Published in final edited form as:

Expert Opin Med Diagn. 2008 November 1; 2(11): 1205–1219. doi:10.1517/17530059.2.11.1205.

New Technologies for Diagnosing Pediatric Tumors Expert Opinion on Medical Diagnostics

Jun S. Wei, Thomas C. Badgett, and Javed Khan

Oncogenomic Section, Pediatric Oncology Branch, Advanced Technology Center, Center for Cancer Research, National Cancer Institute, National Institutes of Health, 8717 Grovemont Circle, Bethesda, Maryland 20892-4605, USA

Abstract

Background—The completion of Human Genome Project (HGP) has paved the way for novel, more detailed and accurate molecular diagnostic classification of cancer. With the information from the HGP, cancers can be categorized not only on the morphology or limited immunohistological markers, but according to their “molecular fingerprints” such as gene expression profiles. Technologies detecting these signatures have been developed to simultaneously measure multiple genes or proteins in one assay with high sensitivity and specificity.

Objective—To evaluate potential innovative novel methods of diagnosis and prognosis in pediatric cancers.

Methods—We selected a variety of promising new diagnostic technologies utilizing molecular signatures which harness the results from HGP including DNA microarray, bead-based detection system, multiplexed RT-PCR, MesoScale Discovery (MSD), and isotope-coded affinity tag (ICAT), as well as their applications in biomarker discovery for pediatric tumors. Label-free detection technologies and the obstacles for taking these new diagnostic technologies from the bench to the bedside are also discussed.

Conclusion—The use of molecular signatures is gaining acceptance in clinical practice. However, technical challenges need to be addressed before incorporating these new technologies into current diagnostic and prognostic schema.

Keywords

Bead-based detection; comparative genomic hybridization (CGH); diagnostics; isotope-coded affinity tags (ICAT); label-free detection; MesoScale Discovery (MSD); microarray; multiplex PCR; massively parallel DNA sequencing; microRNA (miRNA); prognostics; reverse phase protein array (RPPA)

1. Introduction

The progressive accumulation of genetic and epigenetic alterations is the hallmark of cancer. Traditional diagnostic techniques such as Southern, Northern, Western analyses, cytogenetics, immunohistochemistry and polymerase chain reaction (PCR) have been successful in identifying translocations, losses or gains of large chromosomal regions, and alteration of gene expression involved in cancer. However, these methods are limited by either low-resolution, as in the case of cytogenetics, or insufficient throughput, examining a single gene or

chromosomal region at a time. In addition, these diagnostic tools are prone to false negatives due to the technical difficulties on the limited markers examined, and also do not provide insight into global, non-random, genomic alterations that occur in cancers.

The Human Genome Project (HGP) was essentially completed in 2003 after 13 years' effort through international collaborations [1]. The HGP has heralded the way for high throughput analyses of whole genomes and proteomes with the hope that these techniques will identify the fundamental differences between normal and diseased cells including cancer. Since the completion of sequencing of the human genome, the majority of estimated 20-25,000 protein-coding genes and over 700 non-coding small RNAs termed microRNA (miRNA) have thus far been identified. Much of these resources including physical clones, mapping and sequence data, produced by both the public [1] and the private sectors [2], are available to the public over the internet [3-5]. These resources, combined with rapid increases in computing processing speeds and world wide internet access, have stimulated an explosion of techniques utilizing "Omics"-based tools in cancer diagnostics and prognostics. These new high-throughput approaches, such as DNA microarray technology for detection of nucleic acids (RNA or DNA) and Isotope-coded Affinity Tags (ICAT) for proteins, have the capacity to interrogate tens of thousands of genes and their protein products (the whole human genome) in a single experiment. They have become powerful tools aiding the discovery of novel molecular markers and mechanisms of disease.

With the information from the human genome project, cancers can be categorized by their "molecular fingerprint" or gene expression signatures, rather than by traditional classification schemes, relying on only the morphology and limited sets of disease specific markers. Collaborative efforts such as the Cancer Genome Anatomy Project (CGAP;[6]) between National Cancer Institute (NCI) and National Center for Biotechnology Information (NCBI); and the Cancer Genome Atlas (TCGA;[7]) between NCI and National Human Genome Research Institute (NHGRI) are aiming to decipher the molecular anatomy of cancer cells through the application of genome analysis technologies. With the goal to increase benefits to children from advances in molecular-targeted cancer therapeutics, the NCI and the Foundation for the National Institutes of Health (FNIH) have recently established the Childhood Cancer Therapeutically Applicable Research to Generate Effective Treatment (TARGET) initiative specifically to identify and validate therapeutic targets in childhood cancers using genomic technologies including array-based characterization of genomic and transcriptome profiles, large-scale resequencing, and RNA interference/small molecule screening. Studies through the TARGET initiative have been generating high-resolution genomic and transcriptome profiles for acute lymphoblastic leukemia (ALL) and neuroblastoma (NB), the two pilot projects for the initiative. Discovery of novel biomarkers resulting from these systematic studies as well as rapid advance of technology are expected to make the diagnosis of cancer more precise, faster, and more cost-effective.

This review will provide a brief survey of new technologies such as DNA microarray, bead-based detection, multiplexed PCR and protein detection methods emerging from the "Omics" studies. Although we focus on a subset of pediatric cancers in this review, the new technologies should be applicable to all pediatric malignancies. Some of these technologies are mature enough to be applied in the clinical setting while others require further refinement. These new technologies will revolutionize how clinical medicine is practiced and will make personalized medicine possible in the very near future.

2. Current clinical challenges in diagnosis of pediatric cancers

Despite of the advances in pediatric oncology and steady improvement of overall survival rate in children with cancers over the past several decades [8], there are several clinical challenges

in diagnostics for pediatric tumors. First, accuracy of diagnosis and risk-stratification are crucial, because clinical outcomes are directly linked to the appropriate therapy for the specific malignancy. The importance of accurate diagnosis and characterization of pediatric cancers will become more critical with the introduction of targeted therapies. Even with increasing numbers of tumor-specific markers, some childhood malignancies, such as undifferentiated small round blue cell tumors (SRBCT) with their uniform morphology, occasionally continue to be difficult to diagnose and thus determine the best course of action. Second, the slow speed of diagnosis continues to hinder implementation of treatment plans and often frustrates both families and physicians. Conventional characterization of childhood cancers relies on a multimodal approach often including morphologic review of pathology, immunohistochemistry, cytogenetics, flow cytometry and occasionally analysis of a handful of genes with known prognostic implications. Assembling the results from these diagnostic tests can take days to weeks, delaying therapy. Third, traditional risk-stratification schemes have limited predictive power. Despite significant improvements in overall survival for pediatric malignancies over the past 25 years [8], rates of cure remain low for patients with metastatic disease. Using existing methods for risk-stratification, more than 70% of patients categorized with high-risk disease will eventually succumb to their diseases even with aggressive multimodal therapies [9-12]. Unfortunately, current classification schema cannot delineate which patients, whether high or low risk, will respond to the existing therapies and which will require novel approaches. Tailoring therapy for individual patients in an effort to limit long term side-effects, including neuro-cognitive sequelae, growth delay, infertility, and secondary malignancies, while maintaining high rates of cure, is crucial in pediatric oncology. Pediatric patients are exposed to cytotoxic therapies during normal stages of growth and development, and once cured are expected to live for a half a century or more. The challenges confronting pediatric oncologists require increased precision, speed and predictive power of diagnosis, which can be addressed by new diagnostic technologies with the goal of improving survival and enhancing the quality of life for children with cancer.

3. New Diagnostic Technologies

3.1 DNA Microarray

Since the first report of this technology in 1995 [13], DNA microarray technology has become one of the most important and widely used genomic tools for cancer researches. This high-throughput technology uses known synthetic DNA sequences (features; which can be double-stranded cDNA, short (25nt.), or long (50-70nt.) oligonucleotides) attached to a solid support such as a glass slide or beads to measure the level of nucleic acids including RNA and DNA in a biological sample. A modern whole-genome microarray contains millions of features that can interrogate the entire genome in one experiment. Combined with a plethora of statistical methods and powerful computer algorithms such as artificial intelligence, DNA microarray technology is being rapidly incorporated into many aspects of cancer research including tumor classification, outcome prediction, identification of novel molecular targets, characterization of genes of unknown function, and eventually leading to personalized therapies guided by molecular signatures. In 2007, the first DNA microarray-based prognostic test for breast cancer (MammaPrint) was approved by the Food and Drug Administration (FDA), heralding the advent of signature-based diagnostic tools in the clinical setting. Here, we briefly describe several studies demonstrating the use of DNA microarray analysis in diagnosis and prognosis of pediatric tumors.

3.1.1 RNA-based DNA microarray studies—Gene expression profiling using microarrays has been successfully utilized for classification, diagnosis, and predicting patient outcome in pediatric malignancies [14-17]. As mentioned above, pediatric cancers that lack of morphological distinction such as SRBCTs occasionally continue to pose diagnostic challenges

in clinic. To approach the diagnostic problem of undifferentiated SRBCTs, we performed expression profiling using cDNA microarrays and identified a cancer-specific signature by applying artificial neural networks (ANNs) which are computer-based machine learning artificial intelligence algorithms modeled on the structure and behavior of neurons in human brain capable of being trained to recognize complex patterns through reiterative learning cycles [16]. This expression signature contained 93 genes differentially expressed in four different classes of pediatric cancers which share very similar clinical histology including neuroblastoma, rhabdomyosarcoma, Ewing's sarcoma, and Burkett's lymphoma (Figure 1A, a and b). The performance of ANN using this expression signature is very robust (specificity=100% and sensitivity >93% for all categories) confirming its clinical value as a diagnostic assay (Figure 1A, c). Among these 93 genes, 41 had not been previously reported to associate with these cancers, and they can be potentially used as biomarkers for clinical diagnosis. In addition, because these discriminating genes identified by DNA microarray experiments are specific expressed by a particular cancer, they are thought to be the most biologically relevant to the cancers, and logic targets for therapy. For example, we found that *FGFR4*, a protein-tyrosine kinase receptor, was one of the highly expressed discriminating genes in RMS identified in this study. It is expressed only in the normal developing muscle tissue, but absent in normal mature muscle tissue[18] and represents a druggable target for novel therapy. This study demonstrated that DNA microarray technology can be used as an accurate diagnostic tool to correctly classify cancers into more than two categories without additional clinical information. Furthermore, we used the findings from this study to develop a multiplexed RT-PCR assay for clinical diagnosis of SRBCTs (see below).

Recently, a class of non-coding small RNAs, specifically miRNAs, has been discovered to have important regulatory functions in plants and animals [19]. These highly conserved, ~21nt. RNAs regulate the expression of genes through translational inhibition or RNA degradation by preferentially binding to the 3'-untranslated regions (3'-UTR) of specific mRNAs [19]. Each miRNA is thought to target ~200 or more genes, and multiple miRNAs can target a single gene. The entire human genome is predicted to have approximately 1000 miRNA genes, and so far 722 miRNAs (version 10.0) have been reported to be expressed in the human cells [20]. miRNAs have been described to have diverse functions involved in embryogenesis, metabolism, cell growth, differentiation, and apoptosis [19]. In recent years, there has been an explosion of miRNA studies in the literature demonstrating the alteration of miRNA expression in various human cancers [21-32], which can be developed into diagnostic biomarkers.

miRNA expression patterns represent an innovative mechanism for characterizing human cancers. Using miRNA expression profiling as a diagnostic test has several advantages over messenger RNA profiles (mRNA). First, the expression patterns of miRNAs can classify cancers and may be better predictors of diagnosis of human cancers than mRNA-based methods. In a recent study, Lu *et al.* demonstrated miRNA profiles could be used to classify cancers according to their developmental origins and mechanisms of transformation [32] (Figure 1B, a and b). Furthermore, when comparing miRNA to mRNA profiles in distinguishing poorly differentiated metastatic tumors, they correctly classified 12 out of 17 tumors using miRNA profiles, but they could only correctly classified one tumor using mRNA profiles [32] (Figure 1B, c). Second, due to the relatively small number of miRNA species (estimated around a thousand) compared to mRNAs (estimated in the tens of thousands), miRNA expression patterns are expected to be less complex. Finally, because of the short length of miRNAs, they are more stable and thus more reliably extracted from paraffin sections [33]. However, since the miRNA field is still new, there are many technical issues which need to be worked out, such as methods of detection and normalization, before miRNAs can be used as a diagnostic tool.

In addition to the diagnosis of pediatric tumors, prognostic prediction has is critical for treating pediatric cancer patients with modern multimodal therapies, because the outcome is directly dependent on the accurate risk-stratification. We embarked on a study to identify a prognostic gene expression signature for NB (one of the SRBCTs) patients using DNA microarray technology (Figure 2A) [17]. Patients with NB in North America are currently stratified by the Children's Oncology Group (COG) into high-, intermediate- and low-risk based on age, tumor staging (i.e. international neuroblastoma staging system; INSS), Shimada histology, *MYCN*-amplification, and DNA ploidy [34]. Despite careful stratification, the survival rate for patients with high-risk NB remains less than 30%, and it is currently not possible to predict which of these high-risk patients will survive or succumb to the disease. We hypothesized that the mRNA expression profiles of neuroblastoma tumors contain "prognostic information" at presentation, which can be detected by computer algorithms to predict outcome of individual patients. We have therefore performed a pilot study to identify prognostic gene expression signatures using cDNA microarrays and artificial neural networks (ANNs). Utilized an ANN-based gene minimization strategy we identified 19 genes, including two previously reported prognostic markers, *MYCN* and *CD44*, whose expression signature correctly predicted the outcome for 98% of all patients (Figure 2A, a and b), and compared favorably with current COG risk-stratification (Figure 2A, c). In particular, these 19 predictor genes were able to sub-partition COG high-risk patients without *MYCN*-amplification into two subgroups according to their survival status (Figure 2A, d), of which no current clinical criteria can predict the outcome of patients in this risk group. Therefore, our findings demonstrated that a gene expression signature can predict the prognosis of patients with neuroblastoma independent of currently known risk factors, and the feasibility of using a small number of genes for outcome prediction which may allow physicians to customize therapy for individual patients according to their molecular profiles. Moreover, these prognostic genes probably play important roles in NB biology, and represent potential targets for novel therapies. Since the publication of this 19-genes signature, there have been four other published studies predicting outcome in patients with NB [35-38]. Comparative analysis of all these five NB prognostic signatures revealed remarkably few overlapping genes among them. This raises an important question whether these profiles will be generalizable and will have clinical utility for predicting prognosis. Therefore, further validation studies need to be performed and are underway before these markers can be taken to the clinic as FDA approved prognostic signatures.

3.1.2 Genomic DNA-based DNA microarray studies—In addition to gene expression profiling, DNA microarray technology has been utilized to study genomic imbalances in cancers known as array-based comparative genomic hybridization (A-CGH), which is a genome-wide scanning technique for detecting gain, amplification or loss of chromosomal regions in cancer. Compared to the traditional metaphase CGH technique which has a poor resolution of 5-10 mega bases, A-CGH can examine genomic DNA at a much higher resolution such as at the exon level (hundreds of bases) on fine-tiling arrays. There are two potential advantages for using genomic DNA signatures rather than those of RNAs for diagnosis and prognosis. First, genomic DNA signatures derived from DNA copy number alterations are simpler when compared with the complex regulatory networks governing mRNA transcription. This is because the normal state of a cell has only two DNA copies (diploid), whereas gene amplifications and deletions are common mechanisms by which cancers gain survival advantages. Another advantage of using genomic DNA signature is the stability of genomic DNA since it is much more resistant to degradation than RNAs, making it easier to acquire high quality genomic DNA. Like its predecessor of meta-phase CGH, A-CGH cannot detect copy-neutral genomic alterations such as translocation and copy-neutral loss of heterozygosity. However, the genome-wide single nucleotide polymorphisms arrays like SNP arrays can identify such copy-neutral loss of heterozygosity, and may prove to be useful in detecting biologically relevant regions. This genome-wide SNP platform has been successfully used in

investigating pediatric ALL [39]. Mullighan *et. al.* used the DNA copy number information to identify genes that are frequently deleted for sequencing [39]. This approach identified deletion, amplification, point mutation and structural rearrangement in genes regulating B cell development and differentiation. They identified the *PAX5* to be the most frequently mutated gene occurring in 31.7% of cases demonstrating the power of these approaches to discover novel tumor suppressor genes [39].

Our laboratory tested if using the chromosomal imbalance signatures could predict outcome for neuroblastoma patients. Currently, several chromosomal imbalances have been identified to be predictive for patient outcome, which include *MYCN*-amplification [40], 1p36 and 11q23 loss [41,42]. However, in the US only *MYCN*-amplification is currently used to stratify patients in clinic, although future trials are being developed incorporating LOH information for stratification. At present, it is impossible to predict which high-risk patients will respond to the current therapies, especially in those of INSS stage 4 neuroblastoma patients older than 500 days without *MYCN*-amplification, which represent the most challenging high-risk group patients with this disease. Using A-CGH, we identified a DNA copy number-based prognostic profile for this group of patients. Patients with gain or loss of two or more chromosomes (whole chromosome change; WCC) were reliably predicted to benefit from the current multimodal therapies; whereas patients in this risk group without this chromosomal trait would likely fail the current standard therapies [43] (Figure 2B, a). Intriguingly, we found that this chromosomal prognostic signature is dependent on the diagnosis date, which $WCC \geq 2$ is significantly correlated with survival for patients diagnosed after 1998, but not before or in 1998 (Figure 2B, b and c). Although lack of detailed treatment information for patients in this study prevented a concrete conclusion, we speculated that introduction of multimodal therapy including the combination of high dose chemotherapy, autologous bone marrow transplantation, and subsequent treatment of 13-*cis* retinoid acid around that time could explain this phenomena. In addition, this study demonstrated that this new copy-number prognostic marker is independent of known prognostic factors, and complements the current risk-stratification (Figure 2B, d). These observations imply that WCC is a potential pharmacogenomic marker for the response of patients in this risk group to the current therapy. After confirmatory validation studies, such biomarker will become extremely useful in clinic to guide treatment, and will make personalized therapy possible.

3.2 Bead-based detection technology

Bead-based detection technology is a variation of the microarray technology using the principals of flow cytometry (Figure 3). Instead of cells in the flow chamber, polystyrene beads with addressable color codes are utilized to immobilize DNA probes or antibodies for detection of nucleic acids or antigens on the functionalized bead surface. When the beads are forced through the flow chamber as a single string, two laser beams simultaneously interrogate the identity of beads and the quantity of fluorescent signal on the bead surface, allowing multiplexed assays of analytes in a single sample. Unlike microarrays which can measure tens of thousands of genes simultaneously, bead-based detection technology allows measuring nucleic acids or antigens of a limited number (one hundred). There are many advantages for using bead-based detection technology. First, due to the simplicity of the technology and low cost of plastic beads, bead-based detection technology is much cheaper compared to microarray technology which requires special setups and expensive equipment. Second, binding kinetics in liquid are much efficient and faster than on the planar surfaces of microarrays. Third, the bead-based detection has been reported to have wider dynamic range [32]. Therefore, this technology is an ideal cost-effective high-throughput platform for detecting a signature consisting of a limited number of genes or proteins. Bead-based technologies have been used to classify undifferentiated metastatic human cancers [32], and are capable of identifying multiple forms of genetic translocation lesions in acute leukemia for risk-stratification [44].

Currently, the diagnosis and risk-stratification of leukemia can be time-consuming and labor intensive, requiring multiple diagnostic assays to identify the specific chromosomal translocations. In this report, the authors were able to achieve 100% sensitivity and specificity in detecting all assayed translocations within 6 hours of sample collection [44], demonstrating the robustness of the bead-based detection method for clinical use.

3.3 Multiplexed PCR-based technology

Like the bead-based detection technology, there PCR-based technologies for detecting a limited number of genes such as Taqman® low density arrays and multiplexed PCR-based technology. Taqman® low density arrays [45] use microfluidics and automation to perform individual real-time PCR reaction simultaneously. The advantage of this technology is its utilization of Taqman® chemistry which has been well accepted as the accurate and rapid method for measuring gene expression. Multiplexed PCR-based technology is developed for detecting a limited number of nucleic acid species such as gene expression signatures for diagnostic or prognostic purpose in a single PCR reaction. This technology combines the multiplex PCR for detection of specific nucleic acid species and fluorescence capillary electrophoresis techniques for identification of PCR products according to the designed length (Figure 4A). In a multiplexed RT-PCR application, two types of primers were designed for multiplex PCR amplification: chimeric primers and universal primers, so that the size of PCR product from each primer pair is unique to each specific gene. After reverse transcription of RNA, the chimeric primers containing gene specific sequence and universal primer sequence are used in the first stage of amplification. The second stage amplification uses one pair of fluorescent labeled universal primers for the multiplex amplification (Figure 4A, a). The mixture of amplicons is size-separated using capillary electrophoresis to identify the peak location (gene identity) and peak fluorescence intensity (gene expression level) (Figure 4A, b and c). Using 39 genes, a subset of a diagnostic gene signature identified in our previous microarray study of SBRCTs [16], we have adapted this technology combined with artificial intelligence (i.e. ANN) to successfully develop a clinical diagnostic assay for accurate classification of SRBCTs with extremely high specificity (>98%) and sensitivity (>96%) [46] (Figure 4B and C). FDA approval is currently being sought to use this assay in the clinical setting for diagnosis of SRBCTs.

3.4 Protein-based technologies

High throughput protein profiling technology is much more technical challenging, and has lagged behind its genomic counterparts. Nevertheless several new technologies have now emerged and are sufficiently mature enough to allow high throughput screenings to identify biomarkers of clinical importance. These methods include antibody arrays, “reverse phase” protein arrays (RPPA), the MesoScale Discovery (MSD) platforms (Gaithersburg, MD) [47], and Isotope-coded affinity tags (ICAT). Antibody arrays use immobilized antibodies to detect multiple proteins in one sample simultaneously on a supporting surface, whereas in RPPAs a miniaturized dot-blot of protein lysates is probed with antibodies [48]. Using a dilution curve, RPPAs offers a benefit of quantification over an extended dynamic range. As originally conceived of PRRA, samples from a large number of patients can be simultaneously profiled at one time. This approach has been applied successfully to a number of preclinical and clinical problems including rhabdomyosarcoma [49]. However, as this is a miniaturized dot-blot, extensive detail to antibody validation is required, and samples of high homogeneity are required. MSD assay uses electrochemiluminescence tags emitting light when electrochemically stimulated (Figure 5A). In this immunoassay, antibodies coated on the surface of plates are used to capture proteins from whole cell lysates, which can be then detected by an antibody labeled with an electrochemiluminescent tag. The advantage of this method is the ability to perform this assay using minute quantities of protein lysates in a 384-well multiplex format with minimal background due to the decoupling of stimuli (electricity) from

the signal (light). However, this technology is limited by the need for specific antibodies to detect a particular protein. On the contrary, the peptide-sequencing-based ICAT technology allows a quantitative measurement of protein levels in different cell types and tissues without the requirement of antibodies [50]. In this method, proteins are extracted from two tissues (*e.g.* diseased and normal), and are chemically modified using two different isotopes such as deuterium (d(8) and d(0)) respectively (Figure 5B). After labeling, proteins are mixed and digested with trypsin. The modified peptides are then subjected to affinity purification and then combined liquid chromatography/mass spectrometry. The mass spectrometer provides mass information that can be used to assign peptides to proteins of known sequence, and can discriminate the source of the peptides quantitatively on the basis of their difference in the tag mass. However, one limitation of this protein profiling technology is its requirement of cysteine residues in the protein for labeling with the tag. Therefore, approximately 5% of human proteins lack cysteine will be missed by this technology and those with multiple cysteines will generate complex mass spectrums which are difficult to interpret. Another disadvantage is that ICAT requires labeling the proteome with reactive tag molecules which can result in incomplete reactivity. In addition, due to its current high costs, it is now used only as a research tool, not as a routine assay. However, like any other new technologies, the cost will come down quickly once they are widely accepted and proven to be beneficial for patients.

In conclusion, HGP has facilitated a range of new technologies that can be used in diagnosis of pediatric cancers. These novel technologies will mature and be incorporated into clinical routine of diagnosis and risk-stratification, making personalized therapy possible in the near future. The field of pediatric oncology will embrace these new approaches in hopes decreasing morbidity associated with current therapies. Tailoring therapies to fit more precisely defined disease processes will limit the exposure of patients to undue toxicities and their concomitant late effects, with the promise of increasing cure rates and quality of life during and after treatment.

4. Expert opinion

Many of these “Omics” based technologies have proven to be powerful tools in discovery of new biomarkers for clinical applications. The use of molecular signatures as important clinical tools for diagnosis, prognosis and prediction of patient response to a particular therapy is gaining acceptance. However, technical challenges are needed to be addressed before a signature-based assay is incorporated into clinical practice. To be useful and practical in a clinical setting, assays must have a high specificity and sensitivity, simple techniques that can be performed by average lab technicians, rapid turn-around time, and cost-effectiveness. Some of these signature-based technologies are complex and require technical expertise to perform, lengthy sample manipulations and specialized expensive equipment, making them unsuitable for the clinical setting in their current form. Thus, research efforts should focus on improving existing technologies and developing new technologies which can meet the demands of clinical medicine.

In this era of technology explosion, it is difficult to predict which technology will meet the demands required for incorporation into common clinical practice. Regardless of which technologies are available, ultimately identification of clinically relevant biomarkers and their validation on a large scale will be hindered by the rarity of pediatric cancers. Based on SEER data [51], there are more than twice as many adult cancers diagnosed in one week than pediatric cancers diagnosed in one year. Because of the rarity, it is very difficult to perform a large experiment with proper size of discovery and validation cohorts that have statistical power. Genomic data such as that from microarray experiments often contain measurements for tens of thousands of descriptors for each sample, whereas the number of samples in a study is often relatively low, resulting in a risk of over fitting the data where differences between two groups

of patients, *e.g.* survivors *vs.* deceased patients, may be the result of random fluctuation. Small sample number and large number of measurements inherent in genomic studies of pediatric cancers, demand careful design of the training, testing and validation studies in large well-controlled cohorts. Therefore, studies seeking to identify of biomarkers for pediatric cancers are almost impossible and should not be conducted in one center or even one country, but need to involve international collaborative efforts to share patient samples as well as relevant clinical information.

5. Five-year view

New signature-based diagnostic and prognostic technologies are making personalized medicine possible. To date, most of these methods rely on labeling of biological samples with reporter molecules such as fluorescent and molecular tags, or radioactive labels. Often labeling reactions are not very efficient, and these reactions require extensive sample manipulation, and specialized instrumentation for signal detection and quantification. Thus, label-free technologies are promising alternatives to traditional artifactual signal-based methods. One of the labeling-free technologies is electronic-sensing technology, which detects the events of specific hybridization of nucleic acids using electronic circuits (field effect transistor; FET) [52]. In field effect transistors, the current *vs.* voltage characteristics or transconductance between the source and drain electrodes is strongly dependent upon the total charge accumulated at the gate terminal. This effect is well understood and can accurately estimate the amount of charge at the gate from measurement of the transconductance (Figure 5C, a and b). This phenomenon has been used to build carbon nanotube based transistor arrays for label-free detection of specific biological molecules such as DNA (Figure 5C, c) [53]. Another label-free technology is massively parallel DNA sequencing (next-generation sequencing or deep sequencing) technology which directly identify millions or even billions of nucleic acid species in parallel in one sequencing reaction without the requirement of cloning. Due to its large capacity, the massively parallel DNA sequencing technology not only generates sequence information for each nucleic acid strand, but also determines the abundance of each nucleic acid species. Thus, this technology has wide ranging applications including mutation identification, sequence and structure variations, expression profiling, promoter analysis, and epigenetic studies (*e.g.* DNA methylation). Unlike hybridization-based platforms requiring probe designed from reference genomes, this technology does not need prior knowledge of the samples being analyzed. Such properties make it sensitive enough to detect novel structural rearrangement of DNAs or RNAs including balanced translocations, which will not be detected using hybridization-based technologies. In addition, massively parallel sequencing technologies give a digital readout of levels for any sequence, even those at low levels beyond the detection of hybridization-based technologies. Therefore, these label-free technologies are promising tools for providing more accurate and rapid diagnoses essential for future implementation of personalized therapies.

6. Key issues

- Some of the undifferentiated pediatric tumors still occasionally pose a challenge for correct diagnosis.
- Currently lack of clinical tests that can predict individual patient outcome and signature-based new technologies will transform medicine from current generic standard therapy to future personalized medicine.
- Omics studies have generated molecular signatures for diagnosis, prognosis and prediction of response towards therapy for children with cancers.
- The rarity of pediatric cancers requires careful design for identification of robust signatures.

- Technical challenges for Omics technologies need to be addressed before development into clinical assays.
- High cost can be a limiting factor for the wide use of Omics technologies in clinic.
- Label-free detection technologies are likely to be easily adapted for clinical use.

Acknowledgment

We thank Dr. Stephen Hewitt for his helpful comments on the protein array. This manuscript is supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

References

1. <http://www.nhgri.nih.gov/HGP>
2. <http://www.celera.com/>
3. <http://genome.cse.ucsc.edu/>
4. <http://www.ncbi.nlm.nih.gov/>
5. <http://www.ensembl.org/>
6. <http://cgap.nci.nih.gov/>
7. <http://cancergenome.nih.gov/>
8. Linet MS, Ries LA, Smith MA, Tarone RE, Devesa SS. Cancer surveillance series: recent trends in childhood cancer incidence and mortality in the United States. *J Natl Cancer Inst* 1999;91(12):1051–1058. [PubMed: 10379968]
9. Krishnan K, Khanna C, Helman LJ. The biology of metastases in pediatric sarcomas. *Cancer J* 2005;11(4):306–313. [PubMed: 16197720]
10. Maris JM. The biologic basis for neuroblastoma heterogeneity and risk stratification. *Curr Opin Pediatr* 2005;17(1):7–13. [PubMed: 15659956]
11. Ludwig JA. Ewing sarcoma: historical perspectives, current state-of-the-art, and opportunities for targeted therapy in the future. *Curr Opin Oncol* 2008;20(4):412–418. [PubMed: 18525337]
12. Wang LL. Biology of osteogenic sarcoma. *Cancer J* 2005;11(4):294–305. [PubMed: 16197719]
13. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270(5235):467–470. [PubMed: 7569999]
14. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286(5439):531–537. [PubMed: 10521349]
15. Yeoh EJ, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 2002;1(2):133–143. [PubMed: 12086872]
16. Khan J, Wei JS, Ringner M, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001;7(6):673–679. [PubMed: 11385503]
17. Wei JS, Greer BT, Westermann F, et al. Prediction of clinical outcome using gene expression profiling and artificial neural networks for patients with neuroblastoma. *Cancer Res* 2004;64(19):6883–6891. [PubMed: 15466177]
18. deLapeyriere O, Ollendorff V, Planche J, et al. Expression of the Fgf6 gene is restricted to developing skeletal muscle in the mouse embryo. *Development* 1993;118(2):601–611. [PubMed: 8223280]
19. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;116(2):281–297. [PubMed: 14744438]
20. <http://microrna.sanger.ac.uk/>
21. Calin GA, Liu CG, Sevignani C, et al. MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc Natl Acad Sci U S A* 2004;101(32):11755–11760. [PubMed: 15284443]

22. Metzler M, Wilda M, Busch K, Viehmann S, Borkhardt A. High expression of precursor microRNA-155/BIC RNA in children with Burkitt lymphoma. *Genes Chromosomes Cancer* 2004;39(2):167–169. [PubMed: 14695998]
23. Hayashita Y, Osada H, Tatematsu Y, et al. A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res* 2005;65(21):9628–9632. [PubMed: 16266980]
24. Iorio MV, Ferracin M, Liu CG, et al. MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* 2005;65(16):7065–7070. [PubMed: 16103053]
25. Chan JA, Krichevsky AM, Kosik KS. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res* 2005;65(14):6029–6033. [PubMed: 16024602]
26. He H, Jazdzewski K, Li W, et al. The role of microRNA genes in papillary thyroid carcinoma. *Proc Natl Acad Sci U S A* 2005;102(52):19075–19080. [PubMed: 16365291]
27. He L, Thomson JM, Hemann MT, et al. A microRNA polycistron as a potential human oncogene. *Nature* 2005;435(7043):828–833. [PubMed: 15944707]
28. Wei JS, Song YK, Durinck S, et al. The MYCN oncogene is a direct target of miR-34a. *Oncogene* 2008;27(39):5204–13. [PubMed: 18504438]
29. Ota A, Tagawa H, Karnan S, et al. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. *Cancer Res* 2004;64(9):3087–3095. [PubMed: 15126345]
30. Akao Y, Nakagawa Y, Naoe T. let-7 microRNA functions as a potential growth suppressor in human colon cancer cells. *Biol Pharm Bull* 2006;29(5):903–906. [PubMed: 16651716]
31. Kluiver J, Haralambieva E, de Jong D, et al. Lack of BIC and microRNA miR-155 expression in primary cases of Burkitt lymphoma. *Genes Chromosomes Cancer* 2006;45(2):147–153. [PubMed: 16235244]
32. Lu J, Getz G, Miska EA, et al. MicroRNA expression profiles classify human cancers. *Nature* 2005;435(7043):834–838. [PubMed: 15944708]
33. Li J, Smyth P, Flavin R, et al. Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. *BMC Biotechnol* 2007;7:36. [PubMed: 17603869]
34. Brodeur, GM.; Maris, JM. Neuroblastoma. In: Pizzo, PA.; Poplack, DG., editors. *Principles and practice of pediatric oncology*. Lippincott-Raven; Philadelphia: 2002. p. 895-937.
35. Ohira M, Oba S, Nakamura Y, et al. Expression profiling using a tumor-specific cDNA microarray predicts the prognosis of intermediate risk neuroblastomas. *Cancer Cell* 2005;7(4):337–350. [PubMed: 15837623]
36. Schramm A, Schulte JH, Klein-Hitpass L, et al. Prediction of clinical outcome and biological characterization of neuroblastoma by expression profiling. *Oncogene* 2005;24(53):7902–7912. [PubMed: 16103881]
37. Asgharzadeh S, Pique-Regi R, Sposto R, et al. Prognostic significance of gene expression profiles of metastatic neuroblastomas lacking MYCN gene amplification. *J Natl Cancer Inst* 2006;98(17):1193–1203. [PubMed: 16954472]
38. Oberthuer A, Berthold F, Warnat P, et al. Customized oligonucleotide microarray gene expression-based classification of neuroblastoma patients outperforms current clinical risk stratification. *J Clin Oncol* 2006;24(31):5070–5078. [PubMed: 17075126]
39. Mullighan CG, Goorha S, Radtke I, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 2007;446(7137):758–764. [PubMed: 17344859]
40. Schwab M, Alitalo K, Klempnauer KH, et al. Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature* 1983;305(5931):245–248. [PubMed: 6888561]
41. Attiyeh EF, London WB, Mosse YP, et al. Chromosome 1p and 11q deletions and outcome in neuroblastoma. *N Engl J Med* 2005;353(21):2243–2253. [PubMed: 16306521]
42. Guo C, White PS, Weiss MJ, et al. Allelic deletion at 11q23 is common in MYCN single copy neuroblastomas. *Oncogene* 1999;18(35):4948–4957. [PubMed: 10490829]
43. Bilke S, Chen QR, Wei JS, Khan J. Whole Chromosome Alterations Predict Survival in High Risk Neuroblastoma without MYCN Amplification. *Clin Cancer Res*. 2008In Press

44. Wallace J, Zhou Y, Usmani GN, et al. BARCODE-ALL: accelerated and cost-effective genetic risk stratification in acute leukemia using spectrally addressable liquid bead microarrays. *Leukemia* 2003;17(7):1404–1410. [PubMed: 12835731]
45. <http://www3.appliedbiosystems.com/index.htm>
46. Chen QR, Vansant G, Oades K, et al. Diagnosis of the small round blue cell tumors using multiplex polymerase chain reaction. *J Mol Diagn* 2007;9(1):80–88. [PubMed: 17251339]
47. <http://www.mesoscale.com/>
48. Paweletz CP, Charboneau L, Bichsel VE, et al. Reverse phase protein microarrays which capture disease progression show activation of prosurvival pathways at the cancer invasion front. *Oncogene* 2001;20(16):1981–1989. [PubMed: 11360182]
49. Petricoin EF 3rd, Espina V, Araujo RP, et al. Phosphoprotein pathway mapping: Akt/mammalian target of rapamycin activation is negatively associated with childhood rhabdomyosarcoma survival. *Cancer Res* 2007;67(7):3431–3440. [PubMed: 17409454]
50. Yu LR, Conrads TP, Uo T, Issaq HJ, Morrison RS, Veenstra TD. Evaluation of the acid-cleavable isotope-coded affinity tag reagents: application to camptothecin-treated cortical neurons. *J Proteome Res* 2004;3(3):469–477. [PubMed: 15253428]
51. <http://seer.cancer.gov/>
52. Fritz J, Cooper EB, Gaudet S, Sorger PK, Manalis SR. Electronic detection of DNA by its intrinsic molecular charge. *Proc Natl Acad Sci U S A* 2002;99(22):14142–14146. [PubMed: 12386345]
53. Pandana H, Aschenbach KH, Lenski DR, Fuhrer MS, Khan J, Gomez RD. A Versatile Biomolecular Charge-Based Sensor Using Oxide-Gated Carbon Nanotube Transistor Arrays. *IEEE Sensors Journal* 2008;8(6):655–660.
54. Zhu H, Bilgin M, Snyder M. Proteomics. *Annu Rev Biochem* 2003;72:783–812. [PubMed: 14527327]
55. <http://www.annualreviews.org>

Figure 1. Diagnosis of cancers using RNA expression profiles

A, Diagnosis of SRBCTs using mRNA expression profiling and artificial neural networks (ANNs). Yellow, Ewing's sarcoma; Green, neuroblastoma; Red, rhabdomyosarcoma; Blue, Burkett's lymphoma; Black, non-SRBCT test samples. **a**, Hierarchical clustering of 88 SRBCT samples using the 96-clone profiles. *, genes that have not been reported to be associated with these cancers. **b**, The training samples are shown in a multidimensional scaling analysis using the top 96 ANN-ranked cDNA clones. The samples cluster closely according to the 4 different cancer categories. **c**, Classification and diagnosis of the samples. A sample is classified to a cancer category according to its highest committee vote (average of all ANN outputs) and placed in the corresponding plot. The distance between its committee vote and the ideal vote

(for example, for EWS, it is $EWS = 1$, $RMS = NB = BL = 0$) for each sample is plotted for that diagnostic category. Vertical dashed lines represent the boundary where 95% of the training samples are included. Training samples are represented as squares and test samples as triangles. Non-SRBCT samples are colored black. All SRBCT samples, including the 20 tests, were correctly classified. The diagnosis of all 5 non-SRBCT test samples was rejected since they lie outside the dashed lines. Three of the SRBCT samples (EWS-T13, TEST-10 and TEST-20) though correctly classified could not be confidently diagnosed. Adapted by permission from Macmillan Publishers Ltd: *Nature Medicine*, [16], copyright 2001.

B, Hierarchical clustering of tumors using miRNA expression profiles. **a**, miRNA profiles of 218 samples from several different tissues were clustered. Samples are in columns, miRNAs in rows. EP, samples of epithelial origin; GI, samples from the gastrointestinal tract. **b**, Clustering of 73 bone marrow samples from patients with acute lymphoblastic leukemia (ALL). Colored bars indicate the different ALL subtypes. **c**, Comparison of miRNA and mRNA profiles. For 89 epithelial samples from **a** that had mRNA expression data, hierarchical clustering was performed. Samples of GI origin are shown in blue. GI-derived samples largely cluster together when using miRNA expression profiles, but not when mRNA expression profiles are used. Bldr, bladder; Brst, breast; Fcc, follicular lymphoma; Kid, kidney; Lvr, liver; Mela, melanoma; Meso, mesothelioma; Pan, pancreas; Prost, prostate; Stom, stomach; Ut, uterus; AML: acute myelogenous leukaemia; BALL, B-cell ALL; LBL, diffuse large-B cell lymphoma; MF, mycosis fungoides; MLL, mixed lineage leukaemia; TALL, T-cell ALL; Hyper 47-50, hyperdiploid with 47-50 chromosomes; Hyper.50, hyperdiploid with over 50 chromosomes; Normp, normal ploidy. Adapted by permission from Macmillan Publishers Ltd: *Nature*, [32], copyright 2005.

Figure 2. Prognostic prediction in neuroblastoma patients using DNA microarray

A, Prediction by mRNA profiles. Identification of a 19-gene signature that can predict the outcome of neuroblastoma patients. **a**, Prognostic results from artificial neural networks (ANN) for patients using the 19-gene signature, where alive=0 and deceased=1. ANNs were trained using 35 training samples, and then were used to predict the outcome of 21 independent test samples. Horizontal dotted line divides the test (above the line) from the training samples. **b** and **c**, Kaplan-Meier curves for survival probability of the all patients using the 19 gene signature (**b**) or current COG risk-stratification (**c**). Prognostic prediction using the 19 gene signature is favorable over the current COG risk-stratification. **d**, A Kaplan-Meier curve for

13 high-risk patients without *MYCN* amplification were derived from the ANN prediction using the 19 gene signature. Figures are taken from Wei *et. al.* [17] with permission.

B, Prediction by DNA profiles. Whole chromosomal number changes (WCC) predict outcome in neuroblastoma patients in Kaplan-Meier survival analyses. The survival probabilities of patients with $WCC \geq 2$ or $WCC < 2$ are compared in all samples with survival information (**a**, $n=51$). The patients classified using the WCC status had significantly different survival probabilities ($P=0.01$). **b** and **c**, The patients diagnosed before or in 1998 (**b**, $n=23$), and after 1998 (**c**, $n=27$). **d**, The patients diagnosed after 1998 with age > 500 days ($n=15$). Figures are taken from Bilke *et. al.* [43] with permission.

Figure 3. Bead-based detection technology

A, micro plastic beads are color-coded recognizable by a red laser beam. **B**, Probes, nucleic acids or antibodies, are conjugated to specific beads. **C**, Beads with different probes are mixed and hybridized with samples labeled with a fluorescent reporter that can be detected by a green laser beam. **D**, Flow cytometry is performed with beads, where the molecules are identified by the color code of the beads and quantified by the fluorescent reporter. **E**, Expression profiles can be obtained for a hundred of molecules in one test tube.

Figure 4. Multiplexed RT-PCR assay

A, A schematic diagram for multiplexed RT-PCR assay. **a**, RNA samples are first reverse transcribed and partially amplified using a pair of chimeric primers, and then further amplified using a pair of universal primers. **b**, PCR products are separated using fluorescence capillary electrophoresis. A gene is identified by the size of its amplicon, and its expression level is represented by the peak area in a chromatogram. **c**, Four representative chromatograms for four different categories of tumor samples from one multiplex RT-PCR assay.

B, The three-dimensional plot clearly demonstrated that 31 top-ranked genes can accurately classify 4 different categories of SRBCTs in a multidimensional scaling analysis. Blue, lymphoma; yellow, Ewing's sarcoma; red, rhabdomyosarcoma; and green, neuroblastoma.

C, Hierarchical clustering of all 96 SRBCT samples using 31 top-ranked genes after log2 transformation and z-scoring across all samples.

Figures are taken from Chen *et al.* [46], *J. Mol. Diagn.*, 2007, 9:80-88 with permission from the American Society for Investigative Pathology and the Association for Molecular Pathology.

Figure 5. Schematic illustration of MSD, ICAT, and label-free electronic hybridization sensing technologies

A, Meso scale discovery (MSD) technology is a multiplex assay where the levels of multiple proteins can be determined using electrochemiluminescence. This immunoassay uses capture antibodies coated on the surface of plates to capture proteins from whole cell lysates, which can be then detected by an antibody labeled with an electrochemiluminescent tag [47].

B, Isotope-coded affinity tag (ICAT) uses heavy [d(8)] and light [d(0)] labeling reagents to tag protein samples of two states (*e.g.* diseased and normal). The samples are combined, digested with trypsin into peptides, purified with the affinity tags, separated with multidimensional chromatography, and then analyzed by MS and MS/MS for quantification and identification respectively. The relative abundances of labeled peptides are determined by comparison of peak intensities between the light and heavy forms of the peptides, which are separated by 8 Da. Reprinted with permission from the *Annual Review of Biochemistry*, Volume 72 ©2003 by Annual Reviews [54,55].

C, Electronic hybridization sensing using field effect transistors (FET). **a,** principals of a DNA sensing device using FET. In the absence of charge at the gate, no current will flow at any voltage so that the transconductance curve traces a horizontal line (left). Single-stranded DNA oligonucleotides (probe DNA) immobilized at the gate cause a net negative charge from their phosphate backbone. The resulting transconductance curve will exhibit slight but measurable departure from horizontal linearity (middle). If the single-stranded DNA probes were hybridized with a complementary sequence, a net doubling of charge will occur at the gate. This will drive the transistor into forward biased full operation, and will produce large

amplitude and highly nonlinear transconductance curves (right). **b**, Transconductance curves of a FET device (similar to the device on the right panel) demonstrate its ability to distinguish specific DNA hybridization from non-specific DNA interaction. The green arrow represents the change of conductance of the FET from the empty probe state to specific DNA-hybridization state. Empty probe, single-strand DNA probe is immobilized; non-specific, incubated with DNA of unmatched sequence; Specific, hybridized with complementary DNA. **c**, an example of a carbon nanotube (CNT) transistor array on a microchip. CNTs bridge the gap between source (S) and drain (D) electrodes. G is the gate electrode in contact with the hybridization solution.