presence of filterable and nonfilterable mRNA in the plasma of cancer patients and healthy individuals. *Clin Chem* 2002; **48**: 1212-7.
40. Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, Wainscoat JS. Presence of fetal DNA in maternal plasma and serum. *Lancet* 1997; **350**: 485-7.
41. Tsui NB, Chim SS, Chiu RW, Lau TK, Ng EK, Leung TN, et al. Systematic micro-array based identification of placental mRNA in maternal plasma: towards non-invasive prenatal gene expression profiling. *J Med Genet* 2004; **41**: 461-7.
42. Ryan BM, Lefort F, McManus R, Daly J, Keeling PW, Weir DG, Kelleher D. A prospective study of circulating mutant KRAS2 in the serum of patients with colorectal neoplasia: strong prognostic indicator in postoperative follow up. *Gut* 2003; **52**: 101-8.
43. Pornthanakasem W, Shotelersuk K, Termrungruanglert W, Voravud N, Niruthisard S, Mutirangura A. Human papillomavirus DNA in plasma of patients with cervical cancer. *BMC Cancer* 2001; **1**: 2.
44. Kopreski MS, Benko FA, Kwak LW, Gocke CD. Detection of tumor messenger RNA in the serum of patients with malignant melanoma. *Clin Cancer Res* 1999; **5**: 1961-5.
45. Hasselmann DO, Rappl G, Tilgen W, Reinhold U. Extracellular tyrosinase mRNA within apoptotic bodies is protected from degradation in human serum. *Clin Chem* 2001; **47**: 1488-9.
46. Lledo SM, Garcia-Granero E, Dasi F, Ripoli R, Garcia SA, Cervantes A, Alino SF. Real time quantification in plasma of human telomerase reverse transcriptase (hTERT) mRNA in patients with colorectal cancer. *Colorectal Dis* 2004; **6**: 236-42.
47. Goebel G, Zitt M, Zitt M, Muller HM. Circulating nucleic acids in plasma or serum (CNAPS) as prognostic and predictive markers in patients with solid neoplasias. *Dis Markers* 2005; **21**: 105-20.
48. Xu Y, Yao L, Li H, Ouyang T, Li J, Wang T, et al. Presence of erbB2 mRNA in the plasma of breast cancer patients is associated with circulating tumor cells and negative estrogen and progesterone receptor status. *Breast Cancer Res Treat* 2005; 1-7.
49. Bustin SA, Nolan T. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J Biomol Tech* 2004; **15**: 155-66.
50. Haug KBF, Ovstebo R, Kierulf P. How to monitor gene expression. *NBS-nytt* 2003; **4**: 21-5.
51. Wittwer CT, Herrmann MG, Moss AA, Rasmussen RP. Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques* 1997; **22**: 130-8.
52. Ovstebo R, Haug KBF, Lande K, Kierulf P. PCR-based calibration curves for studies of quantitative gene expression in human monocytes: Development and evaluation. *Clin Chem* 2003; **49**: 425-32.
53. ABI Prism 7900HT sequence Detection System and SDS Enterprise Database; User Guide. Basic Operation and Maintenance. Applied Biosystems, 2002.
54. Ririe KM, Rasmussen RP, Wittwer CT. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal Biochem* 1997; **245**: 154-60.
55. Watson A, Mazumder A, Stewart M, Balasubramanian S. Technology for microarray analysis of gene expression. *Curr Opin Biotechnol* 1998; **9**: 609-14.
56. Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, Liu G, et al. Large-scale genotyping of complex DNA. *Nat Biotechnol* 2003; **21**: 1233-7.
57. Churchill GA. Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 2002; **32** Suppl: 490-5.
58. Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 2001; **19**: 342-7.