DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
76

MUTATION DETECTION BY PRIMER EXTENSION ON OLIGONUCLEOTIDE MICROARRAYS

NEEME TÕNISSON

TARTU 2002

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
76

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
76

MUTATION DETECTION BY PRIMER EXTENSION ON OLIGONUCLEOTIDE MICROARRAYS

NEEME TÕNISSON

TARTU UNIVERSITY
PRESS

Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

Dissertation is accepted for commencement of the degree of Doctor of Philosophy (in Molecular Diagnostics) on December 4th, 2002 by the council of the Institute of Molecular and Cell Biology, University of Tartu

Opponent: Prof. Dr. Jörg Hoheisel, Division of Functional Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany

Commencement: January 10th, 2003

Publication of this dissertation is granted by the University of Tartu

© Neeme Tõnisson, 2002

Tartu Ülikooli Kirjastuse trükikoda
Tiigi 78, Tartu 50410
Tellimus nr. 880

To Mailis, Siim and Matis

To my grandfather

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

I Kurg A, **Tõnisson N**, Tollett J, Georgiou I, Shumaker J, Metspalu A. Arrayed Primer Extension: Solid phase four-color DNA resequencing and mutation detection technology. *Genetic Testing 2000; 4(1):1–7*

II **Tõnisson N**, Kurg A, Lõhmussaar E, Metspalu A. Arrayed primer extension on the DNA chip — method and applications. Microarray Biochip Technologies *(M. Schena, ed.) Biotechniques Books. Eaton Publishing, Natick MA 2000: 247–3*

III **Tõnisson N**, Kurg A, Kaasik K, Lõhmussaar E, Metspalu A. Unravelling genetic data by arrayed primer extension. *Clinical Chemistry and Laboratory Medicine 2000; 38(2):165–70*

IV **Tõnisson N** , Zernant J, Kurg A, Pavel H, Slavin G, Roomere H, Meiel A, Hainaut P and Metspalu A. Evaluating the arrayed primer extension resequencing assay of TP53 tumor suppressor gene. *Proc Natl Acad Sci USA 2002; 99(8):5503–8*

V Gemignani F, Perra C, Landi S, Canzian F, Kurg A, **Tõnisson N**, Galanello R, Cao A, Metspalu A, Romeo G. Reliable detection of ß-thalassemia and G6PD mutations by a DNA microarray. *Clinical Chemistry 2002; 48(11):2051–4*

LIST OF SYMBOLS AND ABBREVIATIONS

APEX	– Arrayed Primer Extension
ARMS	– Amplification Refractory Mutation System
ASPEX	– Allele-Specific Primer Extension
ASO	– Allele-Specific Oligonucleotide
bp	– basepair
CCD	– Charge Coupled Device
CCM	– Chemical Cleavage of Mismatch
DASH	– Dynamic Allele Specific Hybridisation
ddNTP	– dideoxyribonucleoside triphosphate
DGGE	– Denaturing Gradient Gel Electrophoresis
DMT	– dimethoxytrityl
dNTP	– deoxyribonucleoside triphosphate
EMC	– Enzyme Mismatch Cleavage
FP	– Fluorescence Polarisation
FRET	– Fluorescence Resonance Energy Transfer
G6PD	– Glucose 6 phosphate dehydrogenase
GBA	– Genetic Bit Analysis
MALDI TOF MS	– Matrix Assisted Laser Desorption / Ionisation Time-of-Flight Mass Spectrometry
PAH	– Phenylalanine hydroxylase
PCR	– Polymerase Chain Reaction
SBE	– Single Base Extension
SBH	– Sequencing By Hybridisation
SNP	– Single Nucleotide Polymorphism
SSCP	– Single Strand Conformational Polymorphism
T_m	– melting temperature, a temperature at which half the DNA strands are single-stranded and half are double-stranded
TIRF	– Total Internal Reflection Fluorescence
TMACl	– tetramethyl ammonium chloride
TTGE	– Temporal Temperature Gradient Electrophoresis
UNG	– Uracil N-Glycosylase

1. INTRODUCTION

All known diseases have presumably both genetic and environmental factors important in their pathogenesis. The role of genetic factors is different in the various types of disorders. While monogenic diseases depend essentially on the variations of one or a few disease genes, genetic background of multifactorial and infectious diseases has incomplete penetrance and the clinical phenotype depends largely on environmental factors.

Medical genetics including the diagnostics and screening as well as basic research need urgently new mutation detection technologies enabling more efficiency and larger analysis scale. Moreover, the choice of a method to study the genetic variability may be equally important with the study design, as the method chosen will largely affect the scale of study and the quality and type of information obtained. The classical mutation detection methods have significant limitations in their throughput and type of information obtained from a single step assay. Direct methods such as dideoxy sequencing, restriction endonuclease digestion and allele-specific PCR have a limited capacity of parallel analysis and, especially automated dideoxy sequencing, also a limited sensitivity (Ahrendt et al., 1999). On the other hand, indirect methods (e.g. SSCP, DGGE, TTGE) give only yes or no type information about the presence and average location of a mutation and second step analysis with a direct method is needed to identify the mutation.

The current study is aimed on development of integrated technology for mutation detection utilising Arrayed Primer Extension (APEX) on DNA chips. The technology consists of glass slide chemistry for immobilisation of the APEX oligonucleotides, template preparation and APEX reaction, four-colour fluorescence imaging system for analysis of DNA chips and, finally, custom tailored genotyping software to analyse the different APEX assays. The technology allows for variable assay scale as well as for high throughput sample processing and has been successfully applied for detection of known mutations in a number of disease genes including β-globin gene (Kurg et al., 2000, Gemignani et al., 2002), BRCA1 gene (Tõnisson et al., 2000a), and detection of unknown mutations in TP53 gene (Tõnisson et al., 2002). The quality of information enables to use APEX in both research and diagnostic screening studies. Last, but not least, the technology has been filed in two international patent applications (Metspalu et al., 2000, Pavel et al., 2000) and has been set into routine use by a spin-off company of Estonian Biocentre, Asper Biotech Ltd. (Tartu, Estonia).

2. REVIEW OF LITERATURE

2.1. The scope and perspectives of genetic testing

Advances in the genetic discovery and healthcare entering to the post-genomic era have continuously increased the demand for genetic testing. With the increased knowledge, genetic information will have to be used throughout the lives of individuals from prenatal and newborn genetic screening to later testing genetic risks for development of complex disorders. Monitoring for individual drug response may also soon become a standard practise (Peltonen and McKusick, 2001). A comprehensive database of genes and mutations associated with approximately 1500 clinical disorders is available online (http://www3.ncbi.nlm.nih.gov/Omim/) and the list is not complete yet.

The complex of genetic and environmental factors have to be continuously studied for full understanding of the basis of diseases. Even "simple" Mendelian disorders have phenotypes apparently inherited as complex traits (Dipple and McCabe, 2000). For example, the classical phenotype of a commonly screened disorder, phenylketonuria, includes high levels of blood phenylalanine and mental retardation of untreated individuals, whereas patients with mild hyperalaninaemia may not require dietary intervention. The levels of phenylalaninaemia and prognosis, if untreated, are only partially correlated to mutations in phenylalanine hydroxylase (PAH) gene. Studies have shown that the increased degradation of the mutant PAH protein may be even more responsible for the increased level of phenylalanine than the decreased catalytic activity of PAH and molecular pathways responsible for protein degradation contribute a lot to the individual differences with people having identical PAH genotype (Scriver and Waters, 1999). Phenotype modifiers have been identified for a number of disorders including the hereditary haemoglobinopathies (e.g. concomitant expression of haemoglobin F for homozygous sickle cell disease). The complex phenotypes fit well with concepts that metabolic pathways are rarely controlled by single rate-limiting steps and more than one step have influence on pathway flux (Dipple and McCabe, 2000).

Identification of genes behind complex diseases like asthma, cancer, hypertension, etc. is a much bigger challenge. The pedigrees do not reveal Mendelian inheritrance and environmental factors such as life style contribute significantly. Risk factors in a family may differ from the relative risk factors at population level. So far, too many loci and not enough gene mutations are implicated in complex diseases (Peltonen and McKusick, 2001).

Association studies and model organisms are the genetic discovery platform for polygenic disorders. The best option to validate associations between the genetic markers and disease are extensive epidemiological studies with samples and data from different populations (Peltonen and McKusick, 2001).

Throughput of genetic testing is not the only limiting factor for performing large-scale studies. In addition, the reliability and quality of results has sometimes been far beyond expected. A well-known study revealed that only 62.5% of laboratories were able to type correctly a limited number of samples for the cystic fibrosis mutations for which they routinely screened and only two out of the nine samples were correctly typed in all 40 participating laboratories (Cuppens and Cassiman, 1995).

Genetic testing is a process involving a number of ethical (the informed consent, problems involving relatives, confidentiality) and laboratory (reliability, costs, positive predictive value) issues. Classical and standard criteria of World Health Organization for introducing a screening test include benefit for the person screened and a treatment being available (Wilson and Jungner, 1968). Two most common newborn screening tests for phenylketonuria and hypothyroidism meet the criteria extremely well. This applies also for susceptibility testing for common disorders (e.g. Familial Adenomatous Polyposis). On the other hand, situation in real life is changing rapidly. Newer tests may only partially fit the classical model and yet be ethically and economically justified if targeted right. For example, tests such as carrier screening of haemoglobinopathies are mostly performed in at-risk populations. For other monogenic disorders like cystic fibrosis, fragile X syndrome, the optimal target group consists of families with affected children.

2.1.1. Background of the genes studied in model assays

- Carrier screening for the thalassemias

Thalassemias are an excellent example of how the knowledge applied to molecular diagnosis, carrier identification and prenatal diagnosis, have resulted in dramatic reduction of the incidence of the homozygous state in several at-risk populations (Cao et al., 1997, Weatherall and Clegg, 2001).

Thalassemias are globally the most common monogenic disorders and also the group of monogenic disorders being first well characterised in the molecular level. The disorders result from reduced or absent production of beta-globin chains which form the haemoglobin tetramere. The highest frequency is observed in the Mediterranean area, the Middle East, India and Far East Asia due to long-term heterozygote selection against malaria, which is endemic in these regions. However, because of population flows, the disorder also occurs in Continental Europe, North and South America, and Australia (Cao et al., 1997). The carrier rates of β-thalassemia mutations range from 1 to 20% in these regions (Weatherall and Clegg, 2001).

The clinical manifestance is extremely variable, but may be still divided into 3 major conditions: carrier state, thalassemia intermedia and thalassemia major. Thalassemia major is a result of homozygosity for severe mutations resulting in markedly low production of beta globin chains and great excess of unassembled alpha globin chains. Without treatment, these patients die from anaemia-related complications in early years of life. Beta-thalassemia carrier state is clinically asymptomatic. Thalassemia intermedia is a large number of conditions ranging from thalassemia carrier state to thalassemia major and is clinically characterised by mild to moderate microcytotic anaemia, splenomegaly, bone manifestations and predisposition to develop gall stones and leg ulcers (Cao et al., 1997).

Beta globin genes are arranged in a cluster on a short arm of chromosome 11. Vast majority of the beta thalassemias are produced by either a single nucleotide substitution or an oligonucleotide addition or deletion that affects the coding region or critical areas for the function of beta globin gene. Among point mutations, about half completely inactivate the beta globin gene. Mutations in the initiation codon ATG or mutations disrupting the splicing sites also result in beta degrees thalassemia. Mutations in the flanking regions of the splice sites affect the efficiency of splicing and produce variable forms of beta thalassemia. More than 200 different molecular defects have been identified in the beta globin gene. Despite the molecular heterogeneity, each at-risk population has a limited set of prevalent mutations, e.g. 8 to 10 in Mediterranean and Asian populations, facilitating the molecular screening and diagnosis (Cao et al., 1997, Weatherall and Clegg, 2001).

The methods commonly used for molecular diagnosis and carrier screening comprise of PCR plus reverse dot-blot analysis, allele specific PCR, restriction endonuclease digestion or denaturing gradient gel electrophoresis followed by direct sequencing if a mutation has not been found with simpler methods (Cao et al., 1997, Martinez di Montemuros et al., 1997, Rosatelli et al., 1992).

Since late 1970's programs of beta thalassemia population screening have been introduced in at-risk populations of Mediterranean basin including Sardinians, Continental Italians, Greeks and Cypriots. The programs include population screening of adults at childbearing age, genetic counselling and prenatal diagnosis. Screening is usually performed at voluntary basis with non-directive counselling. An exception is Cyprus, where premarital testing for beta-thalassemia mutations is mandatory (Revel, 1995). These approaches have resulted in 80 to 100% decline of the birth of affected homozygotes (Weatherall and Clegg, 2001).

- TP53 tumour suppressor gene

TP53 gene is by far the most studied among the tumour suppressor genes. The gene was discovered on 1979 (Linzer and Levine, 1979) and initially described as a presumable oncogene with transforming capability (Eliyahu et al., 1984, Jenkins et al., 1984). The next experiments indicated that wild type TP53 is a suppressor of cell transformation *in vitro* (Eliyahu et al., 1989, Finlay et al., 1989). The normal function of TP53 is regulation of cellular response to various stress factors (Vogelstein et al., 2000, Woods and Vousden, 2001).

The early connection to loss of TP53 function with human cancer was discovered by genetic analysis of colorectal cancer samples. A high rate of heterozygous deletions of the chromosome 17 short arm, where the TP53 gene is located, was first observed (Vogelstein et al., 1988). Sequencing of the remaining TP53 allele showed that it is often affected by point mutations (Baker et al., 1989). Similar genetic changes were detected in the case of lung cancer (Takahashi et al., 1989). Later studies have shown that prevalence of TP53 mutations varies among tumour types, ranging from 0% to 60% or up to 80% for certain histological subtypes. TP53 alterations may be both early and late steps in multistage tumorigenesis process (Greenblatt et al., 1994). The growing evidence shows that specific mutations in the TP53 gene can represent an important factor for the prognosis of cancer and for the response to various types of cytotoxic therapy (Thorlacius et al., 1993, Aas et al., 1996, Wen et al., 1999, Cabelguenne et al., 2000).

Alterations in TP53 gene do not only occur as somatic mutations in human cancers, but also as germline mutations in some cancer-prone families, clinically described as Li-Fraumeni syndrome. This dominantly inherited syndrome presents as a high incidence of a broad spectrum of tumours including osteosarcoma, soft tissue sarcoma, breast cancer, leukaemia or other types of cancer appearing at a very early age. The penetrance of the syndrome is very high. Statistical analysis for these individuals predicts a 50% risk of having a tumour before the age of 30 when only 1% of the general population has cancer. The risk will be increased to 90% for the age of 70. Development of a tumour is in strict correlation with transmission of the mutant allele (Malkin et al., 1990, Malkin, 1994).

The majority of all the alterations found in TP53 gene are point mutations, mainly missense type (Hainaut et al., 1998, Soussi et al., 2000). TP53 mutations may be classified according to the site of mutation and its phenotype (Michalovitz et al., 1991):

1) null mutations with inactive TP53 that does not interfere in transformation;
2) dominant negative mutations with inactive TP53 that is able to interfere with TP53 expressed from the wild-type allele;
3) positive dominant gain-of-function mutations; the mutant TP53 will gain oncogenic activity.

A comprehensive database of TP53 mutations with close to 18,000 entries (Sept. 2002) is maintained by International Agency for Research on Cancer (IARC, Lyon, France, http://www.iarc.fr/p53/) (Hernandez-Boussard et al., 1999, Olivier et al., 2002b). The aim of the database is to provide a tool to classify, sort, compare and analyse the TP53 alterations in order to generate hypotheses about natural course of cancer. 95% of all known mutations fall within DNA binding domain (Hainaut et al., 1998). However, the full coding region is rarely studied and the database may still be incomplete. Analysis of the database reveals mutation fingerprints specific for some exogenous carcinogens including tobacco smoke and tumour-specific mutation spectra (Hussain and Harris, 1999, Hollstein et al., 1999, Olivier et al., 2002b, Hainaut et al., 2001, Hainaut and Pfeifer, 2001).

The search for somatic mutations in the TP53 tumor suppressor gene is usually performed with a combination of two different methods due to several reasons including the labour requirements and sensitivity. At first, indirect analysis is performed by SSCP (Mazars et al., 1991), DGGE (Hamelin et al., 1993), TTGE (Taniere et al., 2001) or DHPLC (Gross et al., 2001). If a mutation is found, a second step is conducted for identification of the mutation by direct sequencing. Massive sequencing of cloned TP53 from tumour samples has been used in a few studies (Baker et al., 1990, Ahrendt et al., 1999). This is probably the utmost laborious technique known and suitable only as a reference method if sensitivity and specificity of other methods is being evaluated. An indirect, but not always correct measure of TP53 mutations is a positive immunostaining due to the fact that mutations in TP53 often stabilise the protein. However, not all the mutations have the stabilizing effect and the level of wild type TP53 may also be upregulated in response to DNA damage or external factors including hypoxia (Wallace-Brodeur and Lowe, 1999).

2.2. Mutation detection methods

2.2.1. Introduction to methods used in screening and clinical diagnostics

Large numbers of different methods and their modifications have been developed for mutation detection in research and diagnostic applications. The most general classification is dividing the methods into indirect and direct methods, according to the type of genetic information obtained (Table 1). Alternative classifications include types like gel-based and non-gel based according to the type of instrumentation needed. An important factor is the capacity to characte-

rise only pre-known or also unknown sequence variations. In addition, the different methods are usually compared by their quality of results including sensitivity and specificity.

The advantages of indirect methods are sensitivity compared to dideoxy sequencing and speed with the common laboratory hardware used. No method mentioned here fulfils completely the current need for throughput and robustness. Precise sequence determination is practically always the ultimate goal. Therefore, most of the modern methods are developed as single step assays for direct identification of known or all possible mutations.

Table 1. Examples of classical mutation detection methods (Landegren, 1996, Day et al., 1995, Cotton, 1997, Underhill et al., 1997)

Method	Classification, Mutations detected	Description	Critical factors
SSCP, SSCP-HA (Single Strand Conformational Polymorphism, may be combined with Heteroduplex Analysis)	Indirect, Known / Unknown	DNA mobility in non-denaturing gel electrophoresis is affected by changes in its tertiary structure. Tertiary structure depends on the primary sequence of amplified PCR fragment. Duplexes of fully complementary and mismatched strands may be formed if wild-type and mutant alleles are present. These heteroduplexes have altered mobility in non-denaturing electrophoresis conditions.	Length of PCR product, constant cooling
DGGE (Denaturing Gradient Gel Electrophoresis)	Indirect, Known / Unknown	The method is based on partial melting of double stranded DNA in increasing gradient of chemical denaturant and heat. Partial melting results in mobility shift during gel electrophoresis.	Amplification probe design, product length, denaturant range
TTGE (Temporal Temperature Gradient Electrophoresis)	Indirect, Known / Unknown	The method is similar to DGGE, but a temperature gradient is used instead of gradient of chemical denaturants.	Amplification probe design, product length, temperature and denaturant range

Method	Classification, Mutations detected	Description	Critical factors
CCM (Chemical Cleavage of Mismatch) **EMC** (Enzymatic Mismatch Cleavage)	Indirect, Known / Unknown	Based on chemical or enzymatic cleavage of heteroduplex, detected by molecule separation on denaturing gel electrophoresis.	Efficient formation of heteroduplexes
DHPLC (Denaturing High-Pressure Liquid Chromatography)	Indirect, Known / Unknown	Based on partial melting of heteroduplexes formed during PCR on high pressure liquid chromatography at partially denaturing conditions	Temperature conditions
Dideoxy sequencing	Direct, Known / Unknown	Chain termination with dideoxy nucleotides during enzymatic synthesis by DNA polymerase. All possible lengths are formed in the same reaction. The fragments are separated by denaturing gel electrophoresis.	Design of sequencing primers, length of analysed sequence, percentage of mutated DNA
ARMS (Amplification Refractory Mutation System)	Direct, Known	PCR with primers specific for wild-type and mutant alleles	PCR primer design, amplification conditions
Restriction endonuclease digestion	Direct, Known or, alternatively, Indirect, Unknown, depending on the set-up of the assay	Mutations will generate or destroy existing restriction sites. Restriction sites may be also introduced artificially during PCR	Presence of restriction sites, available set of restriction endonucleases
Reverse dot blot	Direct, Known	Hybridisation of immobilised oligonucleotide probes on with labelled sample DNA	Probe design, hybridisation conditions

2.2.2. DNA microarray and other novel methods applied to molecular diagnostics

Comprehensive reviews describing most of the current genotyping and mutation detection technologies have been published by different authors (Syvänen, 2001, Kirk et al., 2002).

The conventional methods for screening the sequence variations are largely based on gel electrophoresis. As an alternative to gel-based sequencing, large scale hybridisation with synthetic oligonucleotides was proposed on late 80's by several groups, a strategy which later became known as sequencing by hybridisation (SBH). The strategy emerged mainly in two directions: first, hybridisation of arrays of immobilised large DNA segments with labelled oligonucleotides (format I, similar to dot blot) (Drmanac et al., 1989, Strezoska et al., 1991, Drmanac et al., 1998) and second, hybridisation of arrayed oligonucleotides on solid support with a labelled target sequence (format II, similar to reverse dot blot) (Fodor et al., 1991, Southern et al., 1992). The second format was eventually much more successful as it does not need the large numbers of hybridisation experiments to gain the sequence data. As the probe and target are bound to different phases in the two formats, it's essential to define the probe and target for the current thesis. Probe is the known synthetic sequence, which is used to assess the unknown, target sequence (Phimister, 1999). However, quite often, different authors define the target just as the arrayed sequence and the probe as the sequence in solution phase.

So far, microarrays are the only tools, which allow researchers to handle sequences in a highly parallel format, suitable for analysis in genomic scale (Cho et al., 1999, Patil et al., 2001, Dawson et al., 2002).

- Applications of DNA arrays

The two main directions for use of DNA arrays are the sequence variation and comparative gene expression studies. Oligonucleotide arrays are used for both purposes (Southern et al., 1992, Chee et al., 1996, Hacia et al., 1996, Shumaker et al., 1996, Pastinen et al., 1997).

Oligonucleotide arrays can be used for targeting known (mutation-specific and resequencing arrays) as well as unknown sequences (universal DNA arrays). The analysis types can be classified as "gain-of-signal" and "loss-of-signal" (Hacia, 1999). Oligonucleotides, which hybridise well with the target sequence will raise a detectable signal for analysis. If the target sequence contains a variation, which disrupts hybridisation to the arrayed probes, the variation can be identified by a loss of detectable signal (a "footprint" on the microarray) (Head et al., 1997, Hacia, 1999). The size of footprint depends on

the type and length of variation being smaller with point mutations and larger with frameshift mutations such as insertions and deletions.

Although the DNA arrays were initially proposed for sequencing, their main field of use has moved to gene expression assays, which have been designed to look for absolute (Lockhart et al., 1996) or relative expression levels (Schena et al., 1996).

The most common type of differential gene expression assays are cDNA arrays, where amplified PCR products are bound to solid surface and simultaneously hybridised to mRNA-s from two different sources. Both mRNA samples are specifically labelled and relative abundance of an individual mRNA can be visualized by comparison of the ratio of labels (Schena et al., 1996, Shalon et al., 1996).

- Microarray supports, oligonucleotide *in situ* synthesis and printing

DNA arrays are devices displaying specific oligonucleotides or longer DNA fragments attached in discrete order onto activated solid surface. An ideal support material for DNA array should have certain technical and chemical properties such as mechanical and chemical stability, ability for chemical modification and derivatisation, absence of autofluorescence and low cost. A variety of solid supports (glass, plastics, nylon, polyacrylamide) have been used for DNA arrays. Glass slides are by far the most common supports for microarrays. because they are readily available and inexpensive. Glass gives also a non-porous, relatively homogenous chemical surface, which can be used for modification with silanization chemistry. As liquids do not penetrate glass, the target has direct access to the probes on surface and no time is needed for diffusion (Southern et al., 1999).

The arrays can be made with pre-synthesized and mechanically deposited oligonucleotides (Nikiforov et al., 1994a, Pastinen et al., 1997) and oligonucleotides synthesised *in situ* (Fodor et al., 1991, Southern et al., 1994).

Four approaches have been used for *in situ* synthesis:

1) photolithographic deprotection method with physical masks (Fodor et al., 1991) or electronically guided micromirrors as virtual masks (Singh-Gasson et al., 1999)
2) ink-jet delivery of nucleotide precursors to the surface (Hughes et al., 2001)
3) electrochemical deprotection on semiconductor chip (www.combimatrix.com)
4) *in situ* synthesis by physically locating the nucleotide precursors with different rectangular and circular masks (Southern et al., 1994) or channels (Southern et al., 1992, Maskos and Southern, 1993a) (Figure 1).

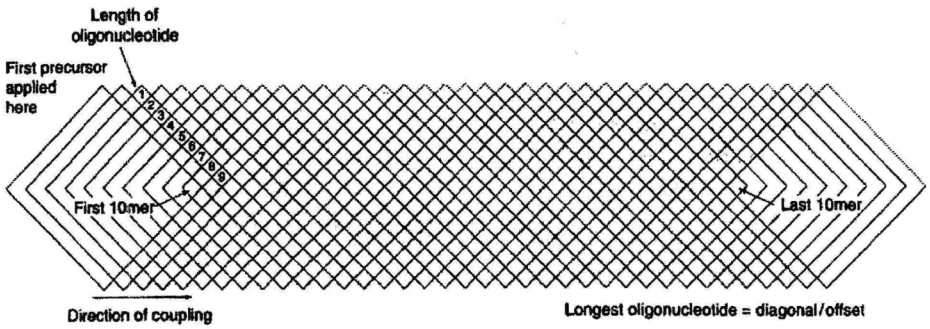


Figure 1. Making of scanning arrays with rectangular mask synthesis. A cell is created by pressing the seal of the mask against the substrate on which the synthesis takes place. After each synthesis cycle the mask is moved along the substrate by a pre-determined offset. Reprinted from Southern, E. M. et al. Arrays of complementary oligonucleotides for analysing the hybridisation behaviour of nucleic acids. (1994) *Nucleic Acids Res,* Vol.22, No.8, p.1369 by permission of Oxford University Press.

In situ synthesis has two main advantages compared to arrays with pre-synthesized oligonucleotides. First, the possibility to reach higher densities of different features per surface unit for large scanning arrays and universal DNA arrays with complete sets of oligonucleotides. The photolithographic synthesis introduced by Affymetrix (Santa Clara, CA, USA) allows to make arrays with 300,000 to > 1,000,000 different sequences per 1.28x1.28 cm area (Lipshutz et al., 1999). Second, while combined with a flexible assay design platform, e.g. micromirror-assisted synthesis, it also creates an excellent testing tool to evaluate different oligonucleotide sequences for large assays before mass production of arrays.

Mechanical deposition of pre-synthesised sequences has become more popular as it is suitable for both oligonucleotides and cDNA-s and because the arrayers are readily available from a range of companies (http://www.lab-on-a-chip.com/suppliers/maequip.html). Current arrayers can deposit more than 100,000 different spots per 25x75mm slide (http://lab-on-a-chip.com/suppliers/comptab.html), which is by far sufficient for most applications.

None of existing oligonucleotide synthesis chemistries provides 100% efficiency and each new synthesis carries risk for variations in the oligonucleotide quality. The efficiency depends on the type of nucleotide precursors and the nucleotides. For example, the photolithographic synthesis with 5'-(α-methyl-2-nitropiperonyl)oxycarbonyl(MeNPOC)-2'deoxynucleoside phosphoramidites has been shown to provide total of 92 to 98% efficiency after first 6 steps of synthesis and with more conventional 5'-(4,4'-dimethoxytrityl)(DMT) monomers a more uniform 98±1% efficiency per each synthesis step (McGall et al., 1997). The synthesis efficiency with MeNPOC phosphoramidites have been only at 88% level of more recently developed deoxynuc-

leoside phosphoramidites with 5'-2-(2-nitrophenyl)-propoxycarbonyl (NPPOC) group per each synthesis cycle (Beier and Hoheisel, 2000).

In situ synthesized oligonucleotide arrays have shown good performance in hybridisation reactions, but the control over the quality of oligonucleotides is difficult. Cleavable linkers, ellipsometry or interferometry have been experimentally used to assess the quality of oligonucleotides synthesized *in situ* (Southern et al., 1999). "Post synthesis" arraying allows the quality of oligonucleotides to be properly checked beforehand, making this approach preferable for mass production and diagnostics.

- Binding chemistry of oligonucleotide probes

The choice of chemical methods to bind DNA to different microarray supports is extremely wide. New methods and modifications are published weekly. The criteria for choosing the type of chemical binding include the support for hybridisation, specific, fast and stable immobilisation. Large-scale production of arrays and oligonucleotides needs also simple and reproducible methods to be used.

To save space, only methods used for arrays on glass will be mentioned here. On glass, the DNA probes are not bound directly to the glass surface silanol groups. Different active groups for DNA binding are generated beforehand by various silanes through universal coating process (Figure 2).

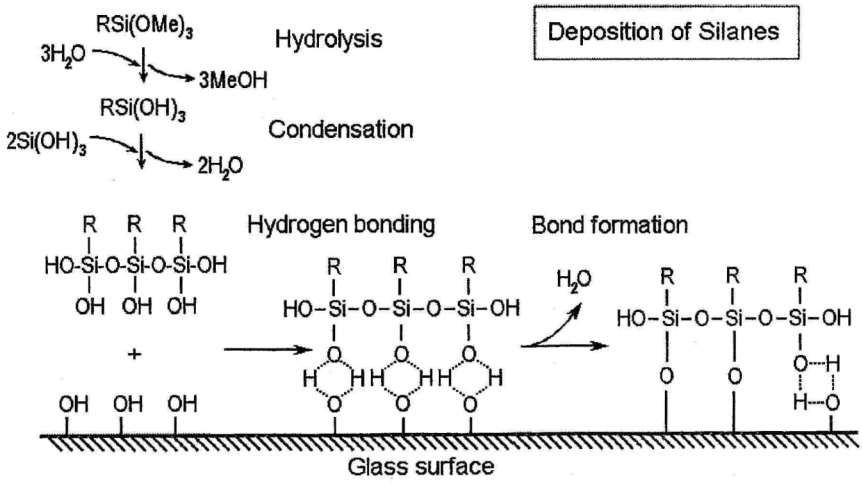


Figure 2. The universal steps of glass silanisation: hydrolysis, condensation, hydrogen bonding and formation of the covalently bound silane layer.

Two commonly used methods for microarray glass slide treatment utilise epoxy- (Maskos and Southern, 1992a, Lamture et al., 1994) and aminosilane (Fodor et al., 1991, Guo et al., 1994). Aminosilane allows for binding of unmodified DNA due to its positive charge and is together with polylysine coating the most used coating for cDNA arrays. The aminosilane can be made amino-reactive by further treatment with excess of phenylene diisothiocyanate linker (Guo et al., 1994). This thiocyanate coating has become a reference method for other methods of oligonucleotide binding and use on glass surface (Lindroos et al., 2002). Both epoxy and thiocyanate coating bind amino modified oligonucleotides (Lamture et al., 1994, Guo et al., 1994). Aminated DNA can also be immobilised on aldehyde- and carboxylated glass surface (Zammatteo et al., 2000).

The synthesis of probes with artificial chemical groups may have advantage over the groups naturally occurring in DNA to promote specific binding via the modified termini of probes. Mercaptosilane may be used to bind disulfide-modified oligonucleotides (Head et al., 1997). Benzaldehyde-modified oligonucleotides are shown to bind and hybridise well on semicarbazide-coated glass (Podyminogin et al., 2001). 5'terminal carboxylic acid functional groups have been introduced to bind probes on aminophenyl or aminopropyl silanes (Joos et al., 1997). Polyacrylamide-mediated binding of oligonucleotides has been reported (Fotin et al., 1998, Rehman et al., 1999).

Instead of silanising the glass surface, it is also possible to silanise the oligonucleotide probes and bind them directly onto unmodified glass (Kumar et al., 2000).

Oligonucleotide *in situ* synthesis has been built up from 3-glycidoxypropyltrimethoxysilane and aminopropyltriethoxysilane coated glass surface using oligoethyleneglycol linker (Maskos and Southern, 1992a) and photolabile DMT-hexaethyleneglycol-(2-cyanoethyl-N,N-diisopropyl) phosphoramidite (McGall et al., 1996, Fodor et al., 1991).

- Molecular interactions on microarrays

DNA arrays have enabled large-scale thermodynamic studies on the probe / target duplex formation, in a way not possible with standard solution-based methods (Maskos and Southern, 1992b, Maskos and Southern, 1993a, Shchepinov et al., 1997, Fotin et al., 1998).

The duplex stability is definitely affected by the base composition as A:T pairs have lower stability compared to G:C pairs. Oligonucleotides of same length may therefore have extreme differences in their T_m. The differences can be minimised and the overall duplex yield significantly increased under high concentration of chaotropic salts, such as 3-4M tetramethylammonium chloride (TMACl) (Maskos and Southern, 1993b). The sequence context is as important

for hybridisation as the base composition as the base stacking depends on the neighbouring base and affects the duplex stability (Maskos and Southern, 1993b, Southern et al., 1999).

Empirical correction parameters have been published to extrapolate the thermodynamic parameters obtained by oligonucleotide array experiments to the experiments performed in solution and *vice versa*. The free energy and T_m of arrayed oligonucleotides are significantly lower compared to free energy and T_m of same oligonucleotides in solution (Fotin et al., 1998) (Figure 3). The exact parameters may depend on the array chemistry used, although some evidence exists that a linker added to oligonucleotide terminus does not significantly affect the duplex stability (Fotin et al., 1998). Hybridisation and duplex stability are affected by mismatch bases, whereas internal mismatches the most significant and terminal mismatches have the lowest effect (Lipshutz et al., 1995). Adding a terminal universal base like 5-nitroindole to the oligonucleotide probe enhances the relatively weak destabilising effect of terminal mismatches (Fotin et al., 1998). The probe bases nearest to the surface are less accessible to the hybridisation than those furthest away if one end of the oligonucleotide is tethered to array surface. This may explain why mismatch at 13th base of 18-mer probes bound to epoxy modified glass surface had the lowest duplex yield from hybridisation with a fluorescently labelled 50 to 77-mer HER2/neu targets compared to the fully complementary probes (M.Cieplik, Univ.of Ulm, Germany, personal communications).

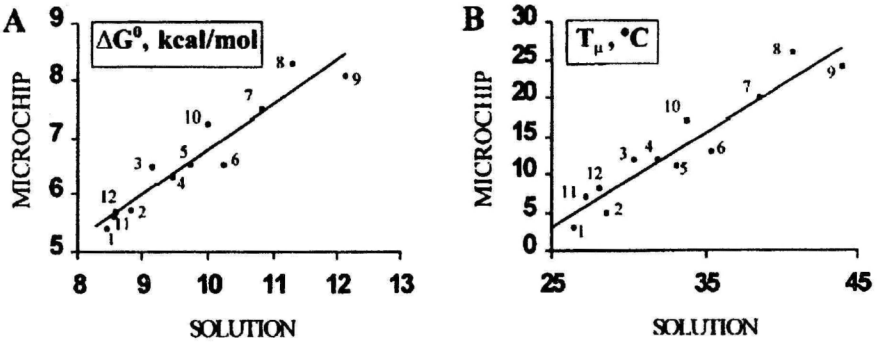


Figure 3. Thermodynamic properties of arrayed oligonucleotides. Both free energy and melting temperature values are lower for oligonucleotides on the microchip compared to the same oligonucleotides in solution. A. plots of free energies. B. Plots of melting temperature. Reprinted from Fotin, A. V. et al. Parellel thermodynamic analysis of duplexes on oligodeoxyribonucleotide microchips. (1998) *Nucleic Acids Res,* Vol.26, No.6, p.1519 by permission of Oxford University Press.

Duplex stability may not be the main factor affecting the yield. The yield is the result of forward reaction in the reversible hybridisation reaction. It is hypothesised that the duplex formation begins from formation of transient nucleation complex, from the interaction of very few base pairs and proceeds, base by base, through zippering process (Southern et al., 1999). At any time the process may be reversed by a mismatch base or a secondary structure. The stability of the nucleation complex and the intermediates of the zippering process affecting the final duplex yield have not yet been systematically studied.

Minimising of secondary structures by preparation of single stranded and short targets are favourable to increase the availability of the complementary sequences to the arrayed probes (Southern et al., 1999). The different methods of target preparation are described below.

Short oligonucleotides are preferably immobilised onto the silanised glass surface via different linker molecules at their 3' or 5' end, which facilitate their mobility and availability to the target molecules (Guo et al., 1994, Afanassiev et al., 2000). The effects of possible steric crowding, linker length and charge have been systematically studied. 30 to 60 atom hydrophilic linkers with low negative charge density were found to be optimal for hybridisation. With the best linkers, the duplex yield was increased by maximum of 150 times (Shchepinov et al., 1997). Poly dT spacer tailing has been used as an additive to the linker molecules to increase accessibility of the probes for hybridisation (Guo et al., 1994). Most of the systematic studies on thermodynamic properties of oligonucleotide probes on microarray have been performed with relatively short oligonucleotides (6- to 10-mers) and can therefore give only a general background to understand the hybridisation process with longer probes used for various applications (18- to 25-mers for mutation detection and 20- to 60-mers for expression profiling). Therefore, each new assay still needs to pass an evaluation process and experimental testing for performance.

- Sequencing by hybridisation on oligonucleotide microarrays

The SBH arrays have been applied to testing of known mutations and unknown variations in the known genes (Hacia, 1999, Wen et al., 2000, Wikman et al., 2000) and also extensive genotyping of single nucleotide polymorphisms (SNP-s) (Cho et al., 1999, Patil et al., 2001). The number of oligonucleotides per identified base varies depending on the type of array. Mutation-specific SBH arrays can be made of allele specific oligonucleotides (ASO) complementary for wild-type and mutant alleles (Saiki et al., 1989). SBH type scanning arrays for resequencing of certain genes consist of large number of oligonucleotides, at least four different oligonucleotides per each base in the

target sequence (Figure 4), and potentially allow for identification of unknown sequence variations.

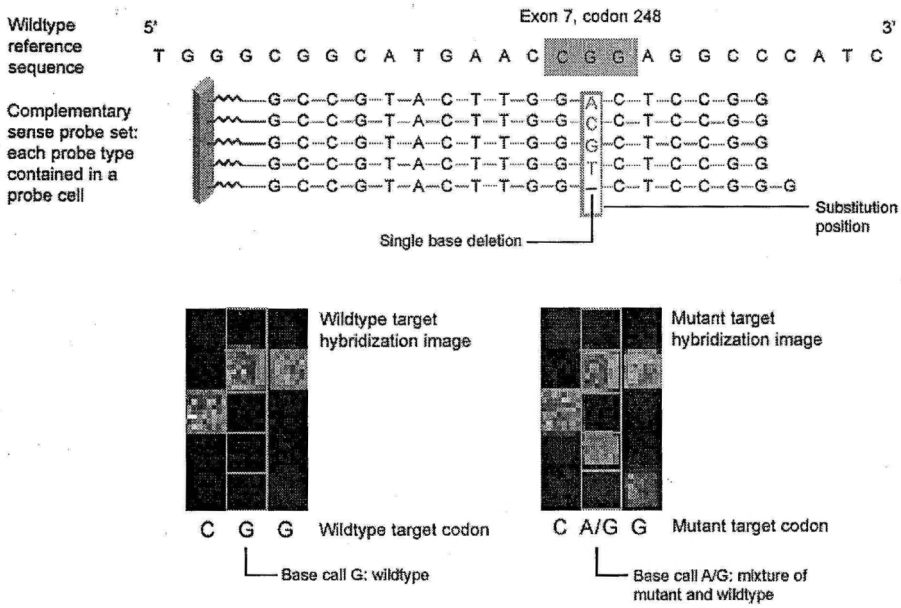


Figure 4. Sequencing by hybridisation. The probe matching the target sequence will generate the strongest signal. Two signals correspond to a heterozygous mutation or a somatic mutation on the background of wild-type sequence. TP53 GeneChip®, Affymetrix, Inc.

However, the scanning arrays have been found insensitive, particularly for heterozygous (Cronin et al., 1996) and somatic (Wen et al., 2000, Wikman et al., 2000) insertions and deletions. Therefore a highly redundant set of oligonucleotides is often used even for the mutation-specific arrays. A mutation-specific array has been designed for CFTR gene with total of 1480 oligonucleotides for 37 mutations. 10 DNA samples were tested for mutations in exons 10 and 11. The microarray showed 100% concordance rate with standard dideoxy sequencing (Cronin et al., 1996). This set of DNA samples consisted mostly of homozygous sequences for the two exons and therefore it is difficult to extrapolate the results for overall performance level. As further advancements of the scanning SBH array, microarrays with combinations of up to 28 oligonucleotides per base pair in the target sequence have been designed (Hacia et al., 1996). A number of studies have been performed to evaluate their feasibility for identification of both known and unknown sequence variations (Hacia et al., 1996, Kozal et al., 1996, Chee et al., 1996, Wen et al., 2000, Wikman et al., 2000). The composition of oligonucleotides consists of wild type

sequence, base substitution, base insertion and one- to five-base deletions for both DNA strands. A highly complex BRCA1 gene array was evaluated in two-step analysis with fluorescein / phycoerythrin two-colour labelling of the studied and reference samples and both gain-of-signal, and loss-of-signal algorithms involved (Hacia et al., 1996). Seven out of 15 pre-known mutations studied performed well in clear gain- and loss-of-signal results. For the rest of mutations, sensitivity of any of the algorithms was imperfect. Loss-of-signal algorithm was actually found to be more robust as with insertions and deletions a significant cross-hybridisation occurred with the wild-type sequence and no clear signal was detected for gain-of-signal analysis. On the other hand, in addition to the known mutations, seven unknown polymorphisms were detected plus a previously unreported base substitution. The minimum acceptable mutant / wild-type signal ratios for calling the mutation in gain-of-signal analysis was found to be 1.3. In heterozygous targets, the ratios were between 0.6 to 1.4 and the genotype assignment criteria were therefore not fully refined (Cronin et al., 1996). In case of the loss-of-signal analysis a cut-off value 1.2 for the signal ratios between the wild-type reference and mutant samples has been found optimal (Hacia et al., 1996).

A recent study has shown the good sensitivity of TP53 gene SBH scanning microarray, GeneChip™ p53 Assay (Affymetrix, Inc.) for the fraction of mutated DNA in case of two point mutations. Even 1% of mutated DNA was successfully identified with the chip score calculated by the analysis software (Wikman et al., 2000).

Universal DNA arrays with complete sets of n-mer oligonucleotides have been designed for *de novo* sequencing. A complete n-mer set consists of 4^n oligonucleotides, e.g 65 536 octameres. The target sequence is constructed from overlapping sequences of hybridised oligonucleotides. The length of unknown sequence, which can be reconstructed from hybridisation to a complete set of oligonucleotides of a given length equals roughly to the square root of the number of oligonucleotides in the complete set (Southern et al., 1992). For example, array of 65 536 octameric oligonucleotides would allow for reconstruction of approximately 200 bases of target sequence. However, in real life the duplex stability is low with short oligonucleotides and only non-stringent hybridisation conditions can be used. Complete sets of longer oligonucleotides become also too expensive for manufacturing (for example, the full set of decameric sequences exceeds one million oligonucleotides). Therefore the SBH experiments with universal DNA arrays have not exceeded the "proof of principle" evaluation state (Strezoska et al., 1991, Southern et al., 1992, Drmanac et al., 1993, Drmanac et al., 1998).

Target preparation for SBH and the labels used

Different target preparation protocols have been used for the SBH arrays. Thermodynamically, short and single-stranded targets have advantages for use with oligonucleotide arrays. For format II arrays, the preparation has to start with PCR (Mullis and Faloona, 1987) at first. The PCR may be complemented with asymmetric PCR (Cronin et al., 1996) or *in vitro* transcription reaction to generate RNA template providing greater duplex stability (Hacia et al., 1996, Kozal et al., 1996).

Hybridisation of the target to the oligonucleotide array is greatly facilitated if its length is in the same range that the length of oligonucleotide probes. The only practical solution to produce short targets is a fragmentation reaction. The choice of target fragmentation methods for SBH arrays includes:

- DNA fragmentation with Uracil-DNA-Glycosylase (Cronin et al., 1996)
- RNA fragmentation with heating at the presence of Mg^{2+} (Chee et al., 1996, Hacia et al., 1996)
- DNase I digestion (Wikman et al., 2000, Patil et al., 2001).

The labels used include single (Cronin et al., 1996) or dual fluorophores, biotin for later colorimetric detection (Saiki et al., 1989) or a radioactive isotope (Southern et al., 1992). In case of dual colour labelling, biotin- and fluorescein-labelled dNTP-s were first incorporated and the second fluorophore, phycoerythrin, is added later as streptavidin conjugate (Hacia et al., 1996). The labelling has been performed either during asymmetric amplification (Cronin et al., 1996), *in vitro* transcription (Hacia et al., 1996) or DNase I fragmentation (Wikman et al., 2000, Patil et al., 2001) steps.

Commercial SBH platforms

AFFYMETRIX GENECHIP™ ARRAYS

Affymetrix GeneChip™ arrays (Affymetrix, Inc., http://www.affymetrix.com) are the most developed *in situ* synthesised oligonucleotide arrays today. The main advantage is in the utmost high number of features per array. The main disadvantage is that only factory pre-made arrays can be used and the usage needs specific hardware from Affymetrix, compatible only with the GeneChip™ arrays. In addition, the oligonucleotides are bound to the surface by the 3' end, which does not allow performing of array-based enzymatic reactions described in the next paragraph. Despite of significant activities to detect DNA variations in mid-90's, most of the *in situ* synthesised GeneChip® products of Affymetrix are now made for differential gene expression analysis, except for TP53 gene

assay (Ahrendt et al., 1999), Cytochrome P450 mutation assay, an assay for 1500 SNP-s and GenFlex™ Tag Arrays (Fan et al., 2000).

WHOLE WAFER ARRAYS

Perlegen, Inc. (Mountain View, CA, http://www.perlegen.com/), the daughter company of Affymetrix has expanded large photolitographically *in situ* synthesised oligonucleotide arrays to chromosome- and genome-wide SNP genotyping and haplotype identification (Patil et al., 2001). The glass wafers have 5 x 5 inch size and carry approximately 60 million different probes each.

HYSEQ

Hyseq Pharmaceuticals, formerly Hyseq, Inc. (Sunnyvale, CA, USA, http://www.hyseq.com/) was initially focused in the use of format I arrays with immobilised targets. The focus was later changed later to use of universal DNA arrays with complete sets of short oligonucleotides (Drmanac et al., 1998). All the efforts to develop SBH products have been passed to their majority owned subsidiary, Callida Genomics.

NANOCHIP® ELECTRONIC MICROARRAYS

Nanogen, Inc. (San Diego, CA, USA, http://www.nanogen.com/) has combined SBH with addressable semiconductor microchip. The manipulation of an electric field on the semiconductor microchip allows for rapid concentration, hybridisation, dehybridisation and detection of DNA molecules at designated test sites (Sosnowski et al., 1997, Edman et al., 1997, Gilles et al., 1999). It has been shown that electric field may speed up the kinetics of hybridisation by at least 25-fold. The method allows to use buffers, which do not support passive hybridisation. NanoChip® electronic microarrays with 100 addressable reaction sites per array are commercially available.

DASH.

Dynametrix Ltd.(UK, http://www.dynametrix-ltd.com/) is utilising method called Dynamic Allele Specific Hybridisation (DASH) (Howell et al., 1999). The method is based on real time hybridisation monitoring with DNA probes on format I macroarrays on dynamic temperature conditions. The melting curves of DNA will be different for perfect matching probes and mismatched probes. Genotypes are deduced from these differences (Figure 5). The hybridisation status may be monitored with either an intercalating dye only, giving signal if double-stranded DNA is present, or by intercalating dye inducing fluorescence resonance energy transfer (iFRET) with oligonucleotide probe labelled with acceptor fluorophore (Howell et al., 2002).

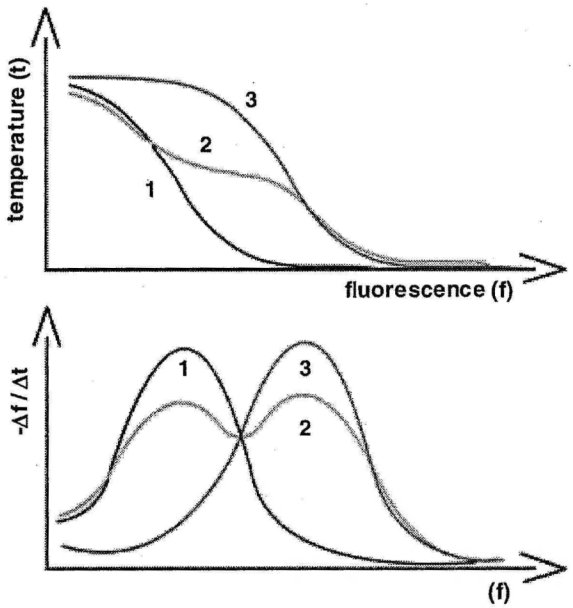


Figure 5. DASH melting profiles and derivatives. DASH melting profiles and the computed derivatives for analysis: 1-homozygous mismatch, 2-heterozygous mismatch, 3-homozygous match

GENIOM® SYSTEM.

Febit AG (Mannheim, Germany, http://www.febit.de/geniom/technology.htm) has integrated light-directed *in situ* oligonucleotide synthesis with the target processing and CCD-based detection onto the same Geniom® workstation. Both synthesis and processing are carried out in three-dimensional micro-channel chip. The light activation is achieved by electronically controlled micromirror system instead of physical masks. Febit's system is currently the most flexible oligonucleotide *in situ* synthesis platform allowing for large scale (up to 80,000 features per chip) and instant changes in chip design. Oligonucleotides may be synthesised at both conventional 3'to 5' and 5'to 3' directions, enabling enzymatic elongation of the oligonucleotide probes.

- Dual selection by hybridisation and enzymatic elongation on oligonucleotide microarrays

Hybridisation is an essential step in all DNA array platforms. However, the high complexity of SBH assays due to the large number of oligonucleotide probes per each targeted base, relatively low overall signal to noise ratio and sensitivity to subtle differences in hybridisation conditions, have driven several groups to

look for alternative technologies. A key improvement for microarray technology lies in the use of enzymes, either polymerase (Pastinen et al., 1997, Shumaker et al., 1996) or ligase (Landegren et al., 1988, Gunderson et al., 1998) as additional allelic recognition mechanism. SBH is based on single nucleotide mismatch discrimination between different ASO-s and the target sequence. The technologies utilising polymerase and ligase consist of two main reaction steps: target annealing to the oligonucleotide probes (primers) and their extension with either labelled terminator nucleotides (dideoxyribonucleoside triphosphates, ddNTP), labelled deoxyribonucleoside triphosphates (dNTP) or labelled oligonucleotides complementary to the target sequence. In contrary to SBH technologies, which are designed to discriminate against internal mismatches on the hybridised oligonucleotide probes, the enzymatic reaction is most sensitive to mismatch directly in the probe 3' end. The simultaneous use of hybridisation and allele-specific enzymatic elongation steps have largely reduced the complexity of oligonucleotide microarrays and provided a more than an order of magnitude increase in general signal-to-noise ratio compared to SBH (Kurg et al., 2000, Pastinen et al., 1997).

Primer extension on oligonucleotide microarray

The early versions of solid phase primer extension were performed in microtiter plate format with affinity-captured single stranded target DNA and primers in solution for analysis of SNP-s of human apolipoprotein E (Syvänen et al., 1990). Taq DNA polymerase and Klenow fragment *E.coli* DNA polymerase I were used with radioactively labelled dNTP-s and three competing unlabelled ddNTP-s in the reaction mixture (Syvänen et al., 1990). Introduction of fluorescent labels and gel-based separation of the extension products enabled multiplexing for the first time (Pastinen et al., 1996, Shumaker et al., 1996). Arrayed Primer Extension (APEX) was a promising solution for highly parallel and precise identification of sequence variations (Shumaker et al., 1996). Different formats of primer extension on solid support are shown on Figure 6.

APEX is similar to minisequencing (Syvänen et al., 1990, Pastinen et al., 1997) and Genetic Bit Analysis, GBA (Nikiforov et al., 1994a, Head et al., 1997) in microarray format. In the current version, APEX is based on incorporation of four dye terminators onto oligonucleotide primers by a thermostable DNA polymerase (Kurg et al., 2000). This can be viewed similar to Sanger dideoxy sequencing technology (Sanger et al., 1977). Instead of one primer which is elongated into many fragments and fragment length is used for their separation, APEX is using many primers separated on the array and elongated by only one base.

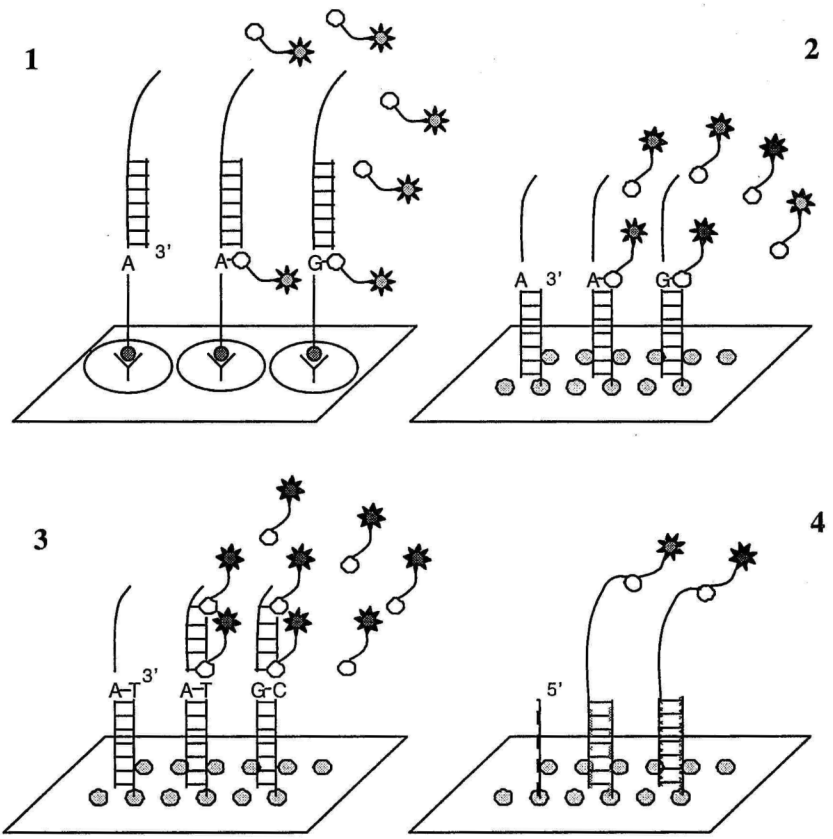


Figure 6. Different formats of primer extension on solid supports.
1 – Format I, affinity-captured target DNA on microtiter plate, primers will hybridise from solution
2 – APEX, N-GBA, Minisequencing on microarray with single base extension
3 – ASO +dNTP primer extension
4 – Extended primers captured by Tag Array. Primer extension takes place in solution.

APEX is capable for identification of different types of mutations and polymorphisms and can be used for resequencing (Head et al., 1997, **Metspalu**, 1999, Tõnisson et al., 2002). Moreover, different genes and diseases may be studied in parallel with the same array. For testing of known mutations, the primers can be always designed so that discrimination between wild-type and mutated alleles is possible with single primer or a pair of ASO primers. For unknown mutations, the only option is to use a resequencing array (Head et al., 1997, Tõnisson et al., 2002).

Only the set of arrayed probes must be changed to apply APEX technology for different applications: mutation detection, SNP genotyping and resequencing. Oligonucleotides are synthesized according to the wild-type sequence with

aminolink or disulfide group at their 5'end to leave 3'end of oligonucleotides free for extension by DNA polymerase. The 3' end of each oligonucleotide is positioned one nucleotide before the nucleotide to be identified. For resequencing on the APEX assay, oligonucleotides are designed with one nucleotide shift on their sequence. For example, 1000 oligonucleotides are needed for identification of 1000 bases in the sample DNA. As APEX allows analysis of both strands in the same reaction, two oligonucleotides are arrayed per each basepair.

Some groups have implemented enzymatic elongation of ASO probes (Gemignani et al., 2002, Pastinen et al., 2000). This modification of primer extension method adds flexibility to oligonucleotide design and allows incorporating multiple labels to the same elongated primer if dNTP-s are used instead of ddNTP-s. In addition, incorporation of dNTP-s is also efficient if reverse transcriptase is used as the enzyme with RNA template (Pastinen et al., 2000). With ARMS, some types of mismatches at 3' end of primers, like G-T and A-C, have been reported to allow for partial elongation. In these cases, apyrase added to the reaction mixture was shown to overcome the difficulties (O'Meara et al., 2002).

The use of primer extension on oligonucleotide microarrays has exceeded the scope of evaluation steps showing its efficiency for candidate gene studies (Pastinen et al., 1998), SNP validation (Lindroos et al., 2002, Dawson et al., 2002) and the characterisation of allelic association in human chromosome 22 (Dawson et al., 2002).

Template preparation for primer extension

Preparation of template (here synonymous to target) DNA for primer extension starts from single- or multiplex amplification by PCR. The further steps have included either preparation of single stranded DNA, RNA or fragmentation of the double stranded DNA. Single stranded DNA has been the template of choice with thermolabile DNA polymerases used for primer extension (Syvänen et al., 1990, Pastinen et al., 1996, Head et al., 1997). The most common method of generating single stranded template is the use of biotin labelled primer in PCR complemented with later strand separation procedure with alkali treatment. As an alternative to biotin labelling, T7 gene 6 exonuclease digestion has been used for PCR products amplified with phosphorothioate primers (Nikiforov et al., 1994b, Head et al., 1997). RNA template has been prepared by *in vitro* transcription using the amplified PCR products and T7 RNA polymerase (Pastinen et al., 1997, Pastinen et al., 2000).

Thermolabile DNA polymerases like Klenow fragment *E.coli* DNA polymerase I, Sequenase and Sequenase Version 2.0 T7 DNA polymerase were the first used in primer extension reactions (Syvänen et al., 1990, Shumaker et al., 1996). Both of the polymerases lack 5'-3' exonuclease activity. The Klenow fragment of DNA polymerase yields significant misincorporation, probably because of its 3'-5' exonuclease activity (Syvänen et al., 1990). Exonuclease-free Klenow fragment DNA polymerase performs much better for primer extension and has been successfully used for array-based sequencing of 33bp region in exon 8 of TP53 gene (Head et al., 1997). The characteristic property of Sequenase Version 2.0 T7 DNA polymerase is its high processivity (Tabor and Richardson, 1987).

The use of thermostable enzymes allows for dynamic control over the hybridisation stringency. *Thermus aquaticus* (Taq) DNA polymerase discriminates against ddNTP-s. Introduction of Phe667Tyr mutation reduced this discrimination 250 to 8000 fold (Tabor and Richardson, 1995). A number of engineered thermostable DNA polymerases have been designed and used for primer extension including Thermo Sequenase™ (Amersham Biosciences, Milwaukee, WI, USA), DyNASeq™ (Finnzymes OY, Espoo, Finland), AmpliTaq® FS (Roche Molecular Systems, Inc., Pleasanton, CA, USA) and others (Pastinen et al., 1997, Chen et al., 1999, Kurg et al., 2000). Tma 31 FS DNA polymerase derived from *Thermotoga maritima* (Roche Molecular Systems, Inc.) is one of the most recent ones, which interestingly has preference to ddNTP-s over dNTP-s. Tma 31 FS DNA polymerase has allowed mass-spectrometric genotype detection, without removal of residual dNTP-s from PCR before primer extension step (Sauer et al., 2002).

Reverse transcriptase (RetroTherm RT™, Epicentre Technologies, Madison, WI, USA) has shown good allele discrimination capabilities with RNA as the template (Pastinen et al., 1997, Pastinen et al., 2000).

Both isotope and fluorescent labels conjugated to dNTP-s, as well as ddNTP-s have been used to detect the primer extension products. Radioactive labelling has two advantages. First, incorporation of the nucleotides is not affected by additional chemical groups, and second, the sensitivity is higher compared to direct fluorescent labelling. The radioactive nucleotides include [α^{35}S]dNTP and [α^{3}H]dNTP (Syvänen et al., 1990), [α^{32}P]dNTP, (Shumaker et al., 1996) and [α^{33}P]ddNTP (Pastinen et al., 1997). Phosphorimaging equipment and scintillation counter were used for detection.

Fluorescent labels are less toxic and have allowed developing assays with multicolour detection. Single dye – four reactions, two dyes – two reactions and four dyes – single reaction set-ups have been used. The use of four colours per reaction permits straightforward simultaneous identification of all possible

sequence variants within the same process (Kurg et al., 2000, Lindroos et al., 2002) (Table 2). The choice of dye labels is determined by three factors:

- the fluorophores must be spectrally well separated;
- the DNA polymerase used must efficiently incorporate the labelled nucleotides and
- the labelled nucleotides must be available.

Intellectual property and cross-patenting reasons have become inhibiting factors for development of multicolour mutation detection and genotyping technologies.

Table 2. Fluophores used in four-colour primer extension on oligonucleotide microarray (Kurg et al., 2000, Lindroos et al., 2002), spectral data is obtained from http://www.perkinelmer.com/nucleotide_analogs.

Fluorescent dye	Excitation maximum (nm)	Extinction Coefficient ($M^{-1} \cdot cm^{-1}$)	Emission maximum (nm)
Fluorescein	497	30,000	517
Cy™3	550	150,000	570
TAMRA	552	91,000	575
Texas Red	593	85,000	612
Cy™5	648	250,000	667

Primer extension is a complex reaction consisting of primer annealing to the target and enzymatic elongation extension reaction with labelled nucleotides. Template hybridisation and primer extension differ in their optimal temperatures. Hybridisation is more effective in lower temperatures, whereas thermostable polymerases work better in higher temperatures. For this reason, some primer extension protocols are designed in two separate steps, first – hybridisation and second – primer extension by appropriate DNA polymerase (Pastinen et al., 1997). Unbound target may be removed by washing between these two steps. For thermolabile polymerases, the templates have been first heated and then annealed to primers by cooling down to room temperature. The DNA polymerase was added in the second step (Syvänen et al., 1990, Shumaker et al., 1996).

In our opinion, the reaction conditions on microarray are better controlled by isothermal set-up, suitable both for the template hybridisation to oligonucleotide primers and enzymatic elongation by the DNA polymerase (Kurg et al., 2000).

In case if generic tag arrays are used for genotyping based on SBE, cyclic extension reaction is performed with primers in solution. This has become an efficient way of analysing full-length double stranded PCR products (Fan et al., 2000, Lindroos et al., 2002).

The mutation detection process includes typically a semiquantitative analysis step where signals must first be normalised.
With equal target amounts hybridised, the primer extension signal intensities depend on:

- The incorporation efficiency of labelled nucleotides. Dye-labelled nucleotides are each incorporated with different efficiency by the DNA polymerase used.
- The brightness of the fluorophore. A fluorophore brightness is proportional to its extinction coefficient and quantum efficiency. Quantum yield or efficiency shows the ratio of emitted and absorbed photons and is affected by photobleaching. The fluorescence is strongly influenced by local environment such as also the fluorophores conjugation to nucleotides and proteins, and the resulting quantum yields vary as well.
- The excitation power and efficiency of detection system at particular wavelength.

Due to the multiple factors involved, the actual signal intensities have to be experimentally determined and used for empirical normalisation (Lindroos et al., 2002).

- Other novel methods combining hybridisation with enzymatic reactions

TAG ARRAYS

Tag arrays (Figure 6) were first introduced by Affymetrix, as a useful improvement to *in situ* synthesised high-density oligonucleotide arrays (Fan et al., 2000). The method has three major advantages over the ordinary SBH arrays:

- Increased flexibility of the assay design. *In situ* photolithographic synthesis has a rigid set-up due to the high cost of masks used. The generic tags allow the same arrays to be used for variable applications and assays.
- The possibility to introduce enzymatic elongation step as the primers specific to the target sequence have 3' end free for extension.
- The possibility to avoid steric factors during hybridisation as the elongation step is performed in solution phase. In addition, cyclic primer extension may be easily implemented, which reduces the effects of primer secondary structures and allows to analyse full length PCR-amplified templates.

The main disadvantages are the additional tag probes needed and the availability of large number of tag sequences. Each identified base needs one tag. Two tags are required for a basepair. The tags must be fully unique, and not

to occur in the target sequence. Synthesis of 32.000 tags has been reported, which may not be sufficient for whole genome applications. Nevertheless the tag arrays gain popularity and are now also commercialised by Orchid Biosciences (Princeton, NJ, http://www.orchid.com) (Bell et al., 2002).

BEADARRAY™ TECHNOLOGY

Single base primer extension, allele-specific primer extension and oligonucleotide ligation assays have been successfully combined with modified zip-code detection in high throughput BeadArray™ genotyping platform (Illumina, San Diego, CA, USA, http://www.illumina.com) (Oliphant et al., 2002, Steemers et al., 2000). The arrays consist of glass fibre bundles with chemically etched termini. Each fibre carries its own isolated optical signal from one end to the other due to total internal reflection. As the fibres are packed, the whole array can be simultaneously detected. Arrays are prepared by randomly distributing a mixture of microsphere sensors, each with unique optical signature because of entrapped dyes, Texas red cadaverine and europium (III) thenoyltrifluoroacetate-$3H_2O$. Each polystyrene microsphere carries also unique tag sequence that is used for capturing and detection of the oligonucleotide elongation products.

PRIMER EXTENSION ON MICROSPHERES

Instead of separation on DNA microarray, uniquely tagged microspheres may be used for separation of different SNP-s by a system similar to fluorescence activated cell sorter. The microbeads have been used for cyclic primer extension in solution phase. Each microsphere carries unique optical code and by carrying unique sequence tag, identifies one SNP. The label from primer extension reaction identifies the genotype (Taylor et al., 2001). This platform of up to 100 different analytes in a single well has been commercialised by Luminex Corporation (Austin, TX, USA, http://www.luminexcorp.com).

FLUORESCENCE POLARISATION

When a fluorescent molecule is excited with polarised light, it emits fluorescent light into a plane related to the rotational freedom of molecule itself. The rotational freedom is proportional to the molecule volume and mass. According to this principle, the fluorescence polarisation (FP) will occur in the result of enzymatic incorporation of fluorescent labels into probes complementary with the target sequence. FP is expressed as the ratio of fluorescence detected in the vertical and horizontal axes.

For a system in which the fluorophore is attached to a low molecular weight nucleotide and then is incorporated into the oligonucleotide probe at the allelic site, the degree of fluorescence polarization is described by the equation

$$P = P\max[\mathrm{ddNTP}]b + P\min[\mathrm{ddNTP}]i - [\mathrm{ddNTP}]b$$

where *P*max is the polarization for dye-labeled ddNTP incorporated onto the oligonucleotide probe, *P*min is the polarization of the unincorporated dye-labelled ddNTP, [ddNTP]*i* is the initial concentration of dye-labelled ddNTP, and [ddNTP]*b* is the concentration of incorporated dye-labeled ddNTP. FP was shown to work well with single base primer extension method (Chen et al., 1999). The degree of FP increases linearly up to 10,000 Daltons before it levels off. Since the molecular weight (MW) of fluorescent nucleotides is approximately 1000 Daltons and the 25- to 30- mer probes have MW 10,000 Daltons, the weight range fits well detection by FP.

The advantage of fluorescence polarisation is that it enables homogeneous assay set-up. No washing steps are required before detection. FP was recently also set up as four-colour assay with four spectrally separated fluorophores (Kwok, 2002).

MASS-SPECTROMETRY FOR GENOTYPING AND MUTATION DETECTION

Combination of primer extension method with mass spectrometric detection has enabled to develop label-free assays, as molecule mass of extended probes may be the tag for genotype detection. Matrix Assisted Laser Desorption / Ionization – Time-of-Flight mass spectrometry (MALDI TOF) is one of the most suitable mass spectrometry methods for biomolecules. The samples are loaded on matrix with light absorbing crystals. The sample and matrix are vaporised by a pulse hit from laser. Negatively charged primer extension products are accelerated towards detector by electric field applied. The time between the laser pulse and detection is in precise correlation with the molecule mass of extended probes.

Mass detection permits for homogeneous assay set-up as the unused nucleotides do not interfere with detection of extension products.

Two versions of primer extension reaction are commonly used for mass detection:

First, primer extension with individually optimised mixture of dNTP-s and ddNTP-s so that elongation products from each known allele are made with different length. The advantage is the bigger mass difference of different extension products. This version has been commercialised by Sequenom (San Diego, CA, USA, http://www.sequenom.com) as MassARRAY platform (Jurinke et al., 2002).

The second, option is to use partial phosphodiesterase digestion of extended primers and charge tags. MW difference of ddNTP-s is low and without the tags and cleavage the different genotypes would overlap in their mass spectra. Both negative and positive charge mode detection has been used. This type is known as GOOD assay (Sauer et al., 2000a, Sauer et al., 2000b). GOOD assay was recently complemented with Tma 31 FS DNA polymerase (Roche Molecular Systems), which has preference for ddNTP-s over dNTP-s (Sauer et al., 2002). If applied to detection of short extension products from cyclic dideoxy

sequencing reactions, MALDI TOF mass spectrometry allows rapid sequencing of 15 to 20 nucleobases (Nordhoff et al., 2000).

OLIGONUCLEOTIDE LIGATION REACTION

Oligonucleotide ligation complementary to the target DNA is even more discriminative to terminal mismatches than primer extension due to stacking interactions between the two oligonucleotides (Parinov et al., 1996). Two principally different solid phase oligonucleotide ligation methods have been published:

- Ligation of single stranded or fragmented target DNA to duplex or stem-loop oligonucleotide probes with terminal overhangs (Broude et al., 1994, Gunderson et al., 1998, Broude et al., 2001). Either target carries the label or additional primer extension reaction is performed on the ligated complex for detection.
- Ligation of a "zip-coded" and a labelled oligonucleotide according to the target sequence (Gerry et al., 1999). This method is similar to the use of tag arrays in primer extension.

The main disadvantage of ligation assays is the need for additional oligo-nucleotide probes.

PYROSEQUENCING™

Pyrosequencing™ is a label-free non-electrophoretic genotyping and mutation detection method where short fragments of the target are sequenced with real time detection (Garcia et al., 2000). The method has been developed and commercialised by Pyrosequencing AB (Uppsala, Sweden, http://www.pyrosequencing.com). The incorporation of bases by DNA polymerase is transformed into light by a mixture of two enzymes: ATP sulfurylase, and luciferase. Apyrase continuously degrades the unincorporated bases and excess ATP. The light impulse is proportional to the number of bases incorporated. The reactions are performed in 96- and 384-well microtiter plate format. Pyrosequencing is an automated sequential process where a new mixture of enzymes is added after each signal detection step. The main disadvantages are the limited parallelisation capacity and relatively high price due to the use of multiple enzymes.

INVADER® ASSAYS

A number of genotyping and mutation detection assays have been developed by Third Wave Technologies (Madison, WI, http://www.twt.com), which utilise structure-specific enzymes, flap endonucleases (Cleavase™). Cleavase cuts junctions between the single- and double-stranded regions of DNA and generates linear amplification process(Brow et al., 1996). The assays are based on the observation that single strands of DNA form highly individual higher

order structures by folding on themselves. Invader assays have shown to be highly allele-specific, robust and reproducible and have recently been set up with dual colour detection using FRET (Olivier et al., 2002a). This is one of a few methods available, which allows to avoid PCR and detect genetic variations starting from genomic DNA directly (Ryan et al., 1999). The main disadvantage is that the amplification process has to take place in individual wells and the assays may therefore be set up only in microtiter plate format.

5' NUCLEASE ASSAY – TAQMAN®

Although not based on microarrays, TaqMan has been one of the most widely used genotyping methods for small and medium scale genetic studies for last couple of years. The method is based on enzymatic cleavage of hybridised dual labelled TaqMan probes by Taq DNA polymerase during PCR because of its 5'-3' exonuclease activity(Holland et al., 1991). TaqMan probes are designed as ASO-s with dual fluorophore labelling, donor (reporter) and an acceptor (quencher). Due to fluorescence resonance energy transfer (FRET), the acceptor label acts as a quencher, and no fluorescence occurs. After cleavage, the donor will be separated and the fluorescence is detected without any additional manipulation with the reaction mixture. In addition to detection of sequence variations, the quantitative detection has allowed TaqMan to be used for gene expression studies and DNA copy number estimation, which of extreme importance in prenatal genetic testing.

TaqMan method and PCR equipment for real time fluorescence detection are commercialised by Applied Biosystems (now being a part of Applera Corporation, Foster City, CA, USA, http://www.appliedbiosystems.com/).

3. PRESENT INVESTIGATIONS AND DISCUSSION

3.1. Objectives

The current study was aimed on:

- Development of integrated technology platform for Arrayed Primer Extension
- Evaluation of the technology with a number of assays at different scale, different mutation types included and different performance requirements.

The following aspects were studied during development of the integrated APEX technology platform:

- Selection and optimisation of suitable chemical methods for binding oligonucleotide primers onto the microarray slides.
- Development of optimal template preparation methods and APEX reaction conditions.
- Development of a suitable fluorescence detection system for collection of images from four different fluorescence channels, each corresponding to a different dye terminator.
- Development of software package for semiautomatic analysis of the APEX images.

3.2. Development of integrated technology platform for arrayed primer extension

3.2.1. Slide chemistry for oligonucleotide arrays

Publications I, II, IV

The attachment chemistry of oligonucleotides on glass support must provide a reproducible coating process of the slides and functionality of the surface-bound DNA probes for hybridisation. The reliable chemical linkage must also be stable, produce specific binding and eliminate undesired steric hindrance on the glass support.

We have evaluated two most commonly used silane-modified surfaces for APEX array preparation: epoxysilane-modified (Lamture et al., 1994) and aminosilane-modified surfaces (Guo et al., 1994) (Figure 7). Oligonucleotide primers are attached to epoxy-activated glass by secondary amine formation

between epoxy-derivatised glass surface and 5' amino linker. Primers are diluted in sodium hydroxide and spotted onto the activated surface (Shumaker et al., 1996). For amino derivatisation the slides are silanised with 3-amino-propyltrimethoxysilane. To convert the amino groups to amino-reactive phenylisothiocyanate groups the slides are treated with 1,4-phenylene diisothiocyanate. Amino modified oligonucleotide primers are diluted in sodium carbonate/bicarbonate buffer (pH 9.0) and spotted onto the activated surface. The slides are later blocked with NH_4OH, washed with water, air dried and stored at room temperature. In our hands the slides have been stable for at least for six months. Both binding methods were suitable for APEX. Epoxy derivatisation process has less steps and has therefore potentially better reproducibility. On the other hand, the hybridisation signal intensities were up to ten times higher and the same primer extension signal intensities at least two times higher on thiocyanate-coated slides (unpublished data). Therefore the epoxy-modified slides are no longer used for APEX.

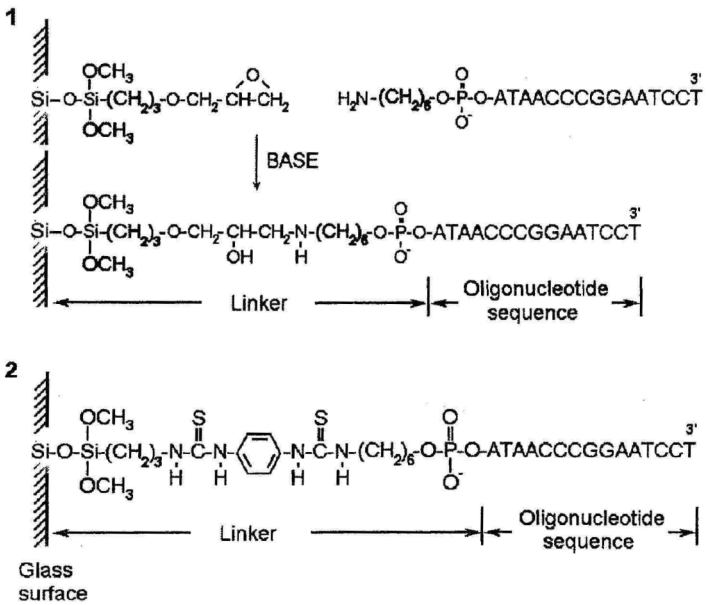


Figure 7. Epoxy- (1) and aminosilane (2) coatings and oligonucleotide attachment (modified from Lamture J. *Nucleic Acids Res,* 1994 22, 2121–5. and Guo Z. *Nucleic Acids Res.* 1994, 22, 5456–65.)

Our early arraying equipment consisted of a standard pipetting robot (Tecan RSP5031, Tecan AG, Hombrechtikon, Switzerland) and a custom-manufactured 96-channel parallel spotting device (Kurg et al., 2000, Tõnisson et al., 2000a, Tõnisson et al., 2000b). Both options were suitable for low-density arrays. The main disadvantage was loss of oligonucleotides due to the relatively large spot

volume (approximately 0.5µl per spot) and large dead volume of the arraying equipment. Microarrays were produced with Genetic MicroSystems (now Affymetrix) 417 pin and ring type arrayer (Tõnisson et al., 2000b). This arrayer enables nanoliter volume spotting and smaller centre-to-centre distance of spots (200-375 µm), increasing the maximum number of oligonucleotides per microarray by more than two orders of magnitude.

3.2.2. Oligonucleotide design and selection

Publications I, IV, V

We have used 18-, 20- and 25-mer oligonucleotides in APEX assays. More intense signals were obtained with longer oligonucleotides, i.e. the signals from 25-mers were 10 to 50% stronger than from 20-mers for the same sequences (Kurg et al., 2000). It may be explained by both with increased duplex stability and accessibility of probes for hybridisation.

Vast majority of the oligonucleotides performs well in APEX. Unfortunately, longer oligonucleotides have also more possibilities to form stable secondary structures (hairpins or oligonucleotide dimers), which may result in either loss of signal or a false positive signal due to self-extension. Although no perfect method exists for precise prediction of overall hybridisation and self-extension behaviour of arrayed primers, the potentially harmful structures are predictable by computer algorithms and the oligonucleotides may be designed beforehand to reduce stability of the secondary structures. In case of TP53 gene APEX resequencing assay, 5.9% of the oligonucleotides were re-designed by introducing a mismatch base to reduce the stability of the predicted dimers and avoid self-priming (Table 3). After modification, 62% of these oligonucleotides were found to generate signals only in the presence of target DNA and not from oligonucleotide dimers. The rest of modified primers did not give constant signals in the APEX reactions. Importantly, none of the modified oligonucleotides were found to generate false-positive signals in the absence of the target DNA fitting well with the "yes or no" type general philosophy of APEX analysis (Tõnisson et al., 2002). Mismatches in central parts of oligonucleotides are known to affect most the duplex stability. However, as seen from the table, introduction of one mismatch base into the primer still permits for specific hybridisation. Only 3' end of primers and its proximity does not tolerate mismatches for enzymatic elongation by dye terminators. This factor is used for designing ASO-s for primer extension assays (Pastinen et al., 2000, Gemignani et al., 2002).

APEX primer	Version 1 sequence 5'-3'	Wild-type base	Self-extension	Internal complemenarity enabling self-extension	Version 2 sequence 5'-3', mismatch base underlined
hp53-11917s	CTCTTGTCTTTCAGACTTCCTGAAA	A	**G**	TTTCAG-----CTGAAA	CTCTTGTCTTT**A**AGACTTCCTGAAA
hp53-12040s	CTACAGTCCCCCTTGCCGTCCCAAG	C	**G**	CTTG------CAAG	CTACAGTCCCC**A**TTGCCGTCCCAAG
hp53-12043s	CAGTCCCCCTTGCCGTCCCAAGCAA	T	**G**	TTGC--------GCAA	CAGTCCCCCT**A**GCCGTCCCAAGCAA
hp53-12209as	GGTAGGTTTTCTGGGAAGGGACAGA	A	**A**	TCTG--------CAGA	GGTAGGTTTT**A**TGGGAAGGGACAGA
hp53-12296s	CTGGGACAGCCAAGTCTGTGACTTG	C	**G**	CAAGTC---GACTTG	CTGGGACAGC**A**AAGTCTGTGACTTG
hp53-13082as	ACAGGGCAGGTCTTGGCCAGTTGGC	A	**C**	GCCA--TGGC	ACAGGGCAGGTCTTG**T**CCAGTTGGC
hp53-13088as	AGCTGCACAGGGCAGGTCTTGGCCA	G	**A**	TGGCCA	AGCTGCACAGGGCAGGTCT**A**GGCCA
hp53-13108as	GGGGTGTGGAATCAACCCACAGCTG	C	**T**	CAGCTG	GGGGTGTGGAATCAACCCA**A**AGCTG
hp53-13112as	GGCGGGGGTGTGGAATCAACCCACA	G	**C**	TGTGG--T-A--CCACA	GGCGGGGGT**A**TGGAATCAACCCACA
hp53-13113as	GGGCGGGGGTGTGGAATCAACCCAC	A	**A**	GTGG--T-A--CCAC	GGGCGGGGGTGT**A**GAATCAACCCAC
hp53-13115as	CCGGGCGGGGGTGTGGAATCAACCC	A	**C**	GGGT---------ACCC	CCGGGCGGG**A**GTGTGGAATCAACCC
hp53-13175as	CTCACAACCTCCGTCATGTGCTGTG	A	**A**	CACA-C----G-C----G-TGTG	CT**A**ACAACCTCCGTCATGTGCTGTG
hp53-13201as	AGCAGCGCTCATGGTGGGGGCAGCG	C	**C**	CGCT----------AGCG	AGCAGC**A**CTCATGGTGGGGGCAGCG
hp53-13215as	ACCATCGCTATCTGAGCAGCGCTCA	T	**G**	TGAGC---GCTCA	ACCATCGCTATC**A**GAGCAGCGCTCA
hp53-13266as	CCTGGGGACCCTGGGCAACCAGCCC	T	**A**	GGGC-----GCCC	CCTGGGGACCCT**A**GGCAACCAGCCC

Table 3. Modification of APEX primers by internal mismatch bases to avoid self-extension. The primers were used for resequencing of human TP53 gene.

The alternative options for redesign of APEX primers include modified bases with different T_m than of the natural bases and modified bases hybridising selectively with the natural bases (Pavel et al., 2000).

3.2.3. Template preparation, APEX reaction

Publications I, II

By our experience, both single- or double-stranded template DNA can be used in APEX reactions. Single-stranded DNA has advantages of missing complementary strand, competing for hybridisation sites with arrayed primers. The main disadvantage is that preparing sufficient amounts of single stranded template is laborious, expensive and one potentially looses 50% of information from the assay. Our aim was therefore to develop protocols suitable with the use of a double-stranded target, because it permits to identify changes from both strands in the same reaction and increasing reliability of the analysis. The important issue in APEX assay is the length of the target. PCR products need fragmentation before use in the APEX reaction (Figure 8). Fragmentation offers several advantages for APEX by reducing both the effects of secondary structures, reducing the melting temperature of target duplexes and permits the analysis of both strands simultaneously. Fragmentation also promotes greater mobility of the template and increases its effective concentration. We have replaced a fraction of the dTTP-s by dUTP-s in the amplification mix followed by Uracil N-Glycosylase (UNG) digestion for the template fragmentation. *In vivo*, UNG acts as one of the most efficient DNA repair enzymes, hydrolysing specifically the N-glycosylic bond connecting uracil to the deoxyribose sugar and generating abasic sites in DNA. *In vitro* this reaction can be used for asymmetric fragmentation of the template DNA (Cronin et al., 1996). UNG is highly specific to uracil bases in DNA and therefore the reaction can be controlled by dUTP incorporation during PCR as the substrate-limited step for fragmentation reaction. By changing the dUTP fraction from 15 to 30% of dTTP, one can vary the mean length of the template DNA fragments. None of other methods for DNA target fragmentation, like DNaseI treatment, restriction enzyme digestion, mechanical shearing and chemical degradation has worked more reproducibly in our hands.

APEX principle of single base primer extension reaction can only work if no deoxyribonucleotide triphosphates are carried over from the amplification mixture. A simple method to inactivate the dNTP leftover of from PCR is enzymatic digestion with shrimp alkaline phosphatase (sAP). This is performed in single tube reaction together with UNG treatment and both enzymes are thermally inactivated prior to the APEX reaction.

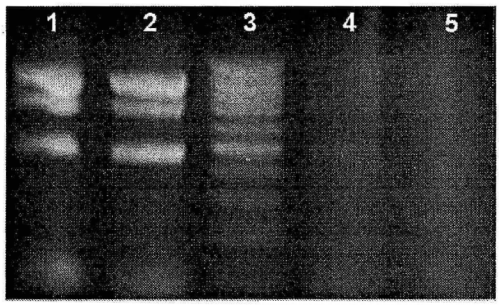


Figure 8. Gel electrophoresis of UNG fragmented DNA. Lanes 1 and 2 – 1/10 of triplex PCR reaction mixture loaded onto 1.5% agarose gel stained with ethidium bromide. Lane 3 – 100 bp DNA ladder (Low Range, Fermentas, Lithuania). Lanes 4 and 5 – the same products after UNG digestion and heat treatment (95°C 10 minutes).

Arrayed Primer Extension is a complex reaction consisting of target annealing to the oligonucleotide array and enzymatic primer extension reaction with fluorescent terminator nucleotides (ddNTP-s). Our goal was to make the APEX assay as robust as possible and therefore we have used single step reactions with simultaneous hybridisation and target dependent extension of arrayed primers. The signal to noise ratio of APEX depends on assay and set of primers, but is often 30/1, 40/1 or 100/1. The specific signals may even be free of detectable background noise. The primer extension reactions are routinely performed at constant temperature at 48°C to 58°C for 20 minutes (Kurg et al., 2000, Tõnisson et al., 2002). Incorporation of labelled terminators is a fast reaction, but hybridisation as an equilibrium process needs longer reaction time.

ThermoSequenase™ (Amersham Biosciences) and 20 to 50 picomoles of each fluorescent terminator per one APEX reaction have shown optimal efficiency and signal to noise ratio in APEX. Our dye labels are spectrally well separated (Table 2). The current set consists of (by spectral order): fluorescein, Cy3, Texas Red and Cy5. All but fluorescein are stable to bleaching. Due to bleaching of fluorescein, we have to use the antifade reagent (SlowFade® Light, Molecular Probes, Eugene, OR, USA) for imaging. An important factor with setting up a four-colour fluorescence assay was availability of the dye terminators. Fluorescein- and Texas Red labelled ddNTP-s are available for research use from PerkinElmer Life Sciences (Boston, MA, USA). Custom synthesis was the only option for implementation of Cy-labelled ddNTP-s.

3.2.4. Four-colour fluorescence detection

Publications I, II

Commercially available detectors are mainly using confocal scanning for array imaging. Confocal scanning readout has a good sensitivity and dynamic range, but is usually slower, compared to total internal reflection fluorescence (TIRF) imaging.

We have developed an automated TIRF based microarray imaging system for fast analysis of APEX reactions. Four lasers of 2-10 mW power, emitting on the wavelengths 488, 543, 594 and 633 nm are used to excite the fluorescence of the arrayed probes. The optical system consists of the light reflectors, rotating prism and cylindrical lenses transforming a beam of each laser into a homogeneously illuminated stripe, which is introduced to the microarray slide from one edge (Figure 9). Due to the total internal reflection on the surfaces of slide, the incoming beam can spread only inside the slide and therefore the intensity of light remains uniform along all its length. The evanescent light field excites the bound fluorophores residing near the surface of the slide. A gated 1.3 megapixel CCD camera, cooled to –25°C, is used to record the emitted fluorescence. The respective narrow-band interference filters fixed on the revolving wheel in front of the CCD camera depress a noise radiation of exciting laser scattered on the slide. Four images corresponding to four laser wavelengths are captured, one for each dye-labelled terminator nucleotide (Kurg et al., 2000).

The closest analogue of the system is a device designed for real-time detection of DNA hybridisation and melting on oligonucleotide microarrays. White light is used instead of lasers for the waveguide excitation and a CCD camera, operating 30 frames per second, collects the signals (Stimpson et al., 1995).

The imaging equipment should produce at least 10 x 10 pixels per each element if errors from a wrong pixel are expected to be in range of one percent. This means, the minimal pixel diameter of detector equipment must be at least 10 microns for a 100 μm centre-to-centre distance array. As most arrayers work in range of 200 μm between centres and we have mostly used 375μm centre-to-centre distance, we consider the spatial resolution of our detector with 20 μm pixel size acceptable.

A special software controls all parameters of the detection procedure switching shutters of the lasers, choosing respective interference filters and setting recording time at each wavelength. Because of precise spectral separation of exciting and emitting wavelengths, the system demonstrates a high signal-to-noise ratio. TIRF excitation gives also a low background. Evanescent field extends only 100 to 200 nm into the solution and only molecules in the close vicinity of the surface are excited and emit fluorescence. The number of detectable fluorescent labels can be adjusted according to assay requirements by

the addition of corresponding lasers and filters. The time required for four fluorescence channels is approximately five minutes enabling ultimate throughput of 10 to 12 samples per hour. The further development of TIRF-based microarray imaging system (Genorama™ QuattroImager) is now available from Asper Biotech Ltd. (Tartu, Estonia, http://www.asperbio.com).

3.2.5. Analysis software for APEX images

A software package was developed to convert the fluorescence signals into DNA sequence data. The analysis starts from addressing the signals on microarray images by overlay with a grid. The grid is set-up manually by selecting four corner marks on the images. The grid from the last set of images analysed is auto-saved enabling faster set-up with images from the same assay. Diameter-independent adaptive algorithms were developed for signal identification. The first parts of APEX image analysis are comparable with software packages designed for microarray-based gene expression analysis. The process of APEX data analysis is comparable to that of an automated four-colour DNA sequencer consisting of intensity comparisons from different fluorescent labels on the same band/dot. The strongest signal is the base called. If the next strongest signal from both strands has intensity level higher than 30 to 50% of the strongest signal, the position is called heterozygous. For mutation detection and resequencing analysis, the sequence is compared with a reference one (Figure 10) and diverging bases are indicated. All divergences and heterozygous positions may be visually verified and the genotypes edited if necessary. The actual extent of manual analysis depends on the type of analysis, being smaller in case of genotyping applications and larger for identification of somatic mutations. Two main windows are used in parallel. The main window for base-calling is designed to analyse all positions of the same sample. The second one allows to compare each position through a database of images giving an excellent overview of performance of the position and enabling automated genotyping via cluster analysis (Figure 10). Both software windows permit various types of data export including MS Excel tables and text tables for both quantitative and genotyping applications.

Genorama™ Genotyping Software package is now available from Asper Biotech.

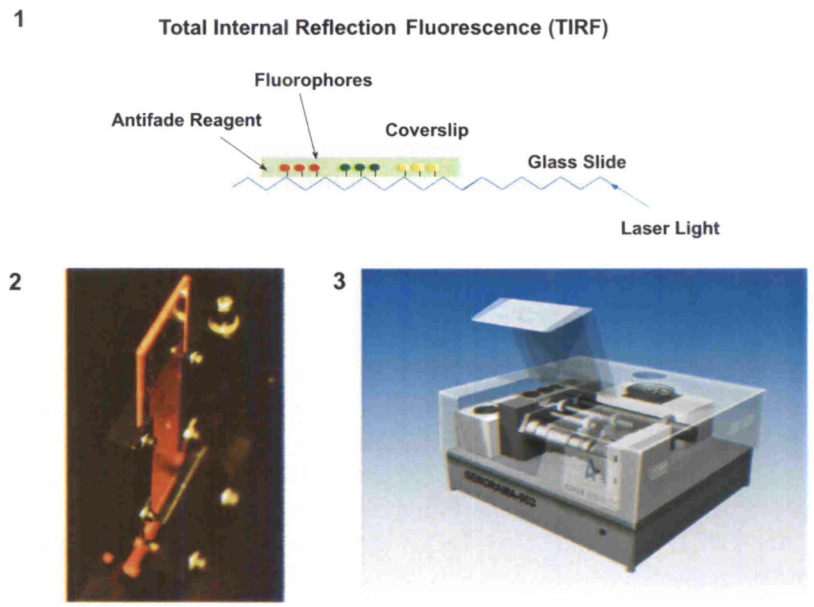


Figure 9. Four-color fluorescence imaging system based on TIRF excitation. 1. Principle of TIRF excitation mechanism. Fluorophores residing near glass surface are excited by evanescent light field. 2. The glass slide becomes a waveguide: the glass slide with red laser switched on. 3. Genorama™ microarray imaging system.

3.3. APEX technology applied for mutation detection and resequencing of disease-associated genes

3.3.1. Detection of common mutations in β-globin gene by APEX

Publication I

As the very first feasibility test of our four-colour mutation detection technology, APEX was applied for identification of 10 common point mutations in the human β-globin gene (Kattamis et al., 1990, Cao et al., 1997) causing β thalassemia in their homozygous state. The mutations studied were typical for Mediterranean region. Wild type DNA and nine DNA samples from patients and carriers were tested with the β-thalassemia APEX assay.

Standard microscope cover glasses (24 x 60 mm) activated by epoxy-silanisation (Southern et al., 1992, Lamture et al., 1994) were used for oligo-nucleotide attachment. An 1.4 kb fragment of the sample DNA was amplified by PCR. After amplification the PCR products were concentrated by ethanol precipitation. dNTP-s carried over from the PCR reaction were degraded enzymatically using shrimp Alkaline Phosphatase (sAP). Simultaneously, the amplicons were treated with UNG. Primer extension reaction conditions were first optimised with a wild-type DNA to achieve signals from all screened positions in the β-globin gene. Specificity of the assays was monitored by immobilised self-elongating marker primers. The marker primers were designed to form self-complementary homoduplexes, permitting for template-independent self-elongation.

The first "all-in-one" test for 10 mutations in beta globin gene was performed successfully. APEX technology including slide preparation, primer design, template preparation and TIRF-based four-colour fluorescence imaging system were evaluated. All primers on the array gave signals as expected from the β-globin gene wild-type sequence. Nine DNA samples from patients and carriers, each of which carries a different mutation and four wild-type DNA samples were correctly identified: from the analysed patient DNA samples, one was homozygous to IVS-I-110 mutation and eight were heterozygous to –87, Codon 5, Codon 6, IVS-I-1, IVS-I-6, IVS-II-1, Codon 39, IVS-II-745 mutations, respectively.

3.3.2. Combined platform for diagnostic screening — APEX plus extension of allele-specific primers (ASPEX)

Publication V

As a further evaluation step of β thalassemia screening, an assay was set up, enabling parallel detection of 17 mutations in β-globin and glycose 6 phosphate dehydrogenase (G6PD) genes commonly found in Mediterraneans. Both arrayed primer extension (APEX) (Shumaker et al., 1996) and allele-specific primer extension (ASPEX) were implemented the same microarray. ASO prmers with variant base at 3' end were used for ASPEX. The extension occurs only when ASO matches the target sequence. APEX and ASPEX were tested on the same slide, same reagent mixture, and templates. The study was focused on extensive validation of the combined APEX / ASPEX assay on Sardinian subjects who were referred to the haematological service and were first genotyped with classical methods.

117 DNA samples were selected to cover the 17 mutations for validation purposes. β-globin gene was amplified from human genomic DNA in two fragments and G6PD gene in five fragments. The PCR products were pooled and purified with Millipore Y30 columns, treated with UNG and sAP and applied in 20μl reaction mixture to the spotted microarray slide for 25 minutes at 58°C. Slides were imaged by Genorama™ four-colour imaging system equipped with Genorama™ analysis software (Asper Biotech).

Each mutation was identified with six different oligonucleotides, two for APEX and four for ASPEX.

The screening results showed to be very precise. Only nine out of 1989 genotypes, were called incorrectly, making the overall error rate as low as 0.45%. All the nine were false positive heterozygotes. A slight over-estimation of heterozygotes may actually benefit a screening program aimed to detect all the mutant alleles in a given population. All the samples can be reconfirmed with another method and identified correctly. In our sample set, two of seventeen mutant alleles, ß-IVSII-844 and G6PD854 were not analysed due to their rarity in Mediterranean area. In most cases, one or two out of six oligonucleotides were enough to determine correctly the genotype. 27 oligonucleotides showed 100% of correct extensions, and other 42 showed correct extensions in more than 90% of experiments. The sensitivity of Thalassochip was 100% for 14 mutations, with lowest value, 90.9 for β-globin codon 6. Specificity, was 100% for 13 mutations, with the lowest value of 98.5% for G6PD codon 563. Redundancy was shown to be favourable if the microarray technology is aimed to diagnostic precision both by a statistical point of view and for the correct interpretation of signals. The results were very promising for genetic screening. Same results would have been reached with much reduced number of oligonucleotides.

3.3.3. Detection of common mutations in BRCA1 gene by APEX

Publication III

Mutations in BRCA1 or BRCA2 gene increase risk to develop breast or ovarian cancer (Phelan et al., 1996). The familiar breast cancer is more polymorphic and contains more mitotic cells than a sporadic cancer. Metastatic breast cancer is also more common on patients with BRCA1 mutations (Xu et al., 1999). There is no clear discrimination between a rear polymorphism and missense mutation increasing risk of developing a cancer (Dunning et al., 1997, Durocher et al., 1996). The spectrum of mutations varies between populations. For example, 185delAG and 5382insC have been found in 90% of Ashkenazi Jewish families with high risk of breast and ovarian cancer (Fodor et al., 1998), but are less common in other populations.

We have studied Estonian families with high risk to breast and ovarian cancer for 42 mutations in the BRCA1 gene. All the subjects were also tested with SSCP-HA (single stranded DNA conformational polymorphism with heteroduplex analysis). Three BRCA1 gene mutations having positive correlation with breast cancer were found in three out of 28 analysed pedigrees.

3.3.4. Resequencing of TP53 tumour suppressor gene by APEX

Publication V

As one of the most recent advancements in APEX technology, assay was developed for the rapid and sensitive detection and identification of mutations in the TP53 gene. The TP53 microarray covers exons 2 to 9 together with flanking splice sites and introns 5 and 8 from both strands (total of 1218 bases) and regions over 98% of all mutations described so far in human cancer (Hernandez-Boussard et al., 1999). Extensive validation was performed for different steps and aspects of the analysis including redesign of APEX primers, suitability for large scale studies, sensitivity for the fraction of mutated DNA and usefulness for mutation detection in tumour samples.

The system allowed for sequencing of 97.5% of the arrayed TP53 gene from either sense or antisense strand, while 81% of the whole sequence was simultaneously analysed from both strands. The length of this simultaneous DNA sequence readout (1.2 Kb from both strands) outmatches the limits of the current standard for mutation detection – automated dideoxy sequencing.

Cluster analysis was implemented into the sequence analysis software facilitating identification of deviations from the wild-type reference signal pattern (Figure 10), indicative of mutations and was used for automated base-

calling. The comparison of a sample with the wild-type reference by the distance measure is comparable with the most advanced, current alternative to dideoxy sequencing today – GeneChip® p53 assay (Affymetrix, Inc.) (Ahrendt et al., 1999, Wen et al., 2000, Wikman et al., 2000). Just one APEX oligonucleotide per each sequenced base and the general low noise makes possible the fast visual inspection at positions where the software gives ambiguous result.

APEX sensitivity for the minimal identified percentage of mutated DNA was evaluated by titrating PCR products from Arg248Trp and Arg273His TP53 cDNA clones at different ratios. Both mutations were detected even if the sample contained as little as 5% of mutant DNA. In fact, 5% of the mutated DNA allowed identification from the analysis software window by eye. APEX sensitivity to detect deletions was titrated with del13-19 TP53 cDNA clone. The deletion was detected with sensitivity equal to a point mutation by analysing first base after the deletion (Head et al., 1997, Tõnisson et al., 2000b). The complementing detection of decreased signal intensities was technically more challenging and required at least 15% mutant sequence.

In order to evaluate performance of the TP53 APEX assay in large-scale studies, 100 normal DNA samples from the Estonian population were tested for common, single nucleotide polymorphisms (SNP) as well as for possible point mutations. A common SNP in exon 4 (Arg72Pro) was found with minor allele frequency of 0.26. We also detected two silent point mutations in codons 36 (CCG to CCA) and 139 (AAG to AAA) (Hernandez-Boussard et al., 1999) in two analysed samples. The results from 100 healthy individuals analysed were encouraging for applying APEX in large scale TP53 studies. The identified Arg72Pro polymorphism has recently been proposed to play a role in tumorigenesis (Storey et al., 1998, Tada et al., 2001).

As a second step of validation, blind test with 11 oesophageal carcinoma samples was performed. Previous study with TTGE plus manual or automated dideoxy sequencing of the extracted heteroduplex band had identified total of 12 point mutations and a two-basepair insertion in TP53 gene. Two silent, six missense, one splice site mutation and an insertion were concordantly identified with APEX and the standard methods. A missense mutation at the codon 290 was found by APEX instead of a silent point mutation as identified by TTGE plus dideoxy sequencing. One missense mutation not previously identified by TTGE plus sequencing was *de novo* identified by APEX. Two samples with missense mutations escaped identification by APEX. However, in these specimens, identification of the mutation was possible only by dideoxy sequencing of a PCR product generated from excised TTGE bands with abnormal migration patterns, indicating that mutant DNA was present only in a tiny fraction of the cells. In conclusion, an APEX-based sequencing test at the scale of the almost complete TP53 coding sequence was developed and evaluated. The evaluation test with tumour samples showed comparable performance with one of the most sensitive and also laborious combination of methods available – TTGE plus dideoxy sequencing.

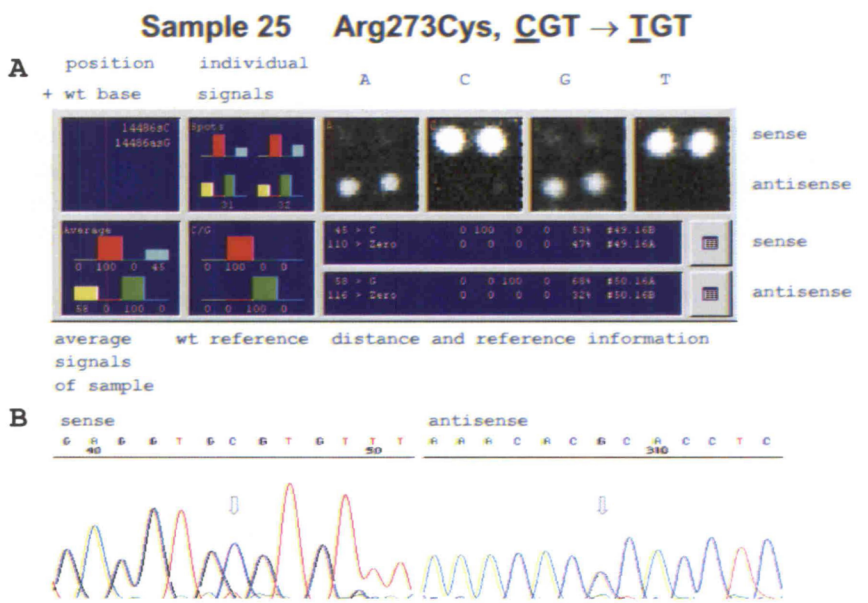


Figure 10. Missense mutation Arg273Cys CGT TGT, difficult to identify by automated dideoxy sequencing. A. 1st base of TP53 codon 273, analysed by APEX-based sequencing. The signals corresponding to T in sense strand and A in antisense strand indicate the presence of the mutation. B. Automated dideoxy sequencing images corresponding to the Arg273Cys mutation from both DNA strands. The indicated mutation-specific peaks are in the range of the background noise and may be easily missed by visual analysis.

4. CONCLUSIONS

An integrated technology platform for arrayed primer extension was developed and set up including testing the silanisation and binding methods of APEX primers, template preparation and APEX reaction conditions, design and building of a custom tailored, efficient and robust fluorescence detector and a useful software to analyse different mutations.

The APEX platform passed step-by step evaluation and validation process in the following assays and studies:

- Proof of principle study by simultaneous detection of 10 mutation sites in the human β-globin gene (Kurg et al., 2000). This was the first study with simultaneous four colour detection at both strands of the target DNA.
- Analysis of mutation sites in a large cancer susceptibility gene, BRCA1 (Tõnisson et al., 2000a).
- Blind study of β-globin mutation detection as a diagnostic screening test. Allele specific primer extension was combined with primer extension on ASO-s in the same assay to generate redundant information for genotypes (Gemignani et al., 2002).
- Proof of APEX for resequencing of TP53 tumour suppressor gene with analysis of DNA samples from 100 healthy individuals (Tõnisson et al., 2002).
- Evaluation of the TP53 gene resequencing assay in the blind test for mutation detection in oesophageal cancer samples (Tõnisson et al., 2002).

The studies showed that our highly scaleable mutation detection technology produces results, comparable with methods previously used even for technically challenging detection of somatic mutations and population screening studies.

Due to the reduced number of steps in wet lab and the possibility of performing automated analysis, APEX is more suitable for developing of high-throughput tests for screening and diagnostic applications.

I am confident that virtually any combination of mutations / loci / diseases; mutation-specific and scanning array set-ups can be successfully applied for different fields of genetic testing.

The APEX technology has a good potential for widespread use and is already commercialised by a spin-off company of Estonian Biocentre, Asper Biotech.

5. REFERENCES

Aas, T., Borresen, A. L., Geisler, S., Smith-Sorensen, B., Johnsen, H., Varhaug, J. E., Akslen, L. A. and Lonning, P. E. (1996) *Nat Med,* **2,** 811–4.

Afanassiev, V., Hanemann, V. and Wolfl, S. (2000) *Nucleic Acids Res,* **28,** E66.

Ahrendt, S. A., Halachmi, S., Chow, J. T., Wu, L., Halachmi, N., Yang, S. C., Wehage, S., Jen, J. and Sidransky, D. (1999) *Proc Natl Acad Sci U S A,* **96,** 7382–7.

Baker, S. J., Fearon, E. R., Nigro, J. M., Hamilton, S. R., Preisinger, A. C., Jessup, J. M., vanTuinen, P., Ledbetter, D. H., Barker, D. F., Nakamura, Y. and et al. (1989) *Science,* **244,** 217–21.

Baker, S. J., Preisinger, A. C., Jessup, J. M., Paraskeva, C., Markowitz, S., Willson, J. K., Hamilton, S. and Vogelstein, B. (1990) *Cancer Res,* **50,** 7717–22.

Beier, M. and Hoheisel, J. D. (2000) *Nucleic Acids Res,* **28,** E11.

Bell, P. A., Chaturvedi, S., Gelfand, C. A., Huang, C. Y., Kochersperger, M., Kopla, R., Modica, F., Pohl, M., Varde, S., Zhao, R., Zhao, X. and Boyce-Jacino, M. T. (2002) *Biotechniques,* **Suppl,** 70–2, 74, 76–7.

Broude, N. E., Sano, T., Smith, C. L. and Cantor, C. R. (1994) *Proc Natl Acad Sci USA,* **91,** 3072–6.

Broude, N. E., Woodward, K., Cavallo, R., Cantor, C. R. and Englert, D. (2001) *Nucleic Acids Res,* **29,** E92.

Brow, M. A., Oldenburg, M. C., Lyamichev, V., Heisler, L. M., Lyamicheva, N., Hall, J. G., Eagan, N. J., Olive, D. M., Smith, L. M., Fors, L. and Dahlberg, J. E. (1996) *J Clin Microbiol,* **34,** 3129–37.

Cabelguenne, A., Blons, H., de Waziers, I., Carnot, F., Houllier, A. M., Soussi, T., Brasnu, D., Beaune, P., Laccourreye, O. and Laurent-Puig, P. (2000) *J Clin Oncol,* **18,** 1465–73.

Cao, A., Saba, L., Galanello, R. and Rosatelli, M. C. (1997) *Jama,* **278,** 1273-7.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S. and Fodor, S. P. (1996) *Science,* **274,** 610–4.

Chen, X., Levine, L. and Kwok, P. Y. (1999) *Genome Res,* **9,** 492–8.

Cho, R. J., Mindrinos, M., Richards, D. R., Sapolsky, R. J., Anderson, M., Drenkard, E., Dewdney, J., Reuber, T. L., Stammers, M., Federspiel, N., Theologis, A., Yang, W. H., Hubbell, E., Au, M., Chung, E. Y., Lashkari, D., Lemieux, B., Dean, C., Lipshutz, R. J., Ausubel, F. M., Davis, R. W. and Oefner, P. J. (1999) *Nat Genet,* **23,** 203–7.

Cotton, R. G. H. (1997) *Mutation Detection,* Oxford University Press, New York.

Cronin, M. T., Fucini, R. V., Kim, S. M., Masino, R. S., Wespi, R. M. and Miyada, C. G. (1996) *Hum Mutat,* **7,** 244–55.

Cuppens, H. and Cassiman, J. J. (1995) *Eur J Hum Genet,* **3,** 235–45.

Dawson, E., Abecasis, G. R., Bumpstead, S., Chen, Y., Hunt, S., Beare, D. M., Pabial, J., Dibling, T., Tinsley, E., Kirby, S., Carter, D., Papaspyridonos, M., Livingstone, S., Ganske, R., Lõhmussaar, E., Zernant, J., Tõnisson, N., Remm, M., Magi, R., Puurand, T., Vilo, J., Kurg, A., Rice, K., Deloukas, P., Mott, R., Metspalu, A., Bentley, D. R., Cardon, L. R. and Dunham, I. (2002) *Nature,* **418,** 544–8.

Day, I. N., O'Dell, S. D., Cash, I. D., Humphries, S. E. and Weavind, G. P. (1995) *Nucleic Acids Res,* **23,** 2404–12.

Dipple, K. M. and McCabe, E. R. (2000) *Am J Hum Genet,* **66,** 1729–35.

Drmanac, R., Drmanac, S., Strezoska, Z., Paunesku, T., Labat, I., Zeremski, M., Snoddy, J., Funkhouser, W. K., Koop, B., Hood, L. and et al. (1993) *Science,* **260,** 1649–52.

Drmanac, R., Labat, I., Brukner, I. and Crkvenjakov, R. (1989) *Genomics,* **4,** 114–28.

Drmanac, S., Kita, D., Labat, I., Hauser, B., Schmidt, C., Burczak, J. D. and Drmanac, R. (1998) *Nat Biotechnol,* **16,** 54–8.

Dunning, A. M., Chiano, M., Smith, N. R., Dearden, J., Gore, M., Oakes, S., Wilson, C., Stratton, M., Peto, J., Easton, D., Clayton, D. and Ponder, B. A. (1997) *Hum Mol Genet,* **6,** 285–9.

Durocher, F., Shattuck-Eidens, D., McClure, M., Labrie, F., Skolnick, M. H., Goldgar, D. E. and Simard, J. (1996) *Hum Mol Genet,* **5,** 835–42.

Edman, C. F., Raymond, D. E., Wu, D. J., Tu, E., Sosnowski, R. G., Butler, W. F., Nerenberg, M. and Heller, M. J. (1997) *Nucleic Acids Res,* **25,** 4907–14.

Eliyahu, D., Michalovitz, D., Eliyahu, S., Pinhasi-Kimhi, O. and Oren, M. (1989) *Proc Natl Acad Sci U S A,* **86,** 8763–7.

Eliyahu, D., Raz, A., Gruss, P., Givol, D. and Oren, M. (1984) *Nature,* **312,** 646–9.

Fan, J. B., Chen, X., Halushka, M. K., Berno, A., Huang, X., Ryder, T., Lipshutz, R. J., Lockhart, D. J. and Chakravarti, A. (2000) *Genome Res,* **10,** 853–60.

Finlay, C. A., Hinds, P. W. and Levine, A. J. (1989) *Cell,* **57,** 1083–93.

Fodor, F. H., Weston, A., Bleiweiss, I. J., McCurdy, L. D., Walsh, M. M., Tartter, P. I., Brower, S. T. and Eng, C. M. (1998) *Am J Hum Genet,* **63,** 45–51.

Fodor, S. P., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T. and Solas, D. (1991) *Science,* **251,** 767–73.

Fotin, A. V., Drobyshev, A. L., Proudnikov, D. Y., Perov, A. N. and Mirzabekov, A. D. (1998) *Nucleic Acids Res,* **26,** 1515–21.

Garcia, C. A., Ahmadian, A., Gharizadeh, B., Lundeberg, J., Ronaghi, M. and Nyren, P. (2000) *Gene,* **253,** 249–57.

Gemignani, F., Perra, C., Landi, S., Canzian, F., Kurg, A., Tonisson, N., Galanello, R., Cao, A., Metspalu, A. and Romeo, G. (2002) *Clin Chem,* **48,** 2051–4.

Gerry, N. P., Witowski, N. E., Day, J., Hammer, R. P., Barany, G. and Barany, F. (1999) *J Mol Biol,* **292,** 251–62.

Gilles, P. N., Wu, D. J., Foster, C. B., Dillon, P. J. and Chanock, S. J. (1999) *Nat Biotechnol,* **17,** 365–70.

Greenblatt, M. S., Bennett, W. P., Hollstein, M. and Harris, C. C. (1994) *Cancer Res,* **54,** 4855–78.

Gross, E., Kiechle, M. and Arnold, N. (2001) *J Biochem Biophys Methods,* **47,** 73-81.

Gunderson, K. L., Huang, X. C., Morris, M. S., Lipshutz, R. J., Lockhart, D. J. and Chee, M. S. (1998) *Genome Res,* **8,** 1142–53.

Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R. and Smith, L. M. (1994) *Nucleic Acids Res,* **22,** 5456–65.

Hacia, J. G. (1999) *Nat Genet,* **21,** 42–7.

Hacia, J. G., Brody, L. C., Chee, M. S., Fodor, S. P. and Collins, F. S. (1996) *Nat Genet,* **14,** 441–7.

Hainaut, P., Hernandez, T., Robinson, A., Rodriguez-Tome, P., Flores, T., Hollstein, M., Harris, C. C. and Montesano, R. (1998) *Nucleic Acids Res,* **26,** 205–13.

Hainaut, P., Olivier, M. and Pfeifer, G. P. (2001) *Mutagenesis,* **16,** 551–3, 555–6.

Hainaut, P. and Pfeifer, G. P. (2001) *Carcinogenesis,* **22,** 367–74.

Hamelin, R., Jego, N., Laurent-Puig, P., Vidaud, M. and Thomas, G. (1993) *Oncogene,* **8,** 2213–20.

Head, S. R., Rogers, Y. H., Parikh, K., Lan, G., Anderson, S., Goelet, P. and Boyce-Jacino, M. T. (1997) *Nucleic Acids Res,* **25,** 5065–71.

Hernandez-Boussard, T., Rodriguez-Tome, P., Montesano, R. and Hainaut, P. (1999) *Hum Mutat,* **14,** 1–8.

Holland, P. M., Abramson, R. D., Watson, R. and Gelfand, D. H. (1991) *Proc Natl Acad Sci U S A,* **88,** 7276–80.

Hollstein, M., Hergenhahn, M., Yang, Q., Bartsch, H., Wang, Z. Q. and Hainaut, P. (1999) *Mutat Res,* **431,** 199–209.

Howell, W. M., Jobs, M. and Brookes, A. J. (2002) *Genome Res,* **12,** 1401–7.

Howell, W. M., Jobs, M., Gyllensten, U. and Brookes, A. J. (1999) *Nat Biotechnol,* **17,** 87–8.

Hughes, T. R., Mao, M., Jones, A. R., Burchard, J., Marton, M. J., Shannon, K. W., Lefkowitz, S. M., Ziman, M., Schelter, J. M., Meyer, M. R., Kobayashi, S., Davis, C., Dai, H., He, Y. D., Stephaniants, S. B., Cavet, G., Walker, W. L., West, A., Coffey, E., Shoemaker, D. D., Stoughton, R., Blanchard, A. P., Friend, S. H. and Linsley, P. S. (2001) *Nat Biotechnol,* **19,** 342–7.

Hussain, S. P. and Harris, C. C. (1999) *Mutat Res,* **428,** 23–32.

Jenkins, J. R., Rudge, K. and Currie, G. A. (1984) *Nature,* **312,** 651–4.

Joos, B., Kuster, H. and Cone, R. (1997) *Anal Biochem,* **247,** 96–101.

Jurinke, C., van den Boom, D., Cantor, C. R. and Koster, H. (2002) *Methods Mol Biol,* **187,** 179–92.

Kattamis, C., Hu, H., Cheng, G., Reese, A. L., Gonzalez-Redondo, J. M., Kutlar, A., Kutlar, F. and Huisman, T. H. (1990) *Br J Haematol,* **74,** 342–6.

Kirk, B. W., Feinsod, M., Favis, R., Kliman, R. M. and Barany, F. (2002) *Nucleic Acids Res,* **30,** 3295–311.

Kozal, M. J., Shah, N., Shen, N., Yang, R., Fucini, R., Merigan, T. C., Richman, D. D., Morris, D., Hubbell, E., Chee, M. and Gingeras, T. R. (1996) *Nat Med,* **2,** 753–9.

Kumar, A., Larsson, O., Parodi, D. and Liang, Z. (2000) *Nucleic Acids Res,* **28,** E71.

Kurg, A., Tõnisson, N., Georgiou, I., Shumaker, J., Tollett, J. and Metspalu, A. (2000) *Genet Test,* **4,** 1–7.

Kwok, P. Y. (2002) *Hum Mutat,* **19,** 315–23.

Lamture, J. B., Beattie, K. L., Burke, B. E., Eggers, M. D., Ehrlich, D. J., Fowler, R., Hollis, M. A., Kosicki, B. B., Reich, R. K., Smith, S. R. and et al. (1994) *Nucleic Acids Res,* **22,** 2121–5.

Landegren, U. (Ed.) (1996) *Laboratory Protocols for Mutation Detection,* Oxford University Press, New York.

Landegren, U., Kaiser, R., Sanders, J. and Hood, L. (1988) *Science,* **241,** 1077–80.

Lindroos, K., Sigurdsson, S., Johansson, K., Ronnblom, L. and Syvanen, A. C. (2002) *Nucleic Acids Res,* **30,** e70.

Linzer, D. I. and Levine, A. J. (1979) *Cell,* **17,** 43–52.

Lipshutz, R. J., Fodor, S. P., Gingeras, T. R. and Lockhart, D. J. (1999) *Nat Genet,* **21,** 20–4.

Lipshutz, R. J., Morris, D., Chee, M., Hubbell, E., Kozal, M. J., Shah, N., Shen, N., Yang, R. and Fodor, S. P. (1995) *Biotechniques,* **19,** 442–7.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E. L. (1996) *Nat Biotechnol,* **14,** 1675–80.

Malkin, D. (1994) *Annu Rev Genet,* **28,** 443–65.

Malkin, D., Li, F. P., Strong, L. C., Fraumeni, J. F., Jr., Nelson, C. E., Kim, D. H., Kassel, J., Gryka, M. A., Bischoff, F. Z., Tainsky, M. A. and et al. (1990) *Science,* **250,** 1233–8.

Martinez di Montemuros, F., Dotti, C., Tavazzi, D., Fiorelli, G. and Cappellini, M. D. (1997) *Haematologica,* **82,** 440–5.

Maskos, U. and Southern, E. M. (1992a) *Nucleic Acids Res,* **20,** 1679–84.

Maskos, U. and Southern, E. M. (1992b) *Nucleic Acids Res,* **20,** 1675–8.

Maskos, U. and Southern, E. M. (1993a) *Nucleic Acids Res,* **21,** 2269–70.

Maskos, U. and Southern, E. M. (1993b) *Nucleic Acids Res,* **21,** 4663–9.

Mazars, R., Pujol, P., Maudelonde, T., Jeanteur, P. and Theillet, C. (1991) *Oncogene,* **6,** 1685–90.

McGall, G., Barone, A. D., Diggellmann, M., Fodor, S., Gentalen, E. and Ngo, N. (1997) *J Am Soc Clin Chem,* **119,** 5081–5090.

McGall, G., Labadie, J., Brock, P., Wallraff, G., Nguyen, T. and Hinsberg, W. (1996) *Proc Natl Acad Sci U S A,* **93,** 13555–60.

Metspalu, A., Berik, J. and Kurg, A. (2000) Asper Biotech Ltd., Estonia, PCT/EE00/00001, WO0063677.

Metspalu, A. a. J. M. S. (1999) In *Microsystem technology: a powerful tool for biomedical studies*(Ed, J.M. Köhler, T. M. a. H. P. S.) Birkhäuser Verlag, Basel, pp. 371–397.

Michalovitz, D., Halevy, O. and Oren, M. (1991) *J Cell Biochem,* **45,** 22–9.

Mullis, K. B. and Faloona, F. A. (1987) *Methods Enzymol,* **155,** 335–50.

Nikiforov, T. T., Rendle, R. B., Goelet, P., Rogers, Y. H., Kotewicz, M. L., Anderson, S., Trainor, G. L. and Knapp, M. R. (1994a) *Nucleic Acids Res,* **22,** 4167–75.

Nikiforov, T. T., Rendle, R. B., Kotewicz, M. L. and Rogers, Y. H. (1994b) *PCR Methods Appl,* **3,** 285–91.

Nordhoff, E., Luebbert, C., Thiele, G., Heiser, V. and Lehrach, H. (2000) *Nucleic Acids Res,* **28,** E86.

Oliphant, A., Barker, D. L., Stuelpnagel, J. R. and Chee, M. S. (2002) *Biotechniques,* **Suppl,** 56–8, 60–1.

Olivier, M., Chuang, L. M., Chang, M. S., Chen, Y. T., Pei, D., Ranade, K., de Witte, A., Allen, J., Tran, N., Curb, D., Pratt, R., Neefs, H., de Arruda Indig, M., Law, S., Neri, B., Wang, L. and Cox, D. R. (2002a) *Nucleic Acids Res,* **30,** e53.

Olivier, M., Eeles, R., Hollstein, M., Khan, M. A., Harris, C. C. and Hainaut, P. (2002b) *Hum Mutat,* **19,** 607–14.

O'Meara, D., Ahmadian, A., Odeberg, J. and Lundeberg, J. (2002) *Nucleic Acids Res,* **30,** e75.

Parinov, S., Barsky, V., Yershov, G., Kirillov, E., Timofeev, E., Belgovskiy, A. and Mirzabekov, A. (1996) *Nucleic Acids Res,* **24,** 2998–3004.

Pastinen, T., Kurg, A., Metspalu, A., Peltonen, L. and Syvanen, A. C. (1997) *Genome Res,* **7,** 606–14.

Pastinen, T., Partanen, J. and Syvanen, A. C. (1996) *Clin Chem,* **42,** 1391–7.

Pastinen, T., Perola, M., Niini, P., Terwilliger, J., Salomaa, V., Vartiainen, E., Peltonen, L. and Syvanen, A. (1998) *Hum Mol Genet,* **7,** 1453–62.

Pastinen, T., Raitio, M., Lindroos, K., Tainola, P., Peltonen, L. and Syvanen, A. C. (2000) *Genome Res,* **10,** 1031–42.

Patil, N., Berno, A. J., Hinds, D. A., Barrett, W. A., Doshi, J. M., Hacker, C. R., Kautzer, C. R., Lee, D. H., Marjoribanks, C., McDonough, D. P., Nguyen, B. T., Norris, M. C., Sheehan, J. B., Shen, N., Stern, D., Stokowski, R. P., Thomas, D. J., Trulson, M. O., Vyas, K. R., Frazer, K. A., Fodor, S. P. and Cox, D. R. (2001) *Science,* **294,** 1719–23.

Pavel, H., Tõnisson, N. and Zernant, J. (2000) Asper Biotech Ltd., Estonia, PCT/EE00/00003.

Peltonen, L. and McKusick, V. A. (2001) *Science,* **291,** 1224–1229.

Phelan, C. M., Rebbeck, T. R., Weber, B. L., Devilee, P., Ruttledge, M. H., Lynch, H. T., Lenoir, G. M., Stratton, M. R., Easton, D. F., Ponder, B. A., Cannon-Albright, L., Larsson, C., Goldgar, D. E. and Narod, S. A. (1996) *Nat Genet,* **12,** 309–11.

Phimister, B. (Ed.) (1999) *The Chipping Forecast, Nat Genet.* **Suppl, 21,** 1–60.

Podyminogin, M. A., Lukhtanov, E. A. and Reed, M. W. (2001) *Nucleic Acids Res,* **29,** 5090–8.

Rehman, F. N., Audeh, M., Abrams, E. S., Hammond, P. W., Kenney, M. and Boles, T. C. (1999) *Nucleic Acids Res,* **27,** 649–55.

Revel, M. (1995) In *International Bioethics Committee of UNESCO (IBC); 3rd session,* Vol. 1 UNESCO, Paris, pp. 9–37.

Rosatelli, M. C., Dozy, A., Faa, V., Meloni, A., Sardu, R., Saba, L., Kan, Y. W. and Cao, A. (1992) *Am J Hum Genet,* **50,** 422–6.

Ryan, D., Nuccie, B. and Arvan, D. (1999) *Mol Diagn,* **4,** 135–44.

Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. (1989) *Proc Natl Acad Sci U S A,* **86,** 6230–4.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc Natl Acad Sci USA,* **74,** 5463–7.

Sauer, S., Gelfand, D. H., Boussicault, F., Bauer, K., Reichert, F. and Gut, I. G. (2002) *Nucleic Acids Res,* **30,** e22.

Sauer, S., Lechner, D., Berlin, K., Lehrach, H., Escary, J. L., Fox, N. and Gut, I. G. (2000a) *Nucleic Acids Res,* **28,** E13.

Sauer, S., Lechner, D., Berlin, K., Plancon, C., Heuermann, A., Lehrach, H. and Gut, I. G. (2000b) *Nucleic Acids Res,* **28,** E100.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O. and Davis, R. W. (1996) *Proc Natl Acad Sci U S A,* **93,** 10614–9.

Scriver, C. R. and Waters, P. J. (1999) *Trends Genet,* **15,** 267–72.

Shalon, D., Smith, S. J. and Brown, P. O. (1996) *Genome Res,* **6,** 639–45.

Shchepinov, M. S., Case-Green, S. C. and Southern, E. M. (1997) *Nucleic Acids Res,* **25,** 1155–61.

Shumaker, J. M., Metspalu, A. and Caskey, C. T. (1996) *Hum Mutat,* **7,** 346–54.

Singh-Gasson, S., Green, R. D., Yue, Y., Nelson, C., Blattner, F., Sussman, M. R. and Cerrina, F. (1999) *Nat Biotechnol,* **17,** 974–8.

Sosnowski, R. G., Tu, E., Butler, W. F., O'Connell, J. P. and Heller, M. J. (1997) *Proc Natl Acad Sci USA,* **94,** 1119–23.

Soussi, T., Dehouche, K. and Beroud, C. (2000) *Hum Mutat,* **15,** 105–13.

Southern, E., Mir, K. and Shchepinov, M. (1999) *Nat Genet,* **21,** 5–9.

Southern, E. M., Case-Green, S. C., Elder, J. K., Johnson, M., Mir, K. U., Wang, L. and Williams, J. C. (1994) *Nucleic Acids Res,* **22,** 1368–73.

Southern, E. M., Maskos, U. and Elder, J. K. (1992) *Genomics,* **13,** 1008–17.
Steemers, F. J., Ferguson, J. A. and Walt, D. R. (2000) *Nat Biotechnol,* **18,** 91–4.
Stimpson, D. I., Hoijer, J. V., Hsieh, W. T., Jou, C., Gordon, J., Theriault, T., Gamble, R. and Baldeschwieler, J. D. (1995) *Proc Natl Acad Sci USA,* **92,** 6379–83.
Storey, A., Thomas, M., Kalita, A., Harwood, C., Gardiol, D., Mantovani, F., Breuer, J., Leigh, I. M., Matlashewski, G. and Banks, L. (1998) *Nature,* **393,** 229–34.
Strezoska, Z., Paunesku, T., Radosavljevic, D., Labat, I., Drmanac, R. and Crkvenjakov, R. (1991) *Proc Natl Acad Sci U S A,* **88,** 10089–93.
Syvänen, A. C. (2001) *Nat Rev Genet,* **2,** 930–42.
Syvänen, A. C., Aalto-Setala, K., Harju, L., Kontula, K. and Soderlund, H. (1990) *Genomics,* **8,** 684–92.
Zammatteo, N., Jeanmart, L., Hamels, S., Courtois, S., Louette, P., Hevesi, L. and Remacle, J. (2000) *Anal Biochem,* **280,** 143–50.
Tabor, S. and Richardson, C. C. (1987) *Proc Natl Acad Sci U S A,* **84,** 4767–71.
Tabor, S. and Richardson, C. C. (1995) *Proc Natl Acad Sci U S A,* **92,** 6339–43.
Tada, M., Furuuchi, K., Kaneda, M., Matsumoto, J., Takahashi, M., Hirai, A., Mitsumoto, Y., Iggo, R. D. and Moriuchi, T. (2001) *Carcinogenesis,* **22,** 515–7.
Takahashi, T., Nau, M. M., Chiba, I., Birrer, M. J., Rosenberg, R. K., Vinocour, M., Levitt, M., Pass, H., Gazdar, A. F. and Minna, J. D. (1989) *Science,* **246,** 491–4.
Taniere, P., Martel-Planche, G., Maurici, D., Lombard-Bohas, C., Scoazec, J. Y., Montesano, R., Berger, F. and Hainaut, P. (2001) *Am J Pathol,* **158,** 33–40.
Taylor, J. D., Briley, D., Nguyen, Q., Long, K., Iannone, M. A., Li, M. S., Ye, F., Afshari, A., Lai, E., Wagner, M., Chen, J. and Weiner, M. P. (2001) *Biotechniques,* **30,** 661–6, 668–9.
Thorlacius, S., Borresen, A. L. and Eyfjord, J. E. (1993) *Cancer Res,* **53,** 1637–41.
Tõnisson, N., Kurg, A., Kaasik, K., Lohmussaar, E. and Metspalu, A. (2000a) *Clin Chem Lab Med,* **38,** 165–70.
Tõnisson, N., Kurg, A., Lõhmussaar, E. and Metspalu, A. (2000b) In *Microarray Biochip Technology* (Ed, Schena, M.) Eaton Publishing, Natick, MA, pp. 247–63.
Tõnisson, N., Zernant, J., Kurg, A., Pavel, H., Slavin, G., Roomere, H., Meiel, A., Hainaut, P. and Metspalu, A. (2002) *Proc Natl Acad Sci U S A,* **99,** 5503–8.
Underhill, P. A., Jin, L., Lin, A. A., Mehdi, S. Q., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L. and Oefner, P. J. (1997) *Genome Res,* **7,** 996–1005.
Wallace-Brodeur, R. R. and Lowe, S. W. (1999) *Cell Mol Life Sci,* **55,** 64–75.
Weatherall, D. J. and Clegg, J. B. (2001) *Bull World Health Organ,* **79,** 704–12.
Wen, W. H., Bernstein, L., Lescallett, J., Beazer-Barclay, Y., Sullivan-Halley, J., White, M. and Press, M. F. (2000) *Cancer Res,* **60,** 2716–22.
Wen, W. H., Reles, A., Runnebaum, I. B., Sullivan-Halley, J., Bernstein, L., Jones, L. A., Felix, J. C., Kreienberg, R., el-Naggar, A. and Press, M. F. (1999) *Int J Gynecol Pathol,* **18,** 29–41.
Wikman, F. P., Lu, M. L., Thykjaer, T., Olesen, S. H., Andersen, L. D., Cordon-Cardo, C. and Orntoft, T. F. (2000) *Clin Chem,* **46,** 1555–61.
Wilson, J. M. G. and Jungner, G. (1968) *Principles and Practises of Screening for Disease,* World Health Organisation, Geneva.
Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M. and Bos, J. L. (1988) *N Engl J Med,* **319,** 525–32.
Vogelstein, B., Lane, D. and Levine, A. J. (2000) *Nature,* **408,** 307–10.

Woods, D. B. and Vousden, K. H. (2001) *Exp Cell Res,* **264,** 56–66.
Xu, X., Wagner, K. U., Larson, D., Weaver, Z., Li, C., Ried, T., Hennighausen, L., Wynshaw-Boris, A. and Deng, C. X. (1999) *Nat Genet,* **22,** 37–43.

MUTATSIOONIDE UURIMINE KASUTADES PRAIMEREKSTENSIOONI OLIGONUKLEOTIIDKIIPIDEL

Kokkuvõte

Käesolev töö käsitleb tehnoloogilisi probleeme, mis on seotud erinevate haiguste geneetilise tausta uurimisega. Geeniuuringuid kasutatakse järjest laialdasemalt nii teadustöös kui meditsiinilises diagnostikas ja sõeluuringutes. Igasugune geeniuuring eeldab sobiva detektsioonimeetodi valikut, kusjuures traditsioonilised meetodid on muutunud parimal juhul suboptimaalseteks nii täpsuse, töömahukuse kui läbitavate etappide rohkuse tõttu.

Töö eesmärgiks oli välja arendada geneetilisteks uuringuteks sobiv kaasaegne tehnoloogia, kasutades praimerekstensiooni oligonukleotiidkiibil. Arendustöö hõlmas nii uuritava materjali ettevalmistamist, kiipide valmistamist, praimerite disaini, praimerekstensioonitingimuste optimeerimist kui neljavärvifluorestsentsdetektsiooni ja analüüsitarkvara väljatöötamist. Tehnoloogiat katsetati edukalt erineva skaala ja täpsusnõuetega mudelsüsteemides, sealhulgas tehnoloogiliselt keerukas somaatiliste mutatsioonide uurimises TP53 geeni näitel. Beeta-talasseemia sõeltesti katsetes saavutati 100% lähedane keskmine tundlikus ja spetsiifilisus.

Töö tulemusena võib väita, et on välja arendatud ja testitud erinevateks rakendusteks ja erinevatesse analüüsiskaaladesse sobiv tehnoloogiline platvorm, mis lubab tuvastada mutatsioone kui teostada DNA järjestuse analüüsi. Arendatud tehnoloogia alusel on esitatud kaks rahvusvahelist patenditaotlust.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof. Andres Metspalu for the interesting topic, continuous support and energising optimism, even if the initial results weren't always perfect;

Ants Kurg, Associate Professor of the Chair of Biotechnology, for excellent guidance in both theoretical and practical questions, for being always available and for critical reading of the manuscript;

Mrs.Krista Liiv for being most helpful with all kind of matters, whenever needed;

My good colleagues in Chair of Biotechnology, IMCB and Asper Biotech for help and support, especially Jana Zernant, M.Sc., for the best possible practical contribution and also for being a good friend;

Mr.Georg Slavin for lessons in computer algorithming and for realising the ideas of mine and other colleagues in a number of software packages;

Hendrik Pavel, M.Sc. for inspiring discussions in Asper;

Indrek Valvas and Viljo Soo for technical assistance and spotting of all the microarray slides;

Dr. John M.Shumaker and Dr.Jeff Tollet for original ideas used for prototype version of the four-channel microarray imaging system; Mr.Jevgeni Berik and Mr.Jevgeni Slivinski for constructing the advanced prototype and the imaging systems now in use;

Elin Lõhmussaar, M.Sc. and Krista Kaasik, M.Sc. for contributing to publications II and III;

Dr.Pierre Hainaut for expert advices and help in development of TP53 gene resequencing assay;

Aune Meiel and Hanno Roomere, M.Sc. for the practical help during evaluation process of TP53 resequencing assay;

Dr.Arnold Kristjuhan and Viljar Jaks, M.Sc. for materials used in the sensitivity test of TP53 resequencing assay;

Tõnis Org for help with artwork;

My beloved wife Mailis for the complete understanding and unconditional support throughout my studies;

Professors Aavo-Valdur Mikelsaar and Jüri Kärner for being my first supervisors;

Doctors Andrei Syritsa and Lev Levkov for teaching the basic and universal skills of laboratory work when I was working in Women's Clinic of Tartu University.

PUBLICATIONS

I

Kurg A, **Tõnisson N**, Tollett J, Georgiou I, Shumaker J, Metspalu A. Arrayed Primer Extension: Solid phase four-color DNA resequencing and mutation detection technology. *Genetic Testing 2000; 4(1):1–7*

GENETIC TESTING
Volume 4, Number 1, 2000
Mary Ann Liebert, Inc.

Arrayed Primer Extension: Solid-Phase Four-Color DNA Resequencing and Mutation Detection Technology

ANTS KURG,[1] NEEME TÕNISSON,[1,4] IOANNIS GEORGIOU,[2] JOHN SHUMAKER,[3] JEFF TOLLETT,[3] and ANDRES METSPALU[1,4]

ABSTRACT

The technology and application of arrayed primer extension (APEX) is presented. We describe an integrated system with DNA chip and template preparation, multiplex primer extension on the array, fluorescence imaging, and data analysis. The method is based upon an array of oligonucleotides, immobilized via the 5′ end on a glass surface. A patient DNA is amplified by PCR, digested enzymatically, and annealed to the immobilized primers, which promote sites for template-dependent DNA polymerase extension reactions using four unique fluorescently labeled dideoxy nucleotides. A mutation is detected by a change in the color code of the primer sites. The technology was applied to the analysis of 10 common β-thalassemia mutations. Nine patient DNA samples, each of which carries a different mutation, and four wild-type DNA samples were correctly identified. The signal-to-noise ratio of this technology is, on the average, 40:1, which enables the identification of heterozygous mutations with a high confidence level. The APEX method can be applied to any DNA target for efficient analysis of mutations and polymorphisms.

INTRODUCTION

GENETICS AND MOLECULAR MEDICINE have an expanding need for rapid genotyping, mutation analysis, and DNA resequencing technologies that have a clear potential for miniaturization, parallelization, and automation and enable high throughput and ability to identify changes precisely in the patient DNA. Conventional methods for mutation detection, such as single strand conformation polymorphism (SSCP), denaturing gradient gel electrophoresis (DGGE), chemical cleavage, or direct sequencing are time and labor intensive (Cotton *et al.*, 1998) with little parallelization. A promising solution for this technological need is the use of oligonucleotide arrays using nucleic acid hybridization (Chee *et al.*, 1996; Cronin *et al.*, 1996; Hacia *et al.*, 1996) or hybridization coupled with an enzyme-mediated reaction, either by primer extension (Shumaker *et al.*, 1996; Head *et al.*, 1997; Pastinen *et al.*, 1997) or ligation (Landegren *et al.*, 1998). The most developed approach today, hybridization of labeled target to high-density oligonucleotide microarrays (*e.g.*, Affymetrix GeneChip™ arrays), is a revolutionary method for DNA sequence analysis. Feasibility studies of this approach are promising and the first results are impressive. However, the high complexity of the assays due to the large number of oligonucleotide probes per target sequence base, sensitivity to hybridization conditions, complicated data analysis, and high cost are driving several research groups to look for alternative technologies.

Here we present Arrayed Primer EXtension (APEX) technology as an alternative to array-based DNA sequence analysis by hybridization. We describe an integrated system with chip and template preparation, multiplex primer extension on the array, fluorescence imaging, and data analysis. The method is based upon a two-dimensional (2D) array of oligonucleotides, immobilized via the 5′ terminal amino group onto an epoxy-silanized glass support. This method can be viewed as dye-terminator sequencing of DNA, but instead of using one primer and analyzing hundreds of extension products in polyacrylamide gel electrophoresis (PAGE), we can use hundreds to thousands of primers that are spatially separated and extend each by only one dye-labeled nucleotide. The single-tube sample preparation protocol consists of PCR amplification followed by a DNA fragmentation reaction. After hybridization of the

[1]Institute of Molecular and Cell Biology, Tartu Childrens Hospital, University of Tartu, Estonian Biocentre, Tartu 51010, Estonia.
[2]Medical School, University of Ioannina, Ioannina 45500, Greece.
[3]Baylor College of Medicine, Houston, TX, 77030.
[4]Asper Ltd., Tartu 51014, Estonia.

target DNA to the array, target-dependent oligonucleotide extension by a DNA polymerase is used to incorporate fluorescently labeled dideoxy terminators onto the primers (Head *et al.,* 1997).

We have developed a total internal reflection fluorescence (TIRF) excitation mechanism combined with a charge coupled device (CCD) detector for high-throughput image acquisition. The signals from the spectrally separated dyes are resolved to better than 2% cross-talk with dual selection by laser excitation and bandpass filtering of the emitted fluorescence. Imaging is followed by a software analysis to convert the fluorescence information into sequence data. The APEX method combines the advantage of both high information content of the oligonucleotide array and fidelity of the enzymatic primer extension reaction. The enzyme acts as a biological proofreading mechanism, discriminating against 3′ end mismatches. Moreover, by using primer extension, each position on the array identifies a unique base of the sequence as the result of the direct competition of four dye-terminators for the same spot, thus reducing the array complexity by at least a factor of four. Signal-to-noise ratio is improved by the added fidelity of the polymerase. The elimination of intensity comparisons across multiple spots, as is the case for hybridization assays, makes the analysis more robust. Two unique oligonucleotide primers probe the sense and antisense target strands at the same base location. APEX can be used to analyze known point mutations, deletions, and insertions and can identify the presence of unknown polymorphisms.

We applied APEX for the identification of 10 common point mutations in the human β-globin gene (Cao *et al.,* 1997; Huisman *et al.,* 1998), causing β-thalassemia in their homozygous state. β-Thalassemia is a very common autosomal recessive disorder in populations of Mediterranean, Middle Eastern, and Far Eastern descent. It has been estimated that approximately 240 million people worldwide are heterozygotes for β-thalassemia and at least 200,000 affected individuals are born annually (Cao *et al.,* 1997). The β-globin gene is rather short for a human gene (~1.5 kb), but harbors more than 150 mutations worldwide. Despite this heterogeneity, each at-risk population has its own spectrum of 5–10 common mutations (Cao *et al.,* 1997). Due to this phenomenon, an array for identification of 10 common mutations would offer a platform for high-throughput genetic testing of β-thalassemia in a given population. The mutations studied in this work are typical for the Mediterranean region (see Table 1). Extra primers are needed to expand the 10-mutation-specific chip presented here to a human β-globin gene resequencing chip, allowing the analysis of nucleotide changes regardless of patient origin or mutation location.

In this report, we describe the APEX technology and present the results of a β-thalassemia mutation study. Nine DNA samples from patients and carriers, each of which carries a different mutation and four wild-type DNA samples, were correctly identified.

MATERIALS AND METHODS

DNA sequencing

DNA samples were previously sequenced by the Sanger dideoxy chain-termination method.

Template DNA preparation

A 1,420-bp fragment of the human β-globin gene (NID g183829; accession no. M36640) was amplified from human genomic DNA using the following primer sequences:
Forward primer 5′-ACAGGTACGGCTGTCATCAC-3′;
Reverse primer 5′-AGAATAATCCAGCCTTATCCC-3′.

TABLE 1. TEN COMMON POINT MUTATIONS FROM THE β-GLOBIN GENE FOR IDENTIFICATION WITH APEX

Mutation	*Sequence change*	*Class*	*Origin*	*Primers (sense/antisense)*
−87	C → G	Transcriptional mutant	Mediterranean	5′-TAGACCTCACCCTGTGGAGCCACAC 5′-CTGGGAGTAGATTGGCCAACCCTAG
Codon 5	ΔCT	Frameshift	Mediterranean Balkans	5′-ACAGACACCATGGTGCACCTGACTC 5′-CGGCAGTAACGGCAGACTTCTCCTC
Codon 6	ΔA	Frameshift	Mediterranean	5′-GACACCATGGTGCACCTGACTCCTG 5′-CAGGGCAGTAACGGCAGACTTCTCC
IVS-I-1	G → A	Splice junction change	Mediterranean,	5′-AAGTTGGTGGTGAGGCCCTGGGCAG 5′-ACCTGTCTTGTAACCTTGATACCAA
IVS-I-5	G → A	Consensus change	Mediterranean,	5′-TGGTGGTGAGGCCCTGGGCAGGTTG 5′-TTAAACCTGTCTTGTAACCTTGATA
IVS-I-6	T → C	Consensus change	Mediterranean	5′-GGTGGTGAGGCCCTGGGCAGGTTGG 5′-CTTAAACCTGTCTTGTAACCTTGAT
IVS-I-110	G → A	Internal IVS change	Mediterranean	5′-TAGGCACTGACTCTCTCTGCCTATT 5′-GCAGCCTAAGGGTGGGAAAATAGAC
Codon 39	C → T	Nonsense	Mediterranean, European	5′-GCTGCTGGTGGTCTACCCTTGGACC 5′-TCCCCAAAGGACTCAAAGAACCTCT
IVS-II-1	G → A	Splice junction change	Mediterranean, American Black, Asian	5′-GCACGTGGATCCTGAGAACTTCAGG 5′-AAACATCAAGGGTCCCATAGACTCA
IVS-II-745	C → G	Internal IVS change	Mediterranean,	5′-ATTGCTAATAGCAGCTACAATCCAG 5′-ACCATAAAATAAAAGCAGAATGGTA
Markers				5′-TTTAGCCTTAACGCCTXTGACGTCA X = A for self-extending T marker X = C for self-extending G marker, *etc.*

The PCR primers were obtained from Life Technologies, Inc. (Gaithersburg, MD). The amplification mixture was prepared and distributed into 50-μl aliquots The mixture contained: 5 μl of 10× PCR buffer (containing 200 m*M* Tris-HCl, pH 8.4, 500 m*M* KCl (Life Technologies), 2.5 m*M* $MgCl_2$, 0.25 m*M* of each deoxynucleotide triphosphate (dATP, dCTP, dGTP), 0.2 m*M* dTTP, 0.05 m*M* dUTP (Amersham Pharmacia Biotech., Inc., Milwaukee, WI), 40 pmol of each primer, and 1 unit of Platinum Taq DNA Polymerase (Life Technologies). The amplification reactions were performed in a PTC-200 instrument (MJ Research, Inc., Watertown, MA). First, an initial incubation at 94°C for 5 min. was performed, followed by 34 amplification cycles consisting of denaturation at 94°C for 30 sec; primer annealing at 61°C for 30 sec; and extension 72°C for 1 min. The final extension was at 72°C for 5 min.

The amplification products were initially concentrated and purified by ethanol precipitation in the presence of ammonium acetate. Fragmentation and functional inactivation of unincorporated dNTPs was achieved in a one-step reaction by the addition of 1/5 U of shrimp alkaline phosphatase (Amersham Pharmacia Biotech. Inc.) and 1/5 U of thermolabile uracil *N*-glycosylase (Epicentre Technologies, Madison, WI) per one amplification product. The reaction was incubated at 37°C for 1 hour and used directly in primer extension reactions.

Oligonucleotide microchips

Oligonucleotide primers were designed, according to the wild-type sequence of the human β-globin gene, for both sense and antisense directions. 25-Mer oligonucleotides with amino linkers at their 5′ ends were obtained from Genset (Paris, France). All but Codon 5 scanning oligonucleotides were designed to scan 1 bp in the wild-type sequence. To look for the Codon 5 ACT frameshift mutation, an antisense primer with +1 nucleotide shift in the sequence was used. Oligonucleotide primers were attached to an epoxy-activated glass surface via an amino linker at their 5′ end (Southern *et al.*, 1992; Lamture *et al.*, 1994; Shumaker *et al.*, 1996; Pastinen *et al.*, 1997). Glass slides (24 × 60-mm; Fisherfinest Premium Cover Glasses, Fisher Scientific, Pittsburgh, PA) were sonicated in acetone and 100 m*M* NaOH (5 min both), rinsed in MilliQ water, and finally sonicated for 2 min with a solution of 2% (3-glycidoxypropyl)trimethoxysilane (Gelest Inc., Tullytown, PA) in 95% ethanol solution. Unbound silane and residual water was removed by brief rinsing in 100% ethanol. Primers were diluted to 50 μ*M* concentration in 100 m*M* NaOH and spotted onto the activated surface with the TECAN RSP 5031 pipetting robot (TECAN AG, Hombrechtikon, Switzerland) or a custom manufactured 25 Gauge, 96-tip capillary arrayer. The slides were stored in a dust-free environment at 4°C until needed and washed twice in 95°C MilliQ water prior APEX reactions. Slides prepared this way are extremely stable and can be used even after 15 months of storage.

Arrayed primer extension reactions

As estimated by comparison with a Gibco BRL mass ladder, 200–300 ng of the amplified product was used per one APEX reaction. The 20-μl primer extension reactions consisted of 10 μl of fragmented product, 4 U of Thermo Sequenase DNA polymerase (Amersham Pharmacia Biotech.), 2 μl of Thermo Sequenase reaction buffer (260 m*M* Tris-HCl, pH 9.5, 65 m*M* $MgCl_2$) (Amersham Pharmacia Biotech.), and 1 μ*M* final concentration of each fluorescently labeled ddNTP (Amersham Pharmacia Biotech., NEN, Boston, MA). The DNA in buffer was denatured at 95°C, for 5 min. The enzyme and dye were immediately added to the other components and the whole mix was applied to prewarmed slides at 48°C. The reactions were allowed to proceed for 20 min under coverslips and stopped by washing at 80°C for 2× 90 sec in MilliQ water. A droplet of SlowFade® Light Antifade Reagent (Molecular Probes, Eugene, OR) was applied to the chips to limit bleaching of the fluorescein. The signals were acquired by a custom built TIRF-based CCD detector.

TIRF-based image detector

The TIRF-base CCD detector consists of a: (1) set of lasers used to excite the spectrally separable dye set, such as fluorescein, Cy3, Texas Red and Cy5; (2) a mechanism for shuttering, expanding, and launching the lasers sequentially into the glass slides used as microarray substrates; (3) a filter wheel used for sequentially selecting the emission band-pass for each dye; and (4) an imaging lens and CCD imager for recording the spatial fluorescence intensities.

Light from the excitation lasers is directed along common paths. The ribbon of light strikes a continuous linearly recirculating mirror (dither mirror), which deflects the light upward toward the vertex of a prism. The cylinder lens directly in front of the prism has two functions: (1) focusing the ribbon in the narrow dimension at the prism vertex, and (2) producing a continuously changing launch angle together with the dither mirror. The range of angles produced insurers that the excitation is uniform over the slide. Fluorescence is collected by a 60-mm f/2.8 MicroNikkor objective (Nikon, Japan) at a 3:1 imaging ratio. The Quantix CCD camera (Photometrics, Tucson, AZ) is cooled to −25°C and contains a Kodak KAF-1400 CCD chip with 1,037 × 1,315 pixels that are 6.8 μm × 6.8 μm square giving maximum resolution of 20 μm over the slide in the current imaging ratio. For arrays with fewer, larger spots, the full spatial resolution of the camera is not required and the CCD pixels are "binned" in a 2 × 2 fashion permitting 4× faster imaging times. The camera, custom-machined shutters, and FW1 filter wheel (Integrated Scientific Imaging Systems, Inc., Santa Barbara, CA) are controlled by a customized version of Image Pro Plus™ software (Media Cybernetics, Inc., Silver Spring, MD), which automates the acquisition of the sequence of images required for each assay. More information about the TIRF based fluorescence detector can be found at (http://www.asper.ee).

Analysis of data

Images were processed with the Image Pro Plus software. Patterns of all four incorporated nucleotides were recorded under different color codes (A, yellow; T, cyan; C, red; and G, green). Four-color images were generated using a macro.

RESULTS AND DISCUSSION

Design of the assay

An integral part of the assay is the DNA chip. We have used standard microscope cover glasses (24 × 60 mm) activated by

TABLE 2. β-THALASSEMIA APEX SLIDE KEY

	M	*−87*	*Codon 5*	*Codon 6*	*IVS-I-1*	*IVS-I-5*	*IVS-I-6*	*IVS-I-110*	*Codon 39*	*IVS-II-1*	*IVS-II-745*	*M*
Sense	N	C to G	C to T	A to G	G to A	G to A	T to C	G to A	C to T	G to A	C to G	N
Antisense	A	G to C	A to G	T to C	C to T	C to T	A to G	C to T	G to A	C to T	G to C	C
Sense	G	C to G	C to T	A to G	G to A	G to A	T to C	G to A	C to T	G to A	C to G	T
Antisense	N	G to C	A to G	T to C	C to T	C to T	A to G	C to T	G to A	C to T	G to C	N

The left and right columns consist of self-extending marker primers (M) and the middle 10 columns are duplicates of the sense adjacent to antisense primers for the mutation sites listed in Table 1. The letters are showing analysis results expected from a wild-type DNA and mutations, respectively.

epoxy-silanization (Southern *et al.*, 1992; Lamture *et al.*, 1994) for oligonucleotide attachment. The primers were coupled to the activated slides by their 5′ amino linker under alkaline conditions. The average coupling efficiency was 5%, as estimated by immobilization of radioactively labeled oligonucleotide (data not shown).

Template preparation

We have developed a single-tube template preparation protocol, consisting of PCR amplification followed by a DNA fragmentation reaction. A 1.4-kb fragment of the patient DNA was first amplified by PCR. A fraction of the dTTPs was substituted

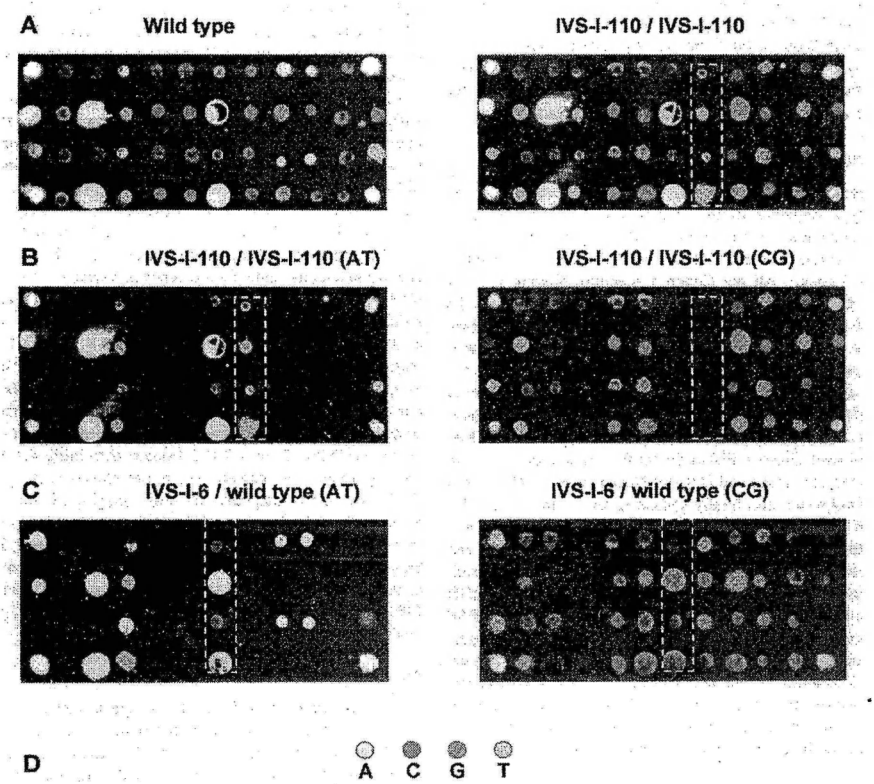


FIG. 1. Four-channel pseudocolor images of β-thalassemia APEX analysis. Table 2 indicates the positions of oligonucleotides on the array. **A.** Wild-type and homozygous IVS-I-110 (G → A; Tables 1 and 2) mutation. Dashed boxes indicate the IVS-I-110 sense and antisense primer locations. **B.** DNA sample carrying homozygous IVS-I-110 (G → A; Tables 1 and 2) mutation. Composed pseudocolor images representing signals from complementary nucleotides are shown for clarity. **C.** DNA sample carrying heterozygous IVS-I-6 (T → C; Tables 1 and 2) mutation. Wild-type allele is apparent in left panel while mutant allele is visible on the right panel. **D.** Color code of the images.

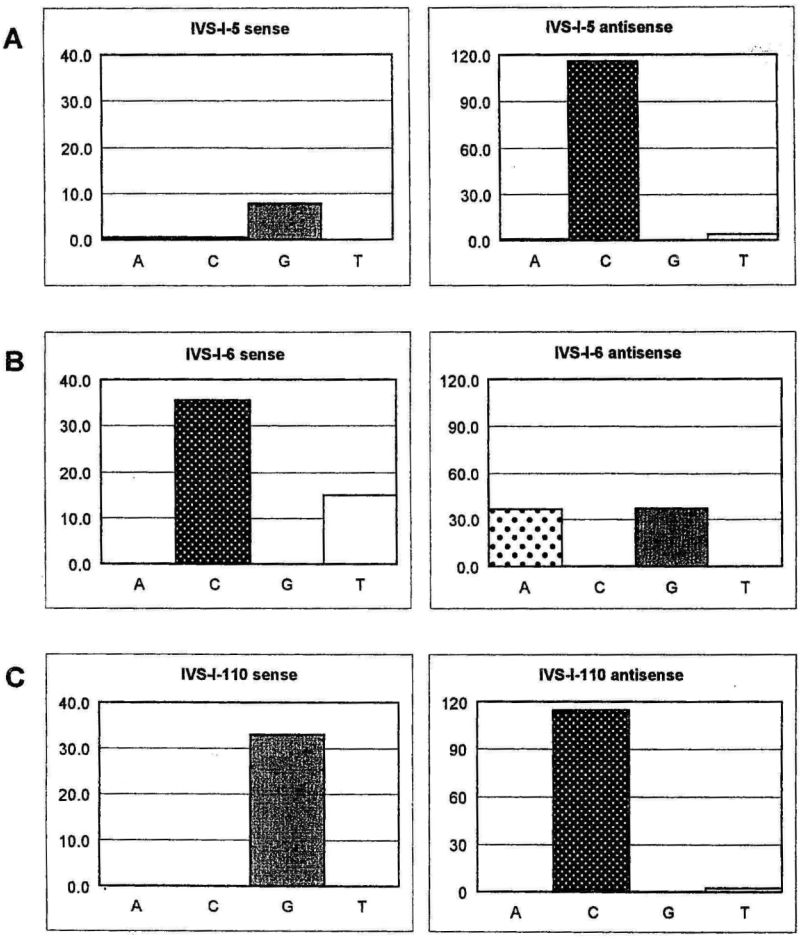


FIG. 2. Fluorescent intensities of three sequential mutation analysis sites on the β-thalassemia oligonucleotide array (Tables 1 and 2). **A.** IVS-I-5 site. Bars represent average fluorescent intensities of three independent experiments with a wild type target DNA. **B.** IVS-I-6 site. Average fluorescent intensities from two different experiments with a heterozygous target DNA. **C.** IVS-I-110 site showing average fluorescent intensities from the same three independent experiments as (**A**).

by dUTPs in the PCR mix allowing for later fragmentation with uracil *N*-glycosylase (UNG) and heat treatment. *In vivo,* UNG acts as one of the most efficient DNA repair enzymes, hydrolyzing specifically the *N*-glycosylic bond connecting uracil to the deoxyribose sugar and generating abasic sites in DNA. *In vitro,* this reaction can be used for asymmetric fragmentation of the template DNA (Cronin *et al.,* 1996). Replacement of 20% of dTTPs was optimal for β-thalassemia APEX. However, other concentrations of dUTP might be needed for templates with different lengths and thymidine content. Fragmentation offers several advantages for APEX by reducing both the effects of secondary structures, reducing the melting temperature of target duplexes, and permitting the analysis of both strands simultaneously. Fragmentation also promotes greater mobility of the template and increases its effective concentration. In addition to the UNG treatment, several other possibilities exist for DNA template fragmentation, such as DNase I treatment, restriction enzyme digestion, and mechanical shearing; however, none of these offers the combination of reproducibility, fragment size, staggered single-sided nicks, and assay flexibility.

After amplification, the PCR products were concentrated by

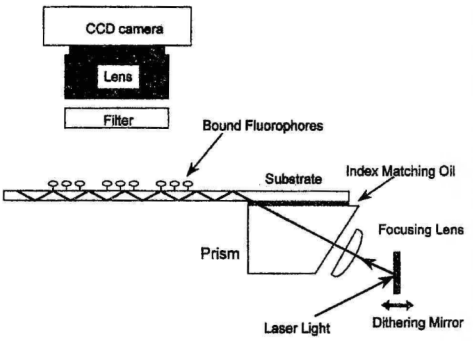


FIG. 3. Scheme of TIRF excitation/CCD imaging system. The excitation beam is trapped in the oligonucleotide array slide by total internal reflection at the slide/air interface. The launch angle is nominally 20° below horizontal. As the excitation beam travels down the oligonucleotide array slide, it expands to fill the slide uniformly. The evanescent field of the trapped beam is used to excite the dye molecules incorporated to the oligonucleotide primers near the surface. Variations in the excitation field are smoothed by the action of the dither mirror.

ethanol precipitation. dNTPs carried over from the PCR reaction are a source of nonspecific extension noise in the APEX reaction and must be removed. The remaining dNTPs were degraded enzymatically using shrimp alkaline phosphatase (sAP). Simultaneously, the amplicons were treated with UNG.

In the case of β-globin gene as the model system, amplification with a single pair of primers was sufficient to evaluate the common mutations. However, to apply the APEX approach to larger genes, amplification with multiple primers may be necessary.

APEX reaction

The majority of the β-thalassemias are caused by either a single nucleotide substitution, or an oligonculeotide addition or deletion that affects the coding region, or critical areas, for the function of the β-globin gene (Cao *et al.*, 1997). Primer extension reaction conditions were optimized with a wild-type DNA to achieve signals from all screened positions in the β-globin gene.

The specificity of the assays was monitored by immobilized self-elongating marker primers (Tables 1 and 2; Fig. 1, A, B and C) that are designed to form self-complementary homoduplexes at the 3′ end, permitting template-independent signals for particular dye terminators. All primers on the array gave signals as expected from the β-globin gene wild-type sequence (Table 2; Fig. 1A). From the analyzed patient DNA, sample 1 was homozygous for the IVS-I-110 mutation (Fig. 1, A and B) and 8 were heterozygous for the −87, Codon 5, Codon 6, IVS-I-1, IVS-I-6 (Fig. 1C), IVS-II-1, Codon 39, and IVS-II-745 mutations, respectively.

The signal-to-noise ratio of primer extension is 40:1, measured as the average fluorescence value for all oligonucleotides on the array from three different experiments (Fig. 2). This favors identification of heterozygous mutations. Covalent bonds between the oligonucleotide and dye terminator allow the slides to be stringently washed, minimizing the nonspecific signals and reducing the background.

Our goal was to make the APEX assay as simple and robust as possible; therefore, we abandoned the two-step assay described earlier (Head *et al.*, 1997; Pastinen *et al.*, 1997). To minimize manual operations with the DNA chips, we have used single-step APEX reactions. Our experiments have shown that both hybridization and template-dependent extension of arrayed primers can be achieved in the same reaction step without removal of unbound template. The same goal applies to the reaction mix; it contains only absolutely necessary components—template DNA, fluorescently labeled dideoxy nucleotides, and high-specificity DNA polymerase in its commercial buffer. Furthermore, the reaction conditions are relatively insensitive to variations in the amount of dye terminator and polymerase.

Some target-dependent primers show self-extension signals, *e.g.*, Codon 39 and IVS-II-1 sense-strand primers gave signals from the wild-type sequence (incorporation of C and G), as well as from self-extension (incorporation of A), respectively, due to formation of primer bridge structures (homodimers) similar to the marker primers used, or to hairpin structures. The primers on the array are designed according to the wild-type sequence of the analyzed gene. Although the 3′ end of the primers cannot be varied, the internal part of the primer may be changed by incorporating a mismatch to reduce primer self-complementarity without seriously affecting the target-specific priming ability. We are screening each mutation from both DNA strands. APEX analysis of the opposite strand does not use the complement of the problematic primer, and thus has a reasonable probability of avoiding the self-priming problem. In addition, there are no sites in the current assay containing such problematic primers for both strands.

TIRF detection system and CCD imaging

We have developed a detection system based on TIRF connected to a CCD image reader (Axelrod *et al.*, 1984; Stimpson *et al.*, 1995). Figure 3 demonstrates the basic idea of the excitation scheme. A laser is deflected by a mirror and focused via a lens through a launch prism and index matching oil onto a glass slide. The intensity of the light field is nearly uniform along the length of the slide. Above the surface of the slide the intensity decreases exponentially, extending for approximately one-quarter of the wavelength of light as an evanescent field, which excites the bound fluorophores residing near the surface. The emitted fluorescence is filtered to reject the background scatter noise, and adjacent dye signals are then collected by a CCD camera. Using one, two, or four dyes, four images are obtained, one for each of the four dye-labeled ddNTPs. The intensities of the imaged spots for each array element are compared and the largest signal will identify the nucleotide in the target sequence. When two signals are present at a location, a heterozygous status is indicated.

The number of detectable fluorescent labels can be adjusted according to assay requirements by the addition (or removal) of corresponding lasers and filters. Two main criteria exist for choice of dye-terminator conjugates. First, they must be spectrally separable from each other and, second, they have to be

incorporated by DNA polymerase. In this report, the "four labels-one reaction" scheme with ddNTPs conjugated to fluorescein, Cy3, Texas Red, and Cy5 was used for β-thalassemia APEX reactions.

The time required for the complete APEX analysis is less than 4 hours, including PCR and sample preparation. However, much of the target preparation and APEX reaction can be performed in parallel, and the detection presents the major limiting step in high-throughput analysis. The TIRF-CCD detector is capable of reading one four-color slide per minute, presenting an ultimate throughput of 60 slides per hour in the present design. For a small number of sites, such as the 10 sites presented here, the assay can be analyzed visually.

We propose that APEX offers a good platform for high-throughput genetic testing. The approach can be applied to any DNA target for analysis. The 40:1 signal-to-noise ratio enables identification of heterozygous mutations with comfortable confidence levels.

ACKNOWLEDGMENTS

The authors wish to thank Heidi Saulep and Viljo Soo for technical assistance. This work was supported by research grants from European Community IC15-CT98-0309, Estonian Science Foundation 2492, and Core Grant 0180518s98 from the Estonian Ministry of Education. J.S. and J.T. were partly supported by a NIST ATP award 70NANB5H1112.

REFERENCES

AXELROD, D., BURGHARDT, T.P., and THOMPSON, N.L. (1984). Total internal reflection fluorescence. Annu. Rev. Biophys. Bioeng. **13,** 247–268.

CAO, A., SABA, L., GALANELLO, R., and ROSATELLI, M. (1997). Molecular diagnosis and carrier screening for β-thalassemia. J. Am. Med. Assn. **278,** 1273–1277.

CHEE, M., YANG, R., HUBBELL, E., BERNO, A., HUANG, X.C., STERN, D., WINKLER, J., LOCKHART, D.J., MORRIS, M.S., and FODOR, S.P.A. (1996). Accessing genetic information with high-density DNA arrays. Science **274,** 610–614.

COTTON, R.G.H., EDKINS, E. and FORREST, S., eds. 1998. *Mutation Detection. A Practical Approach.* Oxford University Press, Inc., New York.

CRONIN, M.T., FUCINI, R.V., KIM, S.M., MASINO, R.S., WESPI, R.M., and MIYADA, C.G. (1996). Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. Hum. Mutat. **7,** 244–255.

HACIA, J.G., BRODY, L.C., CHEE, M.S., FODOR, S.P.A., and COLLINS, F.S. (1996). Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. Nature Genet. **14,** 441–447.

HEAD, S.R., ROGERS, Y.-H., PARIKH, K., LAN, G., ANDERSON, S., GOELET, P., and BOYCE-JACINO, M.T. (1997). Nested genetic bit analysis (N-GBA) for mutation detection in the p53 tumor supressor gene. Nucleic Acids Res. **25,** 5065–5071.

HUISMAN, T.H.J., CARVER, M.F.H., and BAYSAL, E. (1998). *A Syllabus of Thalassemia Mutations.* (ed. by Sickle Cell Anemia Foundation, Augusta, GA).

LAMTURE, J.B., BEATTIE, K.L., BURKE, B.E., EGGERS, M.D., EHRLICH, D.J., FOWLER, R., HOLLIS, M.A., KOSICKI, B.B., REICH, R.K., SMITH, S.R., VARMA, R.S., and HOGAN, M.E. (1994). Direct detection of nucleic acid hybridization on the surface of a charge coupled device. Nucleic Acids Res. **22,** 2121–2125.

LANGEGREN, U., NILSSON, M., and KWOK, P.-Y. (1998). Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. Genome Res. **8,** 769–776.

PASTINEN, T., KURG, A., METSPALU, A., PELTONEN, L., and SYVÄNEN, A.-C. (1997). Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. Genome Res. **7,** 606–614.

SHUMAKER, J.M., METSPALU, A., and CASKEY, T. (1996). Mutation detection by solid phase primer extension. Hum. Mutat. **7,** 346–354.

SOUTHERN, E.M., MASKOS, U., and ELDER, J.K. (1992). Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models. Genomics **13,** 1008–1017.

STIMPSON, D.I., HOIJER, J.V., HSIEH, W.T., JOU, C., GORDON, J., THERIAULT, T., GAMBLE, R., and BALDESCHWIELER, J.D. (1995). Real-time detection of DNA hybridization and melting on oligonucleotide arrays by using optical wave guides. Proc. Natl. Acad. Sci. USA **92,** 6379–6383.

Address reprint requests to:
Prof. Andres Metspalu
Inst. of Molecular and Cell Biology
Chair of Biotechnology
University of Tartu
23 Riia St. Tartu 51010
Estonia

E-mail: andres@ebc.ee

Received for publication September 24, 1999; accepted December 28, 1999.

II

Tõnisson N, Kurg A, Lõhmussaar E, Metspalu A. Arrayed primer extension on the DNA chip — method and applications. Microarray Biochip Technologies *(M. Schena, ed.) Biotechniques Books. Eaton Publishing, Natick MA 2000: 247–3*

12 Arrayed Primer Extension on the DNA Chip: Method and Applications

Neeme Tõnisson[1,2], Ants Kurg[1], Elin Lõhmussaar[1], and Andres Metspalu[1,2]

[1]Institute of Molecular and Cell Biology, University of Tartu, Estonian Biocentre, Tartu, Estonia; [2]Asper Ltd., Tartu, Estonia

INTRODUCTION

The first draft of the complete human genome will be available during the year 2000. Its impact on medicine will be the same as Mendeleev's periodic table on chemistry (i.e., chemistry was differentiated from alchemy). The rapidly accumulating human genetic databases predict an equally rapid increase in the number of genetic tests for both basic biomedical research and medical genetics. Structural (polymorphisms, mutations) and functional data of gene expression patterns are needed to obtain better knowledge about the genetic basis of multifactorial diseases, to allow effective drug discovery, therapy, and personalized medicine. New, cost-effective technologies with the potential of automation and parallelism, together with affordable costs, are needed to satisfy these demands. The rapidly developing DNA array technologies will fulfill a major proportion of these needs.

DNA ARRAY TECHNOLOGIES

DNA Array Formats

A number of reviews on DNA arrays have been published recently; the most comprehensive and recent are collected in the "Chipping Forecast" supplement to *Nature Genetics* (18). An overview of some of the early developments can be

Microarray Biochip Technology
Edited by Mark Schena
© 2000 BioTechniques Books, Natick, MA

found in *Microsystem Technology: A Powerful Tool for Biomolecular Studies* (14). DNA microarrays are devices that contain specific oligonucleotides or longer DNA fragments attached in a discrete order on activated solid surfaces. Three main types of DNA microarrays are used currently: oligonucleotide arrays (15–25 mers), gene expression arrays containing complementary DNAs (cDNAs) and expressed sequence tags (ESTs), and genomic arrays of bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YAC clones). Oligonucleotides on arrays can be either prefabricated and delivered to the surface or synthesized in situ (4,12,22).

Applications of DNA Microarrays

The two main applications for use of DNA arrays are DNA resequencing and comparative gene expression analysis. Oligonucleotide arrays are used for both purposes (2,7,11,22). Differential gene expression patterns can be characterized by cDNA microarrays, when amplified polymerase chain reaction (PCR) products are bound to a solid surface and hybridized simultaneously to mRNA-derived probes from two different sources. The two mRNA-derived samples are labeled with two different fluors, and relative expression levels from two different sources can be obtained by measuring fluorescence intensities; moreover, relative abundance of individual mRNAs can be visualized by comparing the ratios of labels (20). Genomic arrays are used similarly for comparative genomic hybridization (CGH), but are 2 orders of magnitude higher in resolution than the conventional CGH (23).

In this chapter, we focus on the use of oligonucleotide arrays for DNA sequence analysis. This includes resequencing and mutation detection of known genes and single-nucleotide polymorphism (SNP) testing. Although mutation detection is the first obvious application of oligonucleotide arrays, SNP testing to find disease genes underlying complex diseases appears to be the most important one. Moreover, the introduction of personalized drug therapy and early genetic risk assessment will require DNA array techniques to be specific, robust, and cost-effective.

Allelic Discrimination with Enzymatic Reactions

Hybridization is an essential step in all DNA microarray platforms. However, signal-to-noise ratio for single nucleotide discrimination is quite low if the reaction is based only on the DNA hybridization. Single nucleotide discrimination is not overly critical in gene expression monitoring but is extremely important in the detection of heterozygous mutations or polymorphisms and somatic mutations in cancer. This fact has led several groups to look for methods with better allelic discrimination capabilities. A promising alternative lies in the use of enzymes, either polymerase (16,22), ligase (5,11), or cleavase (1), as an additional recognition mechanism when coupled to hybridization.

ARRAYED PRIMER EXTENSION

Arrayed primer extension (APEX) is similar to minisequencing (16,25) and genetic bit analysis (GBA) (15). The APEX method is based on incorporation of four dye terminators into oligonucleotide primers with a DNA polymerase (Figure 1). This can be viewed as similar to Sanger dideoxy sequencing technology.

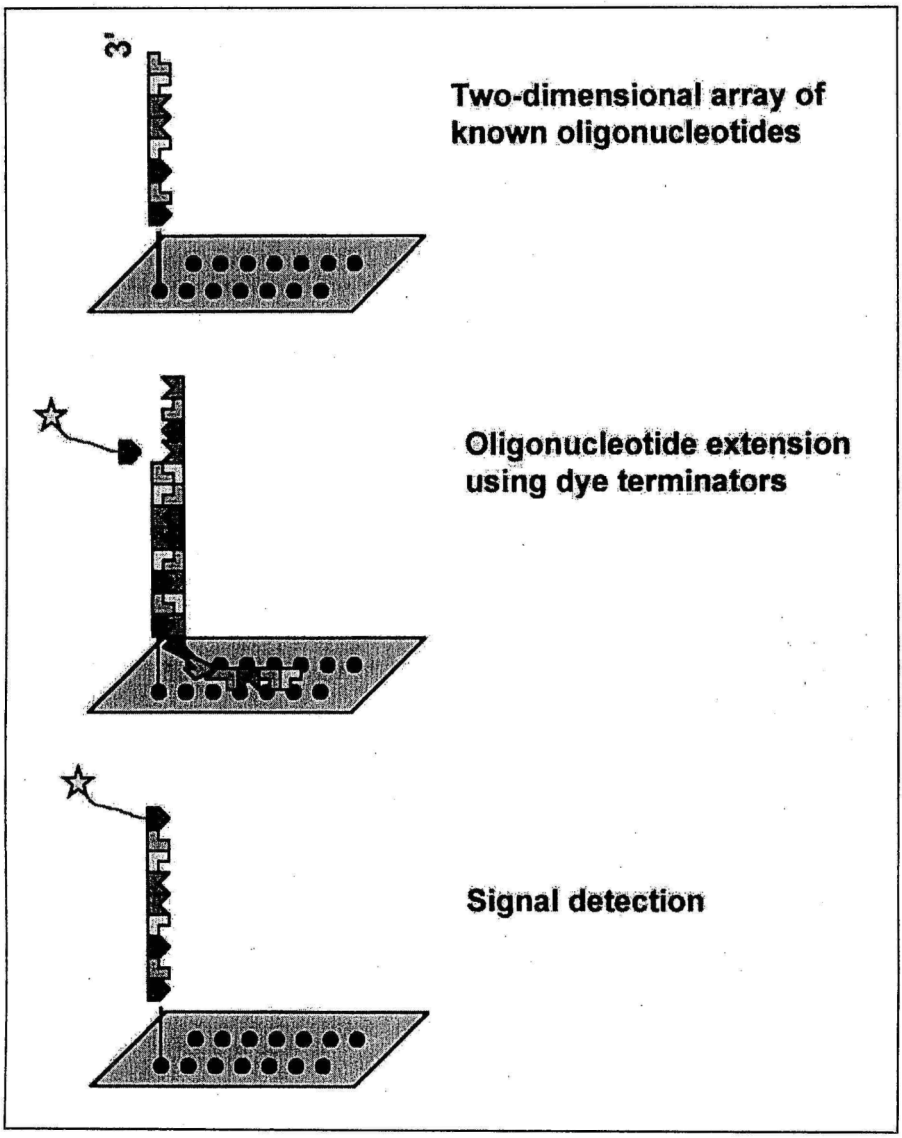


Figure 1. Principle of the arrayed primer extension (APEX) approach.

Instead of one primer being elongated into many fragments and fragment lengths used for their separation, APEX arrays use many primers positioned on the array that are elongated by only one base at each array site.

OLIGONUCLEOTIDE LIGATION REACTION

The oligonucleotide ligation reaction utilizes either one (padlock probe type) or two oligonucleotides with a ligation site located at the nucleotide to be identified (5,11). Both types of ligation assays can be adapted to a microarray platform.

CLEAVASE-BASED ASSAYS

A number of assays have been developed by Third Wave Technologies (http://www.twt.com) that utilize a family of structure-specific enzymes termed Cleavase™. Cleavase cuts junctions between single- and double-stranded regions of DNA (1). The assays are based on the observation that single strands of DNA form distinct higher order structures by folding on themselves. Although they are highly allele specific, robust, and reproducible, the assays are difficult to transform into a microarray format.

APEX ASSAY

The APEX reaction combines the high information content of oligonucleotide microarrays with the specificity of molecular recognition by DNA polymerase. APEX is capable of identifying mutations and polymorphisms (Figure 2), and can be used for resequencing (13,14). For testing known mutations, primers are selected to allow discrimination between wild-type and mutant alleles. For unknown mutations, the only option is to use the DNA resequencing application on the chip. This assay may be one of the most used genetic tests for mutation detection in the future.

Hybridization platforms like Affymetrix require many oligonucleotides and complicated analysis algorithms to analyze signal intensity and make comparisons across multiple features on the chip to obtain valid data. APEX, on the other hand, utilizes only two oligos based on a wild-type sequence to analyze one base pair. The analysis is similar to that of an automated ABI type DNA sequencer that uses normalized intensity comparisons from different fluorescent labels on the same band on a gel.

Oligonucleotide Selection

APEX technology can be tailored to different applications simply by changing the arrayed oligonucleotides. Oligonucleotides are synthesized according to

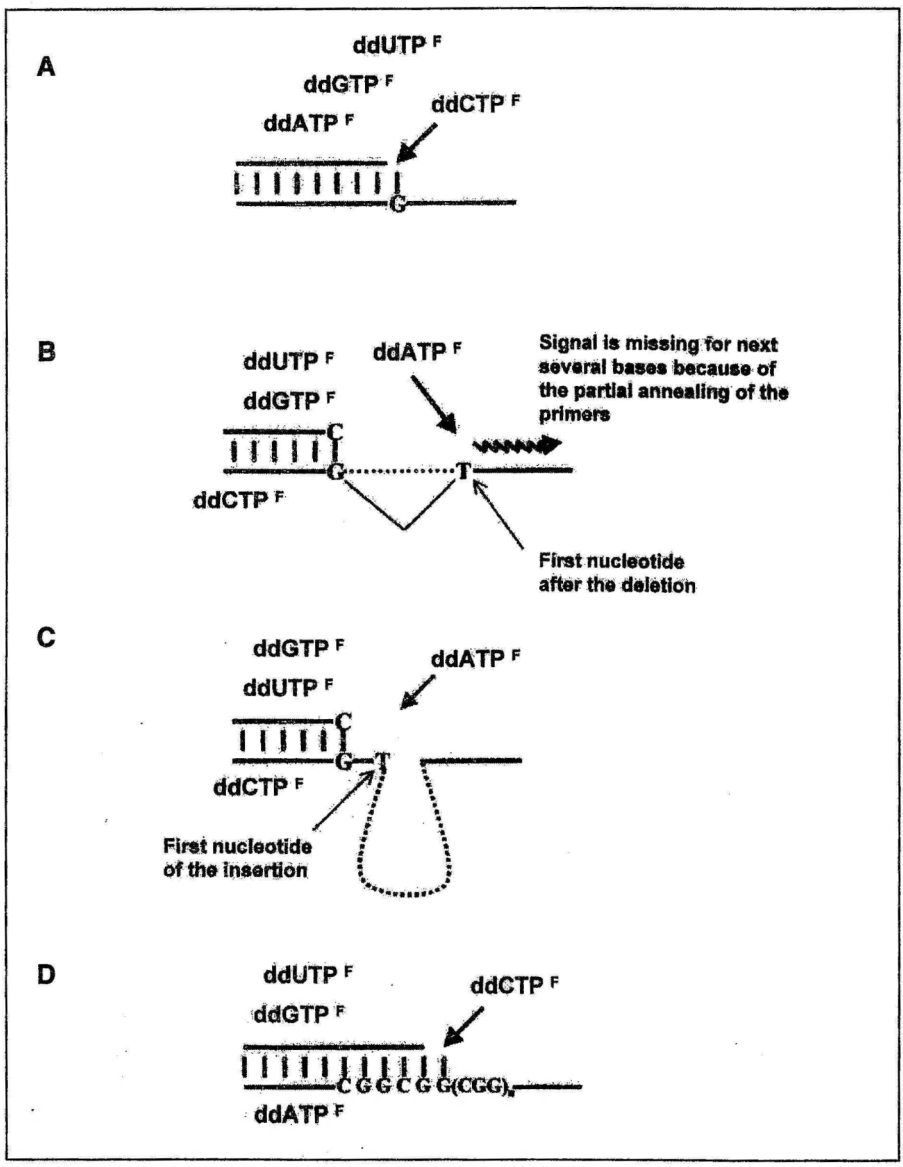


Figure 2. DNA sequence alterations analyzed by APEX. (A) Single-nucleotide polymorphism (SNP) and point mutations. Since all four dye terminators (ddNTPs) are present in the primer extension reaction, oligonucleotides are extended by a complementary nucleotide. (B) A deletion analysis primer immediately upstream of the deletion is extended by a dye terminator complementary to the base after the deletion. This allows detection of the borders of the deletion. By analyzing both strands, nucleotide sequence upstream and downstream of the deletion can be obtained. The length of the detection equals the "footprint" (region with no signal) on the resequencing chip. (C) Insertions. By analyzing both strands, one can detect the first and the last nucleotide of the inserted sequence. Mono- or dinucleotide insertions are identified completely. (D) Repeats. Di-, tri-, tetra (etc.) nucleotide repeats can be analyzed if the repeat length is shorter than the oligonucleotide primer on the oligonucleotide microarray. Long repeats (as in the fragile X case) cannot be analyzed at the present time using this version of the APEX assay.

the wild-type sequence, with an amino linking group at the 5′ end. The amino link is connected to a spacer arm of 12 to 18 carbon atoms and ensures covalent immobilization of the oligonucleotides via their 5′ end. The 3′ end of the oligonucleotides is free for enzymatic extension by DNA polymerase. The length of the spacer arm has a significant impact on oligonucleotide accessibility by the template DNA strand and consequently on hybridization efficiency (24). The 3′ end of each oligonucleotide is positioned one nucleotide upstream of the nucleotide to be analyzed. For resequencing with the APEX assay, oligonucleotides are designed with a one-nucleotide shift in their sequence. For example, 1000 oligonucleotides are needed to identify 1000 bases in the sample DNA. As APEX allows analysis of both strands in the same reaction, two oligos are arrayed per base pair. We have used 20- and 25-mer oligonucleotides in the APEX assay. In our experience, signals with 25-mers are 10% to 50% stronger than with 20-mers. Unfortunately, 25-mers also have more stable secondary structures (hairpins or oligonucleotide dimers), resulting in a loss of specificity due to self-extension. Approximately 10% to 15% of oligonucleotides do not give satisfactory signals on the first attempt and need to be replaced with new oligonucleotides. There is currently no reliable method to predict the overall hybridization and self-extension behavior of arrayed primers prior to running the APEX assay.

Surface Chemistry Used for Oligonucleotide Arrays

A variety of solid supports such as glass, plastic, nylon, and polyacrylamide have been used in DNA chip manufacturing. An ideal substrate for APEX would have certain technical and chemical properties. These would include mechanical and chemical stability, ease of chemical modification and derivatization, absence of autofluorescence, and low cost. We and many other laboratories have chosen glass microscope slides as the substrate material, due to low cost and ready availability. Glass provides a nonporous, relatively homogenous chemical surface that can be used in conjunction with silane chemistries. However, low-cost glass may be heterogeneous, and this variability may cause problems during surface treatment. High-quality glass substrates are now available (http://arrayit.com and http://www.cmt.corning.com), providing greater reliability during surface modification.

The attachment chemistry must also meet several criteria. The accessibility and functionality of the surface-bound DNA for hybridization, the density of attached oligonucleotides, and the reproducibility of the attachment chemistry are the most critical parameters. The chemical linkage must be stable, exhibit minimal nonspecific binding to the reaction components, and provide sufficient clearance from the solid surface to minimize or eliminate steric hindrance.

The two most commonly used silane-modified surfaces for DNA array preparation include epoxysilane-modified (10) and aminosilane-modified surfaces (6). Oligonucleotide primers are attached to epoxy-activated glass surfaces by secondary amine formation between the epoxy-derivatized glass surface and 5′

amino linkers. Amino-linked primers are diluted in sodium hydroxide and spotted onto the activated surface (22).

For amino derivatization, the slides are silanized with 3-aminopropyltrimethoxysilane solution. To convert the amino groups to amino-reactive phenylisothiocyanate groups, the slides are treated with a solution of 1,4-phenylene diisothiocyanate. Amino-modified oligonucleotide primers are diluted in sodium carbonate/bicarbonate buffer (pH 9.0) and spotted onto the activated surface as described above. The slides are later blocked with NH_4OH, washed with water, dried with nitrogen flow, and stored at room temperature. In our experience, they are stable at least for 6 months from the time of preparation.

Template Preparation for APEX

Preparation of template DNA for APEX starts from (multiplex) amplification by the PCR. Both single- and double-stranded target DNA can be used in APEX reactions. Single-stranded DNA has the advantage of lacking a complementary strand, thereby eliminating competing hybridization with arrayed primers. The main disadvantage is that preparing sufficient amounts of single-stranded target is laborious and expensive, and one loses 50% of the genetic information in the assay. The use of a double-stranded target permits single base changes to be identified from both strands in a single reaction, thereby increasing the reliability of the assay.

One important issue in the APEX assay is the length of the template. We have found that the maximum length of a double-stranded PCR product that can be used is approximately 200 bp, but 100 bp is optimal. Longer PCR products require fragmentation before use in the APEX reaction. Fragmentation is performed by replacing a fraction of the dTTPs with dUTPs in the amplification mix, followed by the use of uracil *N*-glycosylase (UNG) to fragment the template (3). UNG is highly specific to uracil bases in DNA, and therefore the reaction can be controlled by dUTP incorporation during PCR. By changing the dUTP concentration from 15% to 30% of dTTP, one can vary the mean length of the template DNA fragments. Several other possibilities exist for DNA target fragmentation, including DNase I treatment, restriction enzyme digestion, and mechanical shearing. However, none of these worked reproducibly in our hands.

The APEX principle for the single-nucleotide extension reaction works only if no deoxyribonucleotide triphosphates are carried over from the amplification reaction. A simple and quick way to inactivate the dNTP leftover from PCR is enzymatic digestion with shrimp alkaline phosphatase (SAP). This is performed in a single-tube reaction together with UNG treatment, and both enzymes are thermally inactivated prior to the APEX reaction (Figure 3).

APEX Reaction

APEX is a complex, single-step reaction consisting of target annealing to the

oligonucleotide array and an enzymatic primer extension reaction with fluorescent dideoxy nucleotides. Engineered DNA polymerases (26) are able to incorporate dye terminators quite efficiently. We use Thermo Sequenase (Amersham Pharmacia Biotech, Piscataway, NJ, USA) and 20 to 50 pmol of each fluorescent terminator per APEX reaction. Other commercially available polymerases are also capable of incorporating labeled dideoxy nucleotides, including DynaSeq

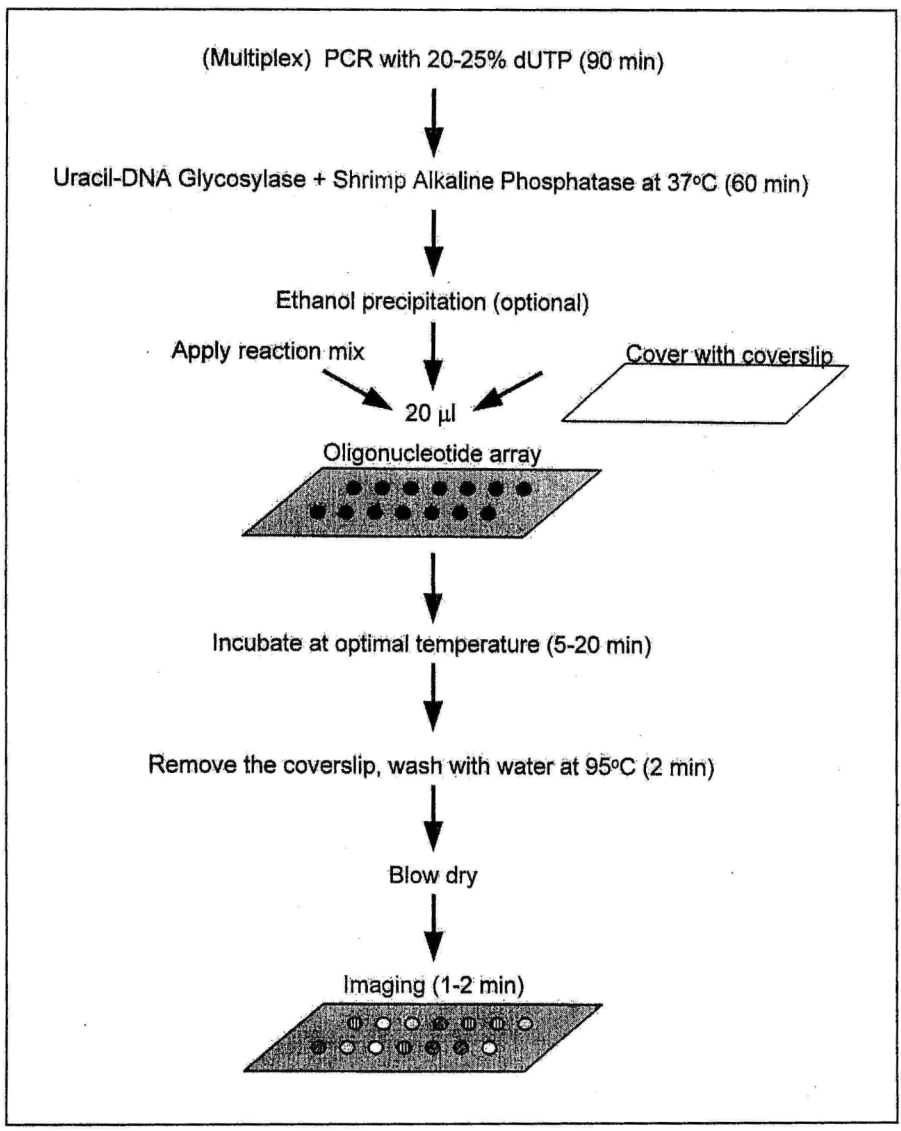


Figure 3. One-tube, one-chip reaction scheme of the APEX reaction. It takes about 3 hours for the entire test, including template amplification by PCR.

(Finnzymes OY, Espoo, Finland), AmpliTaq® FS (Roche Molecular Systems, Branchburg, NJ, USA), and others. Template hybridization to the microarray and the primer extension reaction are different in terms of their optimal temperatures. Hybridization is more effective at lower temperatures, whereas polymerase works better at higher temperatures. For this reason, some primer extension protocols are designed as two separate steps, whereby hybridization is performed first and primer extension with an appropriate DNA polymerase second (16). Unbound target is removed by washing between these two steps. Our goal is to make the APEX assay as robust as possible, and therefore we have used single-step reactions. Hybridization and target-dependent extension of arrayed primers can be performed simultaneously. The signal-to-noise ratio of APEX is currently 30:1 to 100:1. The primer extension reactions are routinely performed at a constant temperature at 48°C for 20 minutes (9). Incorporation of labeled terminators is a rapid reaction, but hybridization is an equilibrium process that requires longer reaction times to obtain strong signals.

The choice of dye labels is determined by two factors: (*i*) the labels must be well separated spectrally, and (*ii*) they must be efficiently incorporated by the DNA polymerase used. Currently, we use the following labels listed in increasing wavelength of emitted light: fluorescein, Cy3, Texas Red, and Cy5. All but fluorescein are very stable to photobleaching. Because of the photosensitivity of fluorescein, we use the antifading reagent SlowFade® Light (Molecular Probes, Eugene, OR, USA) for the imaging step.

Arraying Equipment

Microarrays containing a small number of spots (e.g., 1–2-mm center-to-center spacing) can be made with standard laboratory pipetting robots. Denser arrays require advanced arraying equipment for preparation. Current robotic systems are capable of placing up to 10 000 DNA spots/cm^2 (21). Arraying is done either by mechanical microspotting or ink-jet-type piezoelectric dispensing. A number of arrayers are commercially available from a number of different vendors, and most use different variations of pin spotting (18). An interesting technology developed by Genetic MicroSystems (Woburn, MA, USA; http://www.geneticmicro.com) combines pin spotting and sample loading in a ring. In the Pin and Ring technology, a small-diameter ring is first immersed into the sample solution and then the pin is driven through the ring to deliver nanoliter volumes of the solution to the array surface. One drawback of the Pin and Ring is that the ring holds a rather large volume of sample, requiring a large sample volume for ring loading. Regular spot geometry and uniform density are both very critical parameters for microarray manufacture.

A number of different companies are also offering custom-made microarrays and silanized glass slides, which are probably the best options for small laboratories requiring only a small number but high-quality arrays. Custom microarray

services (see, for example, http://arrayit.com) obviate the need for expensive hardware, sample storage, surface chemistry, labor, and many other issues.

Imaging Equipment

A high-quality imaging system must allow at least 10 × 10 pixels per arrayed element. This means that the minimal pixel diameter of the detector must be at least 10 μm for a 100-μm center-to-center distance array. Some arrayers operate in the 200-μm center-to-center distance, allowing acceptable detector readout with 20-μm pixel size. Commercially available detectors mainly use confocal scanning for array imaging (18). Confocal scanning has a high sensitivity and dynamic range but is significantly slower than total internal reflection (TIR)-based imaging.

We have developed an automatic TIR-based system for fast analysis (FD-003) of APEX reactions (Figure 4). Four lasers of 3- to 10-mW power, with light-emitting wavelengths of 488, 532, 594, and 635 nm, are used to excite the fluorescent labels on the APEX microarrays. The optical system consists of light reflectors and cylindrical lenses that transform a beam from each laser into a homogeneously illuminated stripe that is directed into the interior of the glass slide. Due to TIR of the slide surface, the incoming beam spreads only inside the slide and therefore the intensity of light remains uniform along the entire length of the substrate. The evanescent light field excites the bound fluorophores residing near the surface of the slide, and a gated charge-coupled device (CCD) camera cooled to -25°C records the emitted fluorescence. The respective narrow-band interference filters fixed on the revolving wheel in front of the CCD camera depress noise radiation from scattered laser light on the slide surface. Using one, two, or four dye markers, four images corresponding to four laser wavelengths are obtained, one for each dye-labeled ddNTP.

Custom software controls all the parameters of the detection procedure including switching on and off the shutters of the lasers, choosing respective interference filters and setting the recording time at each wavelength. Because of precise spectral separation of exciting and emitting wavelengths, the system demonstrates a high signal-to-noise ratio. The time required for one fluorescence channel is 10 to 60 seconds, enabling a theoretical throughput of 60 samples per hour. More information on the TIR-based fluorescence detector can be found at http://www.asper.ee. Additional software was developed to convert the fluorescence signals into DNA sequence information.

APEX APPLICATIONS

Detecting Common Mutations in the Human β-Globin Gene

β-Thalassemia is a common autosomal recessive disorder caused by mutations in the human β-globin gene, resulting in imbalanced globin β-chain production.

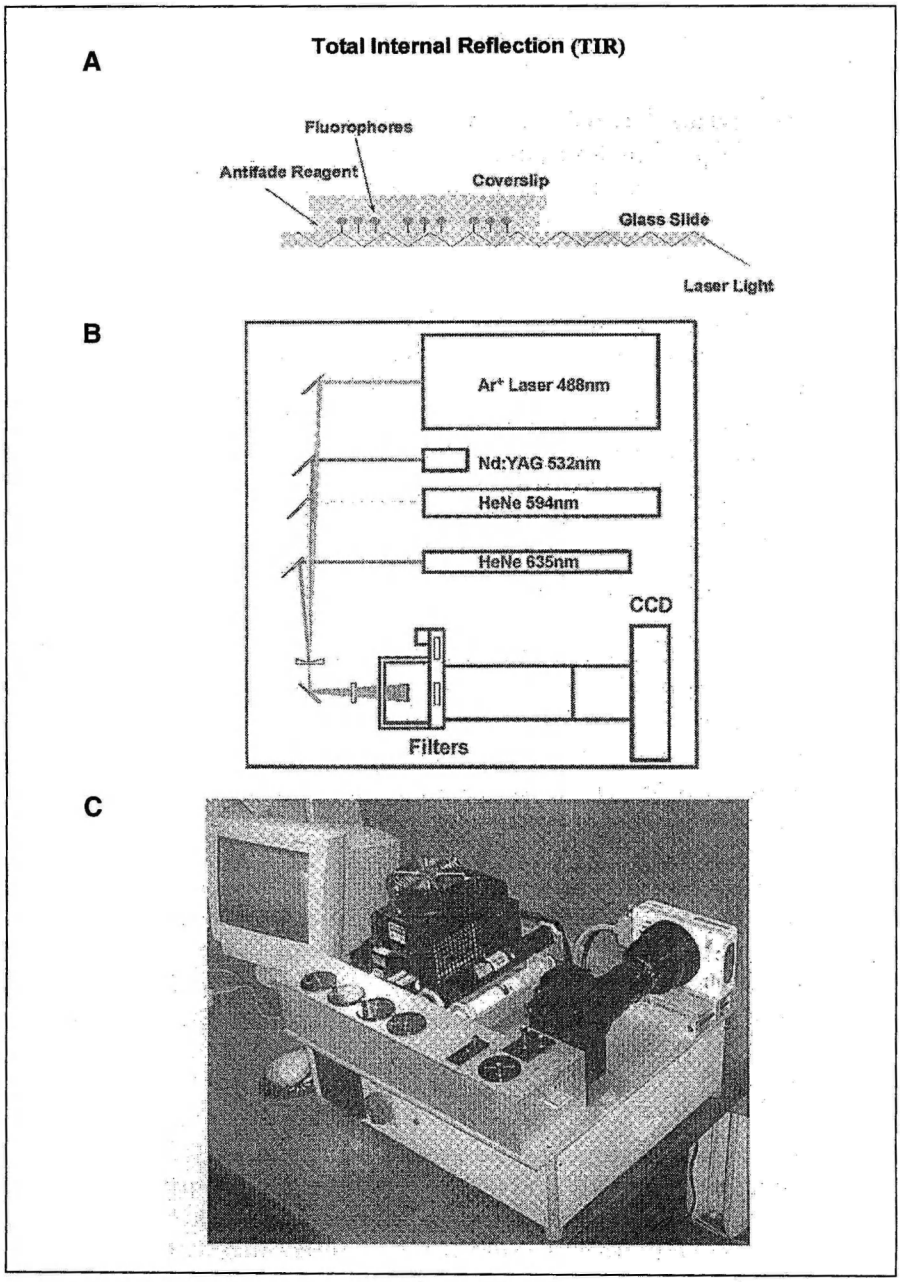


Figure 4. Four-color fluorescence detector based on total internal reflection (TIR). (A) Principle of the fluorophore excitation mechanism based on TIR. The glass slide acts as a waveguide, and fluorophores residing near glass surface are excited by the evanescent light field. (B) Layout of main detector components. Light from four lasers is collected into a common path, transformed into a light stripe, and introduced into the glass slide. Emission is filtered and collected into a cooled CCD camera. Four images are taken per APEX reaction, one for each fluorophore. (C) Photo of the fluorescence detector.

By World Health Organization estimates, approximately 240 million people worldwide are heterozygous for β-thalassemia, and annually at least 200 000 affected individuals are born. Current screening methods are performed via chromatography, protein electrophoresis, and evaluation of the hematological indices. As a model platform for high-throughput genetic testing of β-thalassemia, we used APEX technology to identify 10 β-globin gene mutations, accounting for >95% of all cases found in the Mediterranean region.

To perform the test, a 1.4-kb fragment was amplified from genomic DNA. In the reaction, 25% of the dTTP was replaced with dUTP in the amplification mix. The amplification products were purified and concentrated by ethanol precipitation. Fragmentation of the target and cleanup of the leftover dNTPs was done simultaneously with UNG and SAP. A 200-ng aliquot of the amplified target was used for each APEX reaction. A readout of the β-thalassemia APEX analysis is shown in Figure 5. Nine patient samples containing either homozygous or heterozygous mutations in the β-globin gene were correctly identified using the approach described above (9).

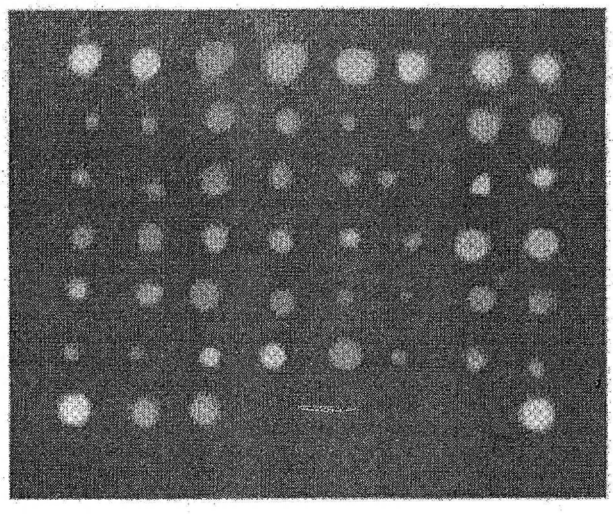
Population Genetics: Finding Risk Factors for Myocardial Infarction

Because it was efficient and fast, a minisequencing method similar to the APEX platform described was used by a Helsinki group to identify genetic risk factors of myocardial infarction in the Finnish population (17). A total of 152 individuals with a history of myocardial infarction and an identical number of control individuals, matched for age, sex, and geography were simultaneously studied for 12 polymorphisms in eight genes known to play roles in platelet adhesion, coagulation, or fibrinolysis. Two alleles were found to be associated with an increased risk of myocardial infarction. Exemplifying the advantage of multiplex genetic analysis, carrier status for both of these polymorphisms conferred additive risk. This is the first study in which minisequencing on oligonucleotide arrays was applied to a large number of individuals. In total approximately 4000 genotypes were identified in this study, greatly exceeding the scope of the model experiment.

SNP Testing

SNPs are the most common type of DNA sequence variation. SNPs are estimated to occur once every 500 to 1000 bp when any two chromosomes are compared (19,27). Recent population modeling demonstrates that one has to map as many as 500 000 SNPs from an individual to perform linkage disequilibrium (LD) analysis (8). SNPs offer a number of advantages with respect to population-based analysis of the human genome. They are very frequent and allow binary assays to be developed, thereby permiting automation and digitalization of the data, which are prerequisites for storing and analyzing the huge amount of genetic information acquired from DNA chip experiments.

A



B

	S 25mer	S 20mer	As 25mer	As 20mer	S 25mer	S 20mer	As 25mer	As 20mer
1	N	A	C	G	T	AC	AG	N
2	IVS-I-110 G to A	IVS-I-110 G to A	IVS-I-110 C to T	IVS-I-110 C to T	Codon 39 C to T	Codon 39 C to T	Codon 39 G to A	Codon 39 G to A
3	IVS-I-1 G to A	IVS-I-1 G to A	IVS-I-1 C to T	IVS-I-1 C to T	IVS-I-6 T to C	IVS-I-6 T to C	IVS-I-6 A to G	IVS-I-6 A to G
4	IVS-II 745 C to G	IVS-II 745 C to G	IVS-II 745 G to C	IVS-II 745 G to C	Codon 6 A to G	Codon 6 A to G	Codon 6 T to C	Codon 6 T to C
5	IVS-II 1 G to A	IVS-II 1 G to A	IVS-II 1 C to T	IVS-II 1 C to T	IVS-I-5 G to A	IVS-I-5 G to A	IVS-I-5 C to T	IVS-I-5 C to T
6	Codon 5 C to T	Codon 5 C to T	Codon 5 A to G	Codon 5 A to G	-87 C to G	-87 C to G	-87 G to C	-87 G to C
7	N	Codon 5 G	Codon 5 G					N

Figure 5. β-Thalassemia APEX with a wild-type DNA sample. (A) Color code of the incorporated nucleotides is shown below the image. (B) Key code of the array. Seven rows and eight columns of oligonucleotides were arrayed onto the slide. Two oligonucleotides (20- and 25-mer) for the sense strand (s) and two for the antisense strand were used for each mutation site. Upper row and four corners are self-elongating marker oligonucleotides. Wild-type signals and signals expected from mutations are indicated at the bottom of each cell. **(See color plate A30.)**

We are working on APEX-based DNA arrays to analyze genome-wide SNP markers—one from each chromosome. We have tested the array with individually amplified SNP-specific templates and provide a complete experimental protocol below. Quantitative results of the SNP tests are shown in Figure 6. Although all the SNPs were tested using specifically amplified templates, we fully understand that it would be impractical to amplify 10 000 to 500 000 fragments for SNP mapping in LD studies. To solve the problem, we are working in two directions: first, we are using whole-genome amplification to generate enough DNA for APEX, and second, we are increasing the sensitivity of the APEX system including optimization of surface chemistry, enzyme reactions, and imaging so that 100 000 to 500 000 copies of the genome obtained from 10 to 100 μL of whole blood would be sufficient for the genetic test.

Template DNA preparation for SNP assay. PCR primers were obtained from Life Technologies (Gaithersburg, MD, USA). The PCR mixture (50 μL) contained the following: 5 μL of 10× PCR buffer [200 mM Tris-HCl (pH 8.4), 500 mM KCl; Life Technologies], 2.5 mM $MgCl_2$, 0.2 mM of each deoxynucleotide triphosphate (dATP, dCTP, dGTP), 0.16 mM dTTP, 0.04 mM dUTP (Amersham Pharmacia Biotech), 40 pmol of each primer, 50 ng of template DNA, and 1 unit of Platinum™ Taq DNA Polymerase (Life Technologies). The amplification reactions were performed in a PTC-200 thermocycler (MJ Research, Watertown,

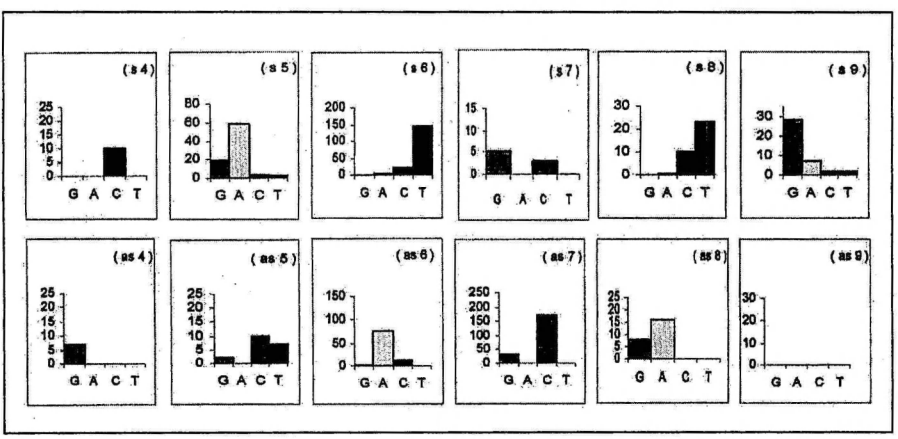


Figure 6. Average fluorescent signals of six sequentially arrayed SNPs from six identical APEX experiments with the same DNA sample. Both sense (s) and antisense (as) strands were analyzed on the same microarray. Positions 5, 7, 8, and 9 are heterozygous. The rest of positions (4 and 6) are homozygous. The signals from correctly incorporated nucleotides are clearly different from other signals. Position 4, for example, is homozygous with a C/G base pair. Position 5 is heterozygous with G plus A signal from the sense strand and corresponding C plus T signal from the antisense strand. Although these signals result only from correct nucleotide incorporation, incorrect nucleotides are sometimes incorporated also, for example, G signal in the antisense position of SNP 5. This misincorporation may result because of oligonucleotide self-extension or low stringency in the hybridization reactions. In some cases, as in SNP 9, the signal was obtained only from the sense strand of the DNA analyzed. The difficulty may arise from poor quality of the oligonucleotides or stable oligonucleotide secondary structures, which require substantially more arrayed oligonucleotide or more template in the reaction mix.

MA, USA). An initial incubation was performed at 95°C for 3 minutes, followed by 35 amplification cycles consisting of denaturation at 95°C for 30 seconds; primer annealing at 56°C for 30 seconds; extension at 72°C for 25 seconds; and the final extension at 72°C for 5 minutes. The amplification products were initially concentrated and purified by ethanol precipitation in the presence of ammonium acetate. DNA fragmentation and functional inactivation of unincorporated dNTPs was achieved in a one-step reaction by addition of 0.5 U of SAP (Amersham Pharmacia Biotech) and 0.5 U of thermolabile uracil *N*-glycosylase (Epicentre Technologies, Madison, WI, USA) per amplification. The reaction mixture was incubated at 37°C for 1 hour and used directly in primer extension reactions.

SNP model array. Oligonucleotides for primer extension were designed according to the wild-type sequence for each SNP in both the sense and antisense orientations. The 25-mer oligonucleotides with 12 carbon amino linkers at their 5′ end were obtained from Genset (Paris, France). Oligonucleotide primers were attached to an epoxy-activated glass surface via amino linkers at their 5′ ends (14). The 24 × 60-mm glass microscope coverslips (Fisherfinest Premium Cover Glasses, Fisher Scientific, Pittsburgh, PA, USA) were sonicated in acetone and 100 mM NaOH for 5 minutes each; rinsed in dH_2O (MilliQ) water; and finally sonicated for 2 minutes in 2% 3-glycidoxypropyltrimethoxysilane (Gelest, Tullytown, PA, USA) in 95% ethanol. Unbound silane and residual water were removed by a brief rinsing of the coverslip in 100% ethanol. Primers were diluted to 50 μM in 0.1 M NaOH and spotted onto the activated surface with a custom-manufactured 25-gauge, 96-channel spotter. The slides were stored in a dust-free environment at room temperature until needed and washed twice in 95°C dH_2O (MilliQ) water prior to the APEX reactions.

APEX reactions for SNP testing. The 20-μL primer extension reactions consisted of 10 μL fragmented DNA (200 ng of each amplified product), 6 U of Thermo Sequenase DNA polymerase (Amersham Pharmacia Biotech), 4 μL Thermo Sequenase reaction buffer concentrate [260 mM Tris-HCl (pH 9.5), 65 mM $MgCl_2$; Amersham Pharmacia Biotech] and 12 μM of each fluorescently labeled ddNTP (FITC-G, Cy3-C, Texas Red-A, and Cy5-U; NEN, Boston, MA, USA, and Amersham Pharmacia Biotech). The DNA in the buffer was denatured at 95°C for 10 minutes. The enzyme and dye terminators were immediately added to other components, and the whole mix was applied to washed and prewarmed slides at 48°C. The reactions were allowed to proceed for 5 to 25 minutes under coverslips and stopped by washing three times at 95°C for 90 seconds in dH_2O (MilliQ) water. A droplet of SlowFade® Light Antifade Reagent (Molecular Probes, Eugene, OR, USA) was applied to the chips before imaging. The signals were acquired with a custom-built TIR-based detector with a CCD camera FD-003 (Asper Ltd., Tartu, Estonia).

Analysis of data. Images were processed with Image Pro Plus™ software (Media Cybernetics, Silver Springs, MD, USA) and custom-written software to obtain the allele data from the analyzed SNPs.

CONCLUSIONS AND PERSPECTIVES

At the end of the first decade of DNA microarray technology development, no doubt remains that the DNA chips are useful and effective in many applications. Although large genomics and pharmaceutical companies have solved many of the initial questions concerning DNA microarrays, some problems remain. We need more standardization, better internal controls, and uniform data formats, all of which will make the results obtained by different groups comparable.

The arrayed primer extension technology (APEX), described in detail here, can be viewed as an advance in hybridization-based microarray technology with an additional allele-discriminating mechanism. We have developed a fully integrated genetic testing system consisting of template preparation, multiplex primer extension on the microarray, fluorescence imaging, and data analysis. One developmental cycle for our laboratory is complete. The next step will be to demonstrate the versatility and usefulness of the APEX technology in daily experimentation.

ACKNOWLEDGMENTS

The authors thank Heidi Saulep, Viljo Soo, and Jana Zernant for technical assistance, Evgeny Berik (Estla Ltd., Tartu, Estonia) for constructing the TIR-based imaging system, Mrs. Krista Liiv for the artwork, and Dr. Hiljar Sibul for critical reading of the manuscript. This work was supported by research grants from the European Community (IC15-CT98-0309), the Estonian Science Foundation (2492), and the Estonian Ministry of Education (core grant 0180518s98).

REFERENCES

1. **Brow, M.A., M.C. Oldenburg, V. Lyamichev, L.M. Heisler, N. Lyamicheva, J.G. Hall, N.J. Eagan, D.M. Olive, L.M. Smith, L. Fors, and J.E. Dahlsberg.** 1996. Differentiation of bacterial 16S rRNA genes and intergenic regions and Mycobacterium tuberculosis katG genes by structure-specific endonuclease cleavage. J. Clin. Microbiol. *34*:3129-3137.
2. **Chee, M., R. Yang, E. Hubbell, A. Berno, X.C. Huang, D. Stern, J. Winkler, D.J. Lockhart, M.S. Morris, and S.P.A. Fodor.** 1996. Accessing genetic information with high-density DNA arrays. Science *274*:610-614.
3. **Cronin, M.T., R.V. Fucini, S.M. Kim, R.S. Masino, R.M. Wespi, and C.G. Miyada.** 1996. Cystic fibrosis mutation detection by hybridization to light generated DNA probe arrays. Hum. Mutat. *7*:244-255.
4. **Fodor, S.P.A., J.L. Read, M.C. Pirrung, L. Stryer, A.T. Lu, and D. Solas.** 1991. Light-directed, spatially adressable parallel chemical synthesis. Science *251*:767-773.
5. **Gunderson, K.L., X.C. Huang, M.S. Morris, R.J. Lipshutz, D.J. Lockhart, and M.S. Chee.** 1998. Mutation detection by ligation to complete *n*-mer DNA arrays. Genome Res. *8*:1142-1153.
6. **Guo, Z., R.A. Guilfoyle, A.J. Thiel, R.Wang, and L.M. Smith.** 1994. Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. Nucleic Acids Res. *22*:5456-5465.
7. **Hacia, J.G., L.C. Brody, M.S. Chee, S.P.A. Fodor, and F.C. Collins.** 1996. Detection of heterozy-

gous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. Nat. Genet. *14*:441-447.

8. **Kruglyak, L.** 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat. Genet. *22*:139-144.
9. **Kurg, A., N. Tõnisson, I. Georgiou, J. Shumaker, J. Tollett, and A. Metspalu.** Arrayed primer extension: solid phase four-color DNA resequencing and mutation detection technology. Genet. Test. (In press).
10. **Lamture, J.B., K.L. Beattie, B.E. Burke, M.D. Eggers, D.J. Ehrlich, R. Fowler, M.A. Hollis, B.B. Kosicki, R.K. Reich, S.R. Smith, et al.** 1994. Direct detection of nucleic acid hybridization on the surface of a charge coupled device. Nucleic Acids Res. *22*:2121-2125.
11. **Landegren, U., R. Kaiser, J. Sanders, and L. Hood.** 1988. A ligase-mediated gene detection technique. Science *241*:1077-1080.
12. **Lockhart, D.J., H. Dong, M.C. Byrne, M.T. Follettie, M.V. Gallo, M.S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton, and E.L. Brown.** 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. *14*:1675-1680.
13. **Metspalu, A., H. Saulep, A. Kurg, and N. Tõnisson.** 1998. Primer extension from two-dimensional oligonucleotide grids for DNA sequence analysis. *In* G.K. Dixon, L.G. Copping and D. Livingstone (Eds.), Genomics Commercial Opportunities from a Scientific Revolution. BIOS Scientific Publishers, Oxford.
14. **Metspalu, A. and J.M. Shumaker.** 1999. DNA resequencing, mutation detection and gene expression analysis by oligonucleotide microchips, p. 371-397. *In* J.M. Köhler, T. Mejevaia and H.P. Saluz (Eds.), Microsystem Technology: A Powerful Tool for Biomedical Studies. Birkhäuser Verlag, Basel.
15. **Nikiforov, T.T., R.B. Rendle, P. Goelet, Y.-H. Rogers, M.L. Kotewicz, S. Anderson, G.L. Trainor, and M.R. Knapp.** 1994. Genetic bit analysis: a solid phase method for typing single nucleotide polymorphisms. Nucleic Acids Res. *22*:4167-4175.
16. **Pastinen, T., A. Kurg, A. Metspalu, L. Peltonen, and A.C. Syvänen.** 1997. Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. Genome Res. *7*:606-614.
17. **Pastinen, T., M. Perola, P. Niini, J. Trewilliger, V. Salomaa, E. Vartiainen, L. Peltonen, and A.-C. Syvänen.** 1998. Array-based multiplex analysis of candidate genes reveals two independent and additive genetic risk factors for myocardial infraction in the Finnish population. Hum. Mol. Genet. *7*:1-10.
18. **Phimister, B. (Ed.)** 1999. The Chipping Forecast. Nat. Genet. Suppl. *21*:1-60.
19. **Picoult-Newberg, L., T.E. Ideker, M.G. Pohl, S.L. Taylor, M.A. Donaldson, D.A. Nickerson, and M. Boyce-Jacino.** 1999. Mining SNPs from EST databases. Genome Res. *9*:167-174.
20. **Schena, M., D. Shalon, R.W. Davis, and P.O. Brown.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*:467-470.
21. **Schena, M., R.A. Heller, T.P. Theriault, K. Konrad, E. Lachenmeier, and R.W. Davis.** 1998. Microarrays: biotechnology's discovery platform for functional genomics. Trends Biotechnol. *16*:301-306.
22. **Shumaker, J.M., A. Metspalu, and C.T. Caskey.** 1996. Mutation detection by solid phase primer extension. Hum. Mutat. *7*:346-354.
23. **Solinas-Toldo, S., S. Lampel, S. Stilgenbauer, J. Nickolenko, A. Benner, H. Dohner, T. Cremer, and P. Lichter.** 1997. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer *20*:399-407.
24. **Southern, E., K. Mir, and M. Schepinov.** 1999. Molecular interactions on microarrays. Nat. Genet. *21*:5-9.
25. **Syvänen, A.-C., K. Aalto-Setälä, L. Harju, K. Kontula, and H.A. Söderlund.** 1990. Primer-guided nucleotide incorporation assay in genotyping of apolipoprotein E. Genomics *8*:684-692.
26. **Tabor, S. and C.C. Richardson.** 1995. A single residue in DNA polymerases of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. Proc. Natl. Acad. Sci. USA *92*:6339-6343.
27. **Taillon-Miller, P., E.E. Piernot, and P.Y. Kwok.** 1999. Efficient approach to unique single-nucleotide polymorphism discovery. Genome Res. *9*:499-505.

III

Tõnisson N, Kurg A, Kaasik K, Lõhmussaar E, Metspalu A.
Unravelling genetic data by arrayed primer extension.
Clinical Chemistry and Laboratory Medicine 2000;
38(2):165–70

Clin Chem Lab Med 2000; 38(2):165–170 © 2000 by Walter de Gruyter · Berlin · New York OFFPRINT

Unravelling Genetic Data by Arrayed Primer Extension

Neeme Tõnisson[1,2], Ants Kurg[1], Krista Kaasik[1], Elin Lõhmussaar[1] and Andres Metspalu[1,2]

[1] Institute of Molecular and Cell Biology, University of Tartu, Estonian Biocentre, Tartu, Estonia
[2] Asper Ltd., Tartu, Estonia

We have developed a method for arrayed primer extension (APEX) on an oligonucleotide microchip together with the 4-color fluoresence imaging equipment and supporting software, that allows analysis of the DNA sequence and changes in it. Mutation analysis of BRCA1 gene and single nucleotide polymorphism (SNP) chip for genotyping were used as a model system. Chip surface chemistry, template preparation and APEX reaction conditions were optimised and the assay is ready to be implemented in variety of DNA analysis from SNP testing to DNA resequencing.

Key words: SNP; Genotyping; Oligonucleotide; Microarray; Primer extension.

Abbreviations: APEX arrayed primer extension; LD linkage disequilibrium; PCR polymerase chain reaction; sAP shrimp alkaline phosphatase; SNP single nucleotide polymorphism; SSCP-HA single-stranded DNA conformational polymorphism with heteroduplex analysis; TIR total internal reflection; UNG uracil N-glycosylase.

Introduction to DNA Array Technologies

The first draft of the Human Genome will be completed by spring 2000. Although there will be no "postsequencing era", the rapidly accumulating amount of human genetic data and knowledge about genotype-phenotype relations will certainly cause a rapid increase in the number of genetic tests. Already the publicly available sequence data demand radically new, automatic sample processing and analysis platforms linked with early genetic risk assessment of multifactorial diseases to allow effective drug discovery or personalised medicine. To be useful in this wide field of genetic analysis, DNA array techniques must be specific, robust and cost effective.

Applications and formats of DNA arrays

DNA arrays are devices displaying specific oligonucleotides or longer DNA fragments attached in a two-dimensional order onto an activated solid surface. The two main areas where DNA oligonucleotide arrays are used include DNA variation studies and comparative gene expression (1–3). In this presentation we will focus on arrayed primer extension (APEX) for DNA sequence analysis. This includes mutation detection or resequencing of known genes and genome-wide single nucleotide polymorphism (SNP) testing. Searching disease- associated mutations has been the first and obvious application of oligonucleotide arrays; however, SNP testing to find the genes associated with complex diseases appears to be the most promising one.

Allelic discrimination by enzymatic reactions

All oligonucleotide array platforms utilise complementarity-based target hybridisation (annealing) to arrayed probes. However, the signal-to-noise ratio remains unsatisfactory if there are no additional allelic discrimination mechanisms involved. Discrimination between the genotypes can be significantly enhanced by introduction of enzymes, either DNA polymerase (3, 4), ligase (5, 6) or cleavase (7) as an additional recognition principle.

From all these enzymes, DNA polymerase is the most widely used one.

Arrayed primer extension

APEX is similar to minisequencing (4, 8) or genetic bit analysis (GBA) (9). The APEX method is based on incorporation of four dye terminators into oligonucleotide primers arrayed on the solid surface by a DNA polymerase. In contrast to Sanger dideoxy sequencing technology where one primer is elongated into many fragments and fragment length is used for their separation, APEX uses many primers separated beforehand as an array and extends them by only one base.

All the primers used in APEX assays are based on the wild-type DNA, with their 3'ends complementary to bases just before the called position. Variations of the primer extension assay include locus-specific oligos with generic tags (10) or systems where either gel electrophoresis or mass spectrometry is used for fragment separation and allele identification (8, 11).

Other enzyme-coupled platforms

The oligonucleotide ligation reaction utilises either one (padlock probe type) or two oligonucleotides with the ligation site located at the nucleotide to be identified (5, 6). Both types of ligation assay can probably be adapted to an array format.

A number of assays has been developed by Third Wave Technologies which utilise a family of structure-specific enzymes, called Cleavase™. Cleavase™ cuts junctions between the single- and double-stranded regions of DNA (7). The assays are based on the observation that single strands of DNA form highly individual higher order structures by folding on themselves. Al-

though highly allele-specific, robust and reproducible, the assays are difficult to transform into the array format.

Specific Features of APEX

The APEX reaction combines both a high information content oligonucleotide array and the specificity of molecular recognition by DNA polymerase. All four dye-terminators, used in the same reaction, allow for simultaneous evaluation of the full nucleotide sequence. APEX is capable of identifying different types of mutations and polymorphisms (Figure 1) and can also be used for gene resequencing (12, 13). To test for known mutations, the primers can always be designed to be allele-specific so that discrimination between wild-type and mutated alleles is possible. For unknown mutations, the only option is to use a resequencing assay. This assay type is likely to become one of the most widely used genetic tests for mutation detection in the future.

Hybridisation-based chip platforms need many allele-specific oligos to get valid data. APEX uses only two wild-type oligos per one analysed basepair. The process of data analysis is comparable to that of an automated ABI type DNA sequencer, involving normalised intensity comparisons from different fluorescent labels on the same band/dot.

Oligonucleotides for APEX technology are synthesised with aminolink at their 5'end. The aminolink is connected to a spacer arm of 12 to 18 carbon atoms and ensures covalent immobilisation of the oligonucleotides via their 5'end. The 3'end of the oligonucleotides is left free for extension by DNA polymerase. The spacer length has a definite impact on oligonucleotide accessibility by the template DNA strand and consequently also on their hybridisation properties (14). For resequencing on the APEX assay, successive oligonucleotides are shifted by one nucleotide step in their sequence towards 3'end. Two oligos are used per each base pair (bp), one for the sense and other for the antisense direction. We have used 20- and 25-mer oligonucleotides in our assays. By our experience, the signals from 25-mers are stronger than those of 20-mers. Approximately 10% of oligonucleotides give signals from self-extension as well as from the target DNA and need to be replaced. Although most of the self-extensions are predictable by computer modelling, the final behaviour of arrayed primers can be validated only in the APEX assay.

Surface chemistry used for oligonucleotide arrays

A variety of solid supports (glass, plastics, nylon, polyacrylamide) have been used in the DNA chip preparation. We and many other laboratories have chosen glass microscope slides as the support material, because glass gives a non-porous, relatively homogenous chemical surface, which can be used for modification with silanisation chemistry. The attachment chemistry must meet several criteria, where the accessibility and functionality of the surface-bound DNA for hybridisation, the density of attached oligonucleotides and the reproducibility of the attachment chemistry are most critical. The two most commonly used silane-modified surfaces for DNA array preparation are epoxysilane-modified (15) and aminosilane-modified surfaces (16).

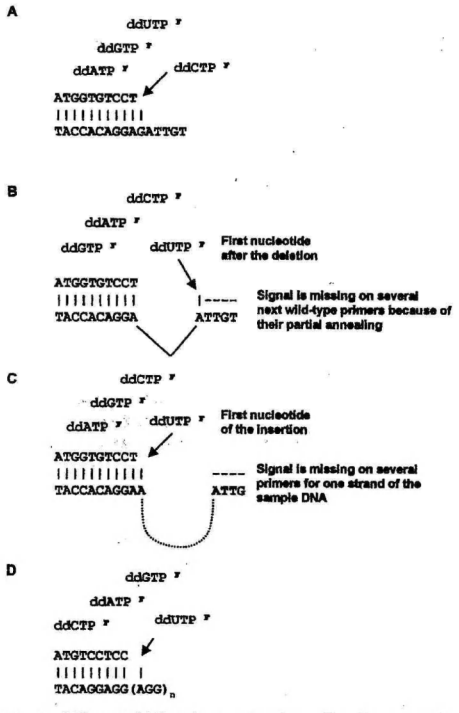


Fig. 1 Different DNA primary structure alterations can be analysed by arrayed primer extension (APEX).
A Point mutations and single nucleotide polymorphism (SNP). All four dye terminators (ddNTP^F) are present in the primer extension reaction and the oligonucleotides will be extended by a complementary nucleotide, which is used for sequence deduction.
B Deletion analysis. Primer before the deletion is extended by dye terminator complementary to the base after the deletion. By analysing both strands, nucleotide sequence before and after the deletion can be obtained. The length of the deletion equals the "footprint" (region with no signal) on the resequencing chip.
C Insertion. By analyzing both strands one can detect the first and the last nucleotide of the inserted sequence. Mono- or dinucleotide insertions are identified completely. Longer insertions produce gap of signals on a resequencing chip, just as deletions. The gap, however does not include the same positions for both strands of the analysed DNA.
D Repeats. Di-, tri-, tetra (etc.) nucleotide repeats can be analysed if repeat length is shorter than oligonucleotide primer on the oligonucleotide array. Long repeats (like in FraX case) are not analysable at the present time using the APEX assay.

Template preparation for APEX

Preparation of template DNA for APEX starts from (multiplex) amplification by polymerase chain reaction (PCR). Both single-and double-stranded target DNA can be used in APEX reactions. Single-stranded DNA has the advantage of being less complex because it lacks the complementary strand, which would compete for hybridisation sites with arrayed primers. The main disadvantages are that preparing sufficient amounts of the single stranded target is laborious and one needs another assay to obtain the second half of the genetic information. The use of a double-stranded target permits identification of changes from both strands simultaneously. An important issue in the APEX assay is the length of the target. We have found that the maximum length of a double-stranded PCR product which can be used directly is approximately 200 base pairs, but 100 bp is optimal. Longer PCR products need fragmentation before use in the APEX reaction. A fraction of the dTTPs is replaced by dUTPs in the amplification mix, allowing for later treatment with uracil N-glycosylase (UNG) (17). UNG is highly specific for uracil bases in DNA and the extent of the reaction can be controlled by dUTP incorporation during PCR. By changing the dUTP fraction, one can vary the mean length of the template DNA fragments. Several other possibilities exist for DNA target fragmentation, like DNaseI treatment, restriction enzyme digestion or mechanical shearing. However, in our hands the UNG treatment has been the most reproducible one.

Single nucleotide extension reaction can only work if no deoxyribonucleotide triphosphates are carried over from the amplification mix. A robust way to inactivate the dNTP leftover from PCR is enzymatic digestion with shrimp alkaline phosphatase (sAP). This is performed simultaneously with UNG treatment and followed by thermal inactivation prior to the APEX reaction (Figure 2).

APEX reaction

Arrayed primer extension is a complex single step reaction consisting of target annealing to the oligonucleotide array and an enzymatic primer extension reaction with fluorescent dideoxy nucleotides. Engineered DNA polymerases (18) are able to incorporate the dye terminators quite efficiently. We use Thermo Sequenase (Amersham Pharmacia Biotech. Inc, Milwaukee, USA) with 20 to 50 picomoles of each fluorescent terminator per reaction. Other commercially available polymerases including AmpliTaq FS (Roche Molecular Systems, Inc., Branchburg, USA) and DyNASeq (Finnzymes OY, Espoo, Finland) are also capable of incorporating labelled dideoxy nucleotides. Template hybridisation to the array and the primer extension reaction are different in their optimal temperatures. Hybridisation is more effective at lower temperatures, whereas polymerase works better at higher temperatures. For this reason, some primer extension protocols are designed in two separate steps, first, hybridisation and second, primer extension by an appropriate

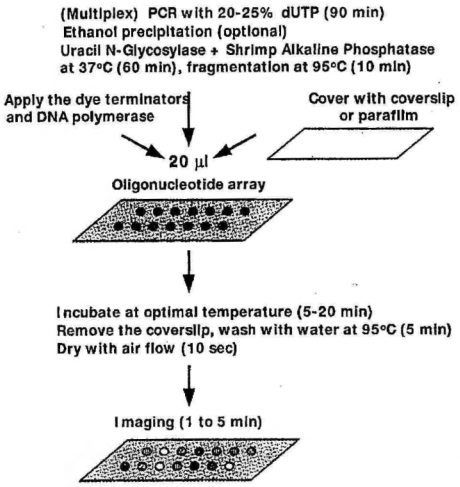


Fig. 2 Single tube, single chip reaction scheme of APEX. It takes about 3 hours (without ethanol precipitation) for the complete test including template amplification by PCR.

DNA polymerase (4). Unbound target is removed by washing between these two steps. Our goal is to make the APEX assay as robust as possible and therefore we have chosen single step reactions. Hybridisation and target-dependent extension of arrayed primers occur simultaneously. The signal-to-noise ratio of APEX is 30/1 in general, but is influenced by all possible primer secondary structures. The APEX reactions are routinely performed at constant temperature (48° C) for 20 minutes. Incorporation of labelled terminators is a very quick reaction, but hybridisation is an equilibrium process and thus it needs time to obtain stronger signals.

The choice of dye labels is determined by two factors: firstly, they must be spectrally well separable and secondly, they must be efficiently incorporated by the DNA polymerase used. Currently we use the following labels (by spectral order): fluorescein, Cy3, Texas Red and Cy5. All but fluorescein are stable to bleaching. Because of the instability of fluorescein, the antifading reagent SlowFade® Light (Molecular Probes, Eugene, USA) is used for imaging.

Arraying and imaging equipment

Current robotic systems are capable of placing up to 10,000 DNA spots per cm^2 (19). However, arrays with a small number of spots (1–2 mm centre to centre distance) have also some value if the aim is to look at the small number of SNPs or at the most common mutations within a single gene. Arraying is done either by mechanical microspotting or ink-jet type piezoelectric dispensing. Most arrayers use different variations of solid pin spotting (20). An interesting modification of

solid pin spotting technology has been developed by Genetic Microsystems, Inc. (owned by Affymetrix, Santa Clara, CA, USA), with a combined pin and ring spotting system. Regular geometry and uniform density of spot are the most critical parameters of the quality of the array.

The imaging equipment should be capable of producing at least 10 x 10 pixels per element if errors from a wrong pixel are expected to be in the range of one percent. This means that the minimal pixel diameter for a 100 μm centre to centre distance array is 10 μm. Most arrayers work in the range of 200 μm between centres, enabling acceptable detector readout at 20 μm pixel size. Commercially available imaging systems mainly use confocal scanning for array imaging (20). This has good sensitivity and dynamic range, but is slower, compared to total internal reflection (TIR)-based imaging.

We have developed an automatic TIR-based system (Genorama 003) for fast analysis of APEX reactions. Four lasers are used to excite the fluorochromes of the arrayed probes. The optical system consists of the light reflectors, scattering elements and cylindrical lenses and transforms the laser beam into a homogeneously illuminated stripe, which enters the slide. Due to total internal reflection on the surfaces of the slide, the incoming beam remains uniform along all the length of the slide. The evanescent light field excites the bound fluorophores residing near the surface of the slide. A gated CCD camera, cooled to –25° C, records the emitted fluorescence. The respective narrow-band interference filters fixed on the revolving wheel in front of the CCD camera depress a noise radiation of laser excitation scattered on the slide. Using one, two or four dye markers, four images corresponding to four laser wavelengths are obtained, one for each dye-labelled ddNTP. Because of precise spectral separation of exciting and emitting wavelengths, the system demonstrates optical cross-talk between the different labels in the range of 2%. The time required for detecting a fluorochrome is currently 10 to 60 seconds.

Genorama software package (Asper Ltd., Tartu, Estonia) is used to convert the fluorescence signals into DNA sequence data. The signal intensities of grayscale pictures from different fluorochromes are first equalised. The strongest signal is the base called. If the next strongest signal from both strands has intensity level higher than 30 to 50% of the strongest signal, the position is called heterozygous. The sequence is compared with a reference one and diverging bases are indicated. All divergences and heterozygous positions are manually verified by comparing signals from different images plus histogram values of the signals.

APEX Applications

Detecting common mutations in the human BRCA1 gene

Mutations in the BRCA1 or the BRCA2 gene increase risk of developing breast or ovarial cancer (21). The familial breast cancer is more polymorphic and contains more mitotic cells than a sporadic cancer. Metastatic breast cancer is also more common in patients with BRCA1 mutations (22). More than 450 mutations have been found in the human BRCA1 gene and over 200 in the BRCA2 gene since 1994 (Breast Cancer Linkage Consortium, Leiden, The Netherlands). There is no clear discrimination between a real polymorphism and a missense mutation increasing risk of developing a cancer (23, 24). The spectrum of mutations varies between populations. For example, 185delAG and 5382insC have been found in 90% of Ashkenazi Jewish families with high risk of breast and ovarian cancer (25)

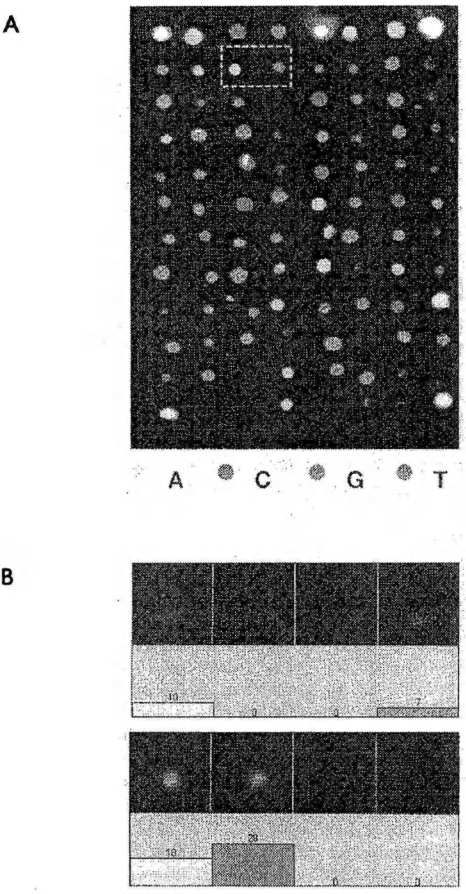


Fig. 3 Example of mutation analysis in the BRCA1 gene.
A Wild-type BRCA1 DNA in APEX reaction. The pseudocolor image shows entire chip scanning for 42 mutation sites across 6 exons. Both sense and antisense strands are analysed. The rectangle shows 185delAG (exon 2) mutation site.
B Heterozygous 185delAG mutation. The upper image shows signals (A and T) from sense and lower (A and C) from antisense strand.

but are less common in other populations. We have studied Estonian families with high risk of breast and ovarian cancer for 42 mutations in the BRCA1 gene. All the subjects were also tested with single-stranded DNA conformational polymorphism with heteroduplex analysis (SSCP-HA). Three BRCA1 gene mutations having positive correlation with breast cancer were found in 3 out of 28 analysed pedigrees. An example of BRCA1 mutation analysis is shown in Figure 3.

SNP testing

SNPs are the most common type of DNA sequence variations. The variations are estimated to occur once every 500 to 1000 bp of the human genome (26). Recent population modeling demonstrated that one has to map up to 500 thousand SNPs from an individual to perform whole genome linkage disequilibrium (LD) analysis (27). SNPs offer several advantages for population-based analysis of the human genome. They are very frequent, allow development of a yes or no type assay and permit easier automation and digitalisation of the data than microsatellite markers.

We are working on an APEX-based assay for genome-wide SNP markers. As a pilot project, we have selected 50 SNPs, two from each chromosome. This number of SNPs is sufficient to be used for human identity testing. We have evaluated the assay with individually amplified SNP-specific templates. A result of the SNP test is shown in Figure 4.

Conclusions

The APEX, described in detail here, can be viewed as an advancement of hybridisation-based array technologies. We have developed a fully integrated genetic testing system consisting of template preparation, multiplex primer extension on the array, fluorescence imaging and data analysis. The APEX approach can be efficiently applied to any other DNA region and is therefore useful both for basic biomedical research and diagnostic assays.

Acknowledgements

The authors wish to thank Heidi Saulep, Viljo Soo, Hendrik Pavel and Jana Zernant for technical assistance, Evgeny Berik (Estla Ltd., Tartu, Estonia) for constructing the TIR-based imag-

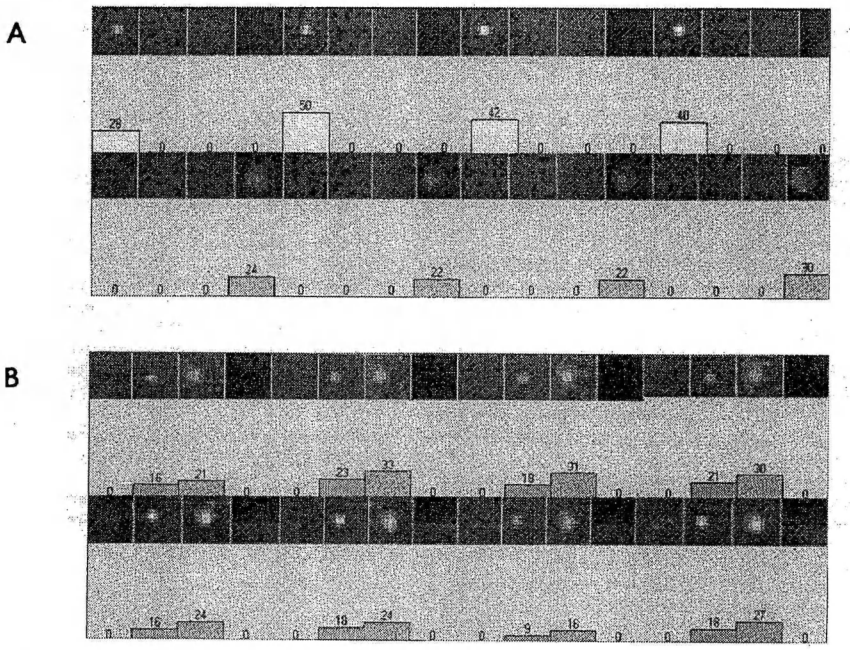


Fig. 4 Fifty SNP loci over the whole genome were selected for analysis.
A Image shows results from a homozygous position. Upper row represents sense strand (A) and lower row antisense strand (T) in the DNA, respectively.
B A heterozygous position is shown. In both alleles, C and G nucleotides were incorporated into sense and antisense strand of DNA. Four parallel analyses on the same chip were carried out.

ing system, and Mrs. Krista Liiv for the artwork. This work was supported by research grants from European Community IC15-CT98-0309, Estonian Science Foundation 2492 and Core Grant 0180518s98 from the Estonian Ministry of Education.

References

1. Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, *et al.* Accessing genetic information with high-density DNA arrays. Science 1996; 274:610-4.
2. Hacia JG, Brody LC, Chee MS, Fodor SP, Collins FS. Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. Nat Genet 1996; 14:441–7.
3. Shumaker JM, Metspalu A, Caskey CT. Mutation detection by solid phase primer extension. Hum Mutat 1996; 7:346–54.
4. Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvanen AC. Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. Genome Res 1997; 7:606–14.
5. Landegren U, Kaiser R, Sanders J, Hood L. A ligase-mediated gene detection technique. Science 1988; 241:1077–80.
6. Gunderson KL, Huang XC, Morris MS, Lipshutz RJ, Lockhart DJ, Chee MS. Mutation detection by ligation to complete n-mer DNA arrays. Genome Res 1998; 8:1142–53.
7. Brow MA, Oldenburg MC, Lyamichev V, Heisler LM, Lyamicheva N, Hall JG, *et al.* Differentiation of bacterial 16S rRNA genes and intergenic regions and Mycobacterium tuberculosis katG genes by structure-specific endonuclease cleavage. J Clin Microbiol 1996; 34:3129–37.
8. Syvanen AC, Aalto-Setala K, Harju L, Kontula K, Soderlund H. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. Genomics 1990; 8:684–92.
9. Nikiforov TT, Rendle RB, Goelet P, Rogers YH, Kotewicz ML, Anderson S, *et al.* Genetic bit analysis: a solid phase method for typing single nucleotide polymorphisms. Nucleic Acids Res 1994; 22:4167–75.
10. Hirschhorn JN, Altshuler D, Sklar P, Bolk S, Lindblad K, Lindgren C, *et al.* Application of high throughput SNP genotyping to complex endocrine diseases [abstract]. Am J Hum Genet 1999; 65:Suppl:A254.
11. Erdos MR, Braun A, Mohlke K, Silander K, Ghosh S, Collins FS. MALDI-TOF mass spectrometry for high throughput SNP detection: application to type 2 diabetes [abstract]. Am J Hum Genet 1999; 65:Suppl:A248.
12. Metspalu A, Saulep H, Kurg A, Tõnisson N. Primer extension from two-dimensional oligonucleotide grids for DNA sequence analysis. In: Dixon GK, Copping LG, Livingstone D, editors. GENOMICS commercial opportunities from a scientific revolution. Oxford, UK: BIOS Scientific Publishers, 1998:217–9.
13. Metspalu A, Shumaker JM. DNA resequencing, mutation detection and gene expression analysis by oligonucleotide microchips. In: Köhler JM, Mejevaia T, Saluz HP, editors. Microsystem technology: a powerful tool for biomedical studies. Basel: Birkhäuser Verlag, 1999:371–97.
14. Southern E, Mir K, Shchepinov M. Molecular interactions on microarrays. Nat Genet 1999; 21(Suppl):5–9.
15. Lamture JB, Beattie KL, Burke BE, Eggers MD, Ehrlich DJ, Fowler R, *et al.* Direct detection of nucleic acid hybridization on the surface of a charge coupled device. Nucleic Acids Res 1994; 22:2121–5.
16. Guo Z, Guilfoyle RA, Thiel AJ, Wang R, Smith LM. Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. Nucleic Acids Res 1994; 22:5456–65.
17. Cronin MT, Fucini RV, Kim SM, Masino RS, Wespi RM, Miyada CG. Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. Hum Mutat 1996; 7:244–55.
18. Tabor S, Richardson CC. A single residue in DNA polymerases of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. Proc Natl Acad Sci U S A 1995; 92:6339–43.
19. Schena M, Heller RA, Theriault TP, Konrad K, Lachenmeier E, Davis RW. Microarrays: biotechnology's discovery platform for functional genomics. Trends Biotechnol 1998; 16:301–6.
20. Phimister BE, editor. The chipping forecast. Nat Genet 1999; 21(Suppl):1–60.
21. Phelan CM, Rebbeck TR, Weber BL, Devilee P, Ruttledge MH, Lynch HT, *et al.* Ovarian cancer risk in BRCA1 carriers is modified by the HRAS1 variable number of tandem repeat (VNTR) locus. Nat Genet 1996; 12:309–11.
22. Xu X, Wagner KU, Larson D, Weaver Z, Li C, Ried T, *et al.* Conditional mutation of BRCA1 in mammary epithelial cells results in blunted ductal morphogenesis and tumour formation. Nat Genet 1999; 22:37–43.
23. Dunning AM, Chiano M, Smith NR, Dearden J, Gore M, Oakes S, *et al.* Common BRCA1 variants and susceptibility to breast and ovarian cancer in the general population. Hum Mol Genet 1997; 6:285–9.
24. Durocher F, Shattuck-Eidens D, McClure M, Labrie F, Skolnick MH, Goldgar DE, *et al.* Comparison of BRCA1 polymorphisms, rare sequence variants and/or missense mutations in unaffected and breast/ovarian cancer populations. Hum Mol Genet 1996; 5:835–42.
25. Fodor FH, Weston A, Bleiweiss IJ, McCurdy LD, Walsh MM, Tartter PI, *et al.* Frequency and carrier risk associated with common BRCA1 and BRCA2 mutations in Ashkenazi Jewish breast cancer patients. Am J Hum Genet 1998; 63:45–51.
26. Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, *et al.* Mining SNPs from EST databases. Genome Res 1999; 9:167–74.
27. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 1999; 22:139–44.

Received 10 November 1999; accepted in present form 10 December 1999

Corresponding author: Andres Metspalu, Institute of Molecular and Cell Biology, University of Tartu, Estonian Biocentre, 23 Riia St., 51010 Tartu, Estonia
Tel.: +372 7 375 029, Fax: +372 7 420 286
Email: andres@ebc.ee

IV

Tõnisson N , Zernant J, Kurg A, Pavel H, Slavin G, Roomere H, Meiel A, Hainaut P and Metspalu A.
Evaluating the arrayed primer extension resequencing assay of TP53 tumor suppressor gene.
Proc Natl Acad Sci USA 2002; 99(8):5503–8

Evaluating the arrayed primer extension resequencing assay of TP53 tumor suppressor gene

Neeme Tõnisson*†, Jana Zernant*, Ants Kurg†, Hendrik Pavel*, Georg Slavin*, Hanno Roomere*†, Aune Meiel*†, Pierre Hainaut‡, and Andres Metspalu*†‡§

*Asper, Ltd., 3 Oru Street, 51014 Tartu, Estonia; †Institute of Molecular and Cell Biology, University of Tartu/Estonian Biocentre, 23 Riia Street, 51010 Tartu, Estonia; and ‡International Agency for Research on Cancer, 150, Cours Albert Thomas, F-69372 Lyon Cedex 08, France

Communicated by C. Thomas Caskey, Cogene Biotech Ventures, Ltd., Houston, TX, February 20, 2002 (received for review July 9, 2001)

Identification of mutations in the tumor suppressor gene TP53 has implications for the molecular epidemiology and for the molecular pathology of human cancer. We have developed and evaluated an arrayed primer extension assay for covering both strands of a region of the coding sequence containing more than 95% of the mutations described so far in TP53. On average, 97.5% of the arrayed TP53 gene sequence can be analyzed from either sense or antisense strands, and 81% from both strands. A patient DNA sample is amplified and annealed to arrayed primers, which then promote DNA polymerase extension reactions with four fluorescently labeled dideoxynucleotides. The TP53 gene chip spans exons 2–9 plus two introns from both strands. The performance of the assay was evaluated by using freshly extracted genomic DNA, as well as DNA extracted from archival (paraffin-embedded) DNA samples. The arrayed primer extension-based TP53 gene test provides an accurate and efficient tool for DNA sequence analysis of this frequently mutated gene for both research and clinical applications.

APEX | oligonucleotide array | chip

The evidence is growing that specific mutations in the TP53 gene can represent important factors for the prognosis of cancer and for the response to various types of cytotoxic therapy. Furthermore, patterns of TP53 mutations have differed considerably from one type of cancer to the other (1–4). However, screening for TP53 mutations gene has yet to become a routine in clinical or epidemiological practice, mainly because current detection technologies are labor-intensive and have prohibitive costs for large-scale prospective studies. Another strong limitation to routine analysis of TP53 mutations resides in the fact that many tumors contain an excess of wild-type TP53 as compared with mutant, resulting from the presence of intact alleles in tumor as well as in noncancer cells (stroma, inflammatory cells, blood vessels).

In this report we describe the development of an arrayed primer extension (APEX) assay for the rapid and sensitive detection and identification of mutations in the TP53 gene. APEX is a genotyping and resequencing technology that combines the advantages of Sanger dideoxy sequencing with the parallelization and high-throughput potential of the microarray format. A DNA sample is amplified, fragmented enzymatically, and annealed to arrayed primers, promoting sites for template-dependent DNA polymerase extension reactions by using four fluorescently labeled dideoxynucleotides. Each base is probed with two primers, one for the sense and another for the antisense strand (5). GENORAMA imaging system and genotyping software (Asper Ltd., Tartu, Estonia, www.asperbio.com) were used for imaging and semiautomatic sequence analysis (Fig. 1).

The principle of sequencing by primer extension on oligonucleotide array has been successfully applied for the systematic identification of all common TP53 mutations in human cancers. The TP53 microarray presented here spans exons 2–9 [containing more than 98% of all mutations described so far in human cancer (6)], together with flanking splice sites and introns 5 and 8 from both strands (total of 1,218 bases; Fig. 2). This system has been designed to allow the detection of most common mutations (missense, nonsense, tandem, insertions, deletions, and complex mutations) and all identified polymorphisms in the TP53 coding sequence. We found that this system allows for sequencing of an average of 97.5% of the arrayed TP53 gene from either sense or antisense strand, whereas 81% of the whole sequence was simultaneously analyzed from both strands. The length of this simultaneous DNA sequence readout (1.2 kb from both strands) outmatches the limits of the current standard for mutation detection, automated dideoxy sequencing. We describe performance of this assay, evaluated by using 100 normal genomic DNA samples from the Estonian population, plus DNA extracted from 11 archival pathology sections (paraffin-embedded resections of primary esophageal cancers), which were demonstrated to contain TP53 mutations by using classical mutation detection methods [temporal temperature gradient gel electrophoresis (TTGE), followed by direct sequencing].

Two silent, six missense, one splice-site mutation, and an insertion were confirmed by both techniques (Table 1). One of the tumors showed a missense mutation at codon 290 by APEX, instead of a silent, point mutation as detected by TTGE plus dideoxy sequencing. In addition, one point mutation, which escaped detection by TTGE plus dideoxy sequencing, was identified by APEX. On the basis of these results, we conclude that the APEX-based TP53 mutation assay provides an accurate and cost-efficient tool for DNA sequence analysis of this frequently mutated gene. Additional oligonucleotides or regions of the TP53 gene can be easily added to the assay. This prototypic assay represents a valuable platform for the development of diagnostic sequencing assays, for TP53 and other genes of interest.

Methods

Template Preparation. Exons 2–9 of the TP53 gene were amplified from genomic DNA in three amplicons: exons 2–4 (with 5′-TGGAAGTGTCTCATGCTGGA and 5′-ATACGGCCAGGCATTGAAGT primers), exons 5–6 (with 5′-TCTGTCTCCTTCCTCTTCCT and 5′-CACTGACAACCACCCTTAAC primers), and exons 7–9 (with 5′-CTCATCTTGGGCCTGTGTTA and 5′-GCCCCAATTGCAGGTAAAAC primers). A 20% fraction of the dTTP in the amplification mixture was substituted by dUTP (5, 7). The amplification products were concentrated and purified by ethanol precipitation in the presence of ammonium acetate. Fragmentation and functional inactivation of the unincorporated dNTPs was achieved in a one-step reaction by addition of shrimp alkaline phosphatase (Amersham Biosciences, Piscataway, NJ) and thermolabile ura-

Abbreviations: APEX, arrayed primer extension; TTGE, temporal temperature gradient gel electrophoresis.

§To whom reprint requests should be addressed. E-mail: andres@ebc.ee.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

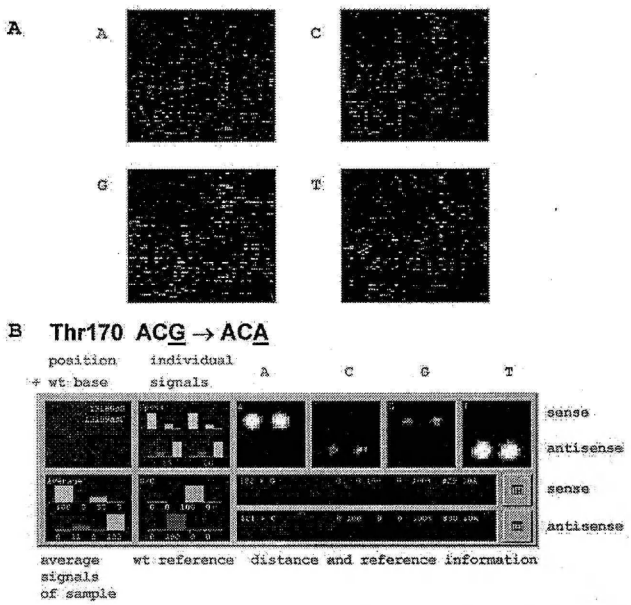


Fig. 1. TP53 APEX-based sequencing assay. (*A*) Grayscale images for each fluorescent dideoxy nucleotide are used for the sequence analysis. (*B*) Silent mutation in the third base of codon 170 of TP53, analyzed by the GENORAMA software. Signals from the analyzed base are averaged and the signal pattern obtained is compared with the wild-type (wt) reference. Grayscale bitmaps corresponding to all four fluorescent dideoxy nucleotides at the base to be determined are shown enabling visual analysis. A signal in the sense area and T in the antisense area are indicative for mutation in the current tumor sample. The distance and reference information consist of: (*i*) the distance measure of the given signal pattern from the wt reference pattern; (*ii*) the wt base with relative signal intensities at four (A, C, G, and T) fluorescence channels; (*iii*) percentage of the signal pattern at the wt reference cluster database for the given base; and (*iv*) index given by the GENORAMA software.

cil *N*-glycosylase (Epicentre Technologies, Madison, WI) (5) and heat treatment.

Oligonucleotide Microchips. APEX primers were designed, according to the wild-type sequence of the human TP53 gene (accession no. U94788) for both sense and antisense directions. The 25-mer oligonucleotides with 12-carbon amino linkers at their 5′ end were obtained from Genset (Paris). Used for spotting the oligonucleotides were 24 × 60 mm aminosilane plus phenylene diisothiocyanate-coated microarray slides (8) (Asper, Ltd.).

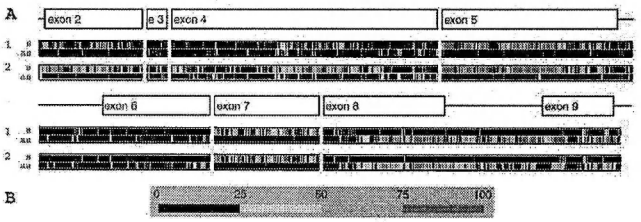


Fig. 2. Performance of the APEX-based sequencing in different regions in the TP53 gene. In some regions of the gene, sense and antisense strands have different performance in APEX. For instance, in exon 7, which can be viewed as an extreme case, the antisense strand signals are very good. At the same time the signals from the sense strand perform below the average level. (*A*) Performance of oligos corresponding to different regions of the p53 gene from sense (s) and antisense (as) strands is shown with data from two independent series of experiments. The upper bar (1) represents APEX performance from 20 repeated APEX reactions with the same wild-type reference DNA, whereas only the automatically clustered signal intensities are used. The lower bar (2) represents visually corrected data from 100 healthy individuals sequenced by APEX. Both of the patterns are highly overlapping. (*B*) Color code and scale for the image. Oligos with signals matching the wt sequence at least in 75% of experiments are shown black. Oligos with zero signals or signals different from the wt sequence are shown yellow, orange, or red.

Table 1. Mutations detected by TTGE and APEX assays

Sample	Mutation: Codon	Mutation: Nucleotide	Mutation: Amino acid	TTGE + sequencing	APEX
8	172	GTT→TTT	Val→Phe	+	−
	315	TCT→TGT	Ser→Cys	−	+
10	175	CGC→CAC	Arg→His	+	+
13	213	CGA→CGG	Arg	+	+
15	170	ACG→ACA	Thr	+	+
16	Intron 5 splice	TAG→TAA		+	+
20	179	CAT→CGT	His→Arg	+	+
22	170	ACG→ACACG	ins 2 bp	+	+*
25	273	CGT→TGT	Arg→Cys	+	+
31	290	CGC→CGA	Arg	+	−
	290	CGC→CCC	Arg→Pro	−	+
48	164	AAG→ACG	Lys→Thr	+	−
	286	GAA→AAA	Glu→Lys	+	+
53	175	CGC→CAC	Arg→His	+	+
	213	CGA→CAA	Arg→Gln	+	+
11		Total		13	12

Concordance with TTGE plus dideoxy sequencing as the reference was 10 of 13. One mutation was identified in the same codon by APEX compared with dideoxy sequencing (sample 31). One mutation was identified by APEX only (sample 8).

*Presence of mutated base (A) determined from the sense strand only.

Primers were diluted to 50 μM concentration in 100 mM carbonate buffer, pH 9.0, and spotted onto the activated surface with Affymetrix 417 arrayer (Affymetrix, Santa Clara, CA). The slides were blocked with 1% ammonia solution and stored at 4°C until needed. Washing steps with 95°C water and 100 mM NaOH were performed before APEX reactions to reduce the background fluorescence and avoid rehybridization of unbound oligonucleotides to the APEX slide.

Genomic DNA Samples from Estonian Population. The genomic DNA samples from healthy individuals were obtained from the Institute of Molecular and Cell Biology, University of Tartu/Estonian Biocentre, and comprised a subset of samples collected within the framework of the project The Influence of Genetical and Environmental Factors on Health of Estonian Population of the Estonian Ministry of Social Affairs. The project had been approved by the ethics committee of University of Tartu. Informed consent was signed by all the participants of the study.

APEX-Based Resequencing. One-third of a product from 50 μl of PCR was used for each primer extension reaction. The APEX mixture consisted of 10 μl of fragmented product, 4 units of Thermo Sequenase DNA polymerase (Amersham Pharmacia), 2 μl of Thermo Sequenase reaction buffer (260 mM Tris·HCl, pH 9.5/65 mM $MgCl_2$) (Amersham Pharmacia), and 2 μM final concentration of each fluorescently labeled ddNTP: Texas Red-ddATP, Cy3-ddCTP, fluorescein-ddGTP, Cy5-ddUTP (Amersham Pharmacia; NEN). The DNA in buffer was denatured at 95°C for 5 min. The enzyme and dye terminators were immediately added to other components, and the whole mixture was applied to prewarmed slides at 58°C. The reactions were allowed to proceed 20 min under parafilm and stopped by washing at 95°C for 2 × 90 s in MilliQ water. A droplet of SlowFade Light Antifade Reagent (Molecular Probes) was applied to the microchips to limit bleaching of the fluorescein. The slides were imaged with the Genorama imaging system (Asper, Ltd.), at 20-μm resolution.

The TP53 gene sequence and mutations were identified by GENORAMA 3.0 genotyping software by using clustered signal patterns from a sequenced wild-type DNA as the statistical reference. The distances of signals from the clusters were used as measures of match with the wild-type gene sequence. Distance (d) of the sample signal pattern compared with the signal patterns in the wild-type reference cluster database were calculated as follows:

$$d = \sqrt{\sum (N_c - N_s)^2},$$

where N_c is the signal intensity of the given nucleotide (A, C, G, and T) in the cluster database, and N_s is the signal intensity of the given nucleotide of the DNA sample.

Results

Oligonucleotide Design. Each base in TP53 is identified by two unique 25-mer oligonucleotides, one for sense and one for antisense strand (total of 2,436 oligonucleotides for the analyzed sequence). The oligonucleotides are based on TP53 wild-type sequence (accession no. U94788), with their 3′ends one base upstream of the base to be identified. The vast majority of these oligonucleotides performs well in APEX. A fraction of the oligonucleotides formed secondary structures, either enabling signals from self-priming or interfering with annealing to test DNA, and therefore needed redesigning. Although the 3′ end and its proximity of the primers cannot be modified, the internal part of the primer may be changed by incorporating a mismatch without seriously affecting the target-specific priming ability. Oligonucleotides for 5.9% of the sequenced bases from either strand were redesigned by introducing a mismatch to reduce the stability of the predicted dimers and avoid self-priming. After modification, 62% of these oligonucleotides generated signals only in the presence of target DNA and not from oligonucleotide dimers; 21% of the modified oligonucleotides did not give any signal either from the target DNA or self-priming because of their reduced hybridization ability; 17% of the modified sequences produced weak or undetectable signals in half of the experiments. None of the modified oligonucleotides generated false-positive signals in the absence of the target DNA.

Some areas of the gene are difficult to sequence from both strands (Fig. 2*A*) for multiple reasons, including sequence repeats, regions with very high GC content, sequences corresponding to oligonucleotide with AT-rich 3′ ends, etc. However, only a very limited number of bases (2.5% on average) were not detected from either strand at the present state of assay development.

Sequence Analysis Algorithm. As a general strategy in APEX, the sequence can be identified either from a single experiment or interpreted on the basis of a statistical analysis. Statistical analysis facilitates identification of deviations from the wild-type reference signal pattern indicative of mutations (Fig. 1*B*). The level of possible secondary signals in the wild-type reference is useful for determining a threshold for acceptance or rejection of signals interpreted as mutations. A sequenced wild-type genomic DNA from a healthy individual was used to create a reference database of signal patterns. The signals from all of the oligonucleotides were analyzed by a clustering algorithm, grouping the signal patterns from four fluorescence channels. Each base in the sample was compared with the wild-type reference, and the value of the distance (see *Methods*) between the signal pattern and the corresponding wild-type base was used as a measure for calling the given base. Zero distance indicates a perfect match between the given base and the wild-type reference base. The analysis was performed in an automated manner, and only a subset of signals needed visual examination.

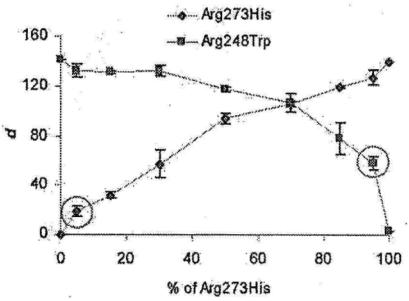


Fig. 3. APEX sensitivity for the fraction of mutated DNA. Relationship between the fraction of mutated DNA and the value of distance measure from the wild-type reference signal pattern. PCR products from two cDNA clones with known missense mutations Arg273His (CGT→CAT) and Arg248Trp (CGG→TGG) were used for the titration. Both mutations are analyzed at the DNA strand with G to A change. Five percent content of mutated DNA is detectable in both cases (indicated by circles). Error bars represent the standard deviations.

APEX Performance in TP53 Sequencing Tests with Numerous Samples. To evaluate performance of the TP53 APEX assay in large-scale studies, 100 normal DNA samples from the Estonian population were tested for common, single-nucleotide polymorphisms and for possible point mutations. A common single-nucleotide polymorphism in exon 4 (Arg-72 → Pro; Arg72Pro) was found with minor allele frequency of 0.26. The identified single-nucleotide polymorphism matches the Hardy–Weinberg equilibrium by the calculated χ^2 value ($P > 0.05$). We also detected two silent point mutations in codons 36 (CCG to CCA) and 139 (AAG to AAA) (6) in two analyzed samples. The first one may correspond to a rare polymorphism, which has been identified in up to 4% of the general population (9). At present, no evidence has been published regarding the status of silent mutation at codon 139, but it cannot be ruled out that it might also correspond to a previously unrecognized, rare polymorphism.

On average, 97.5% of the arrayed TP53 sequence was identified in our current version of the TP53 assay from either sense or antisense strand, and 81% from both strands. In the best cases, respectively, up to 99.8% and 96% of the sequence were analyzed.

Sensitivity for Mutated DNA. DNA extracted from tumor samples always contains a background of normal DNA. APEX sensitivity for the minimal identified percentage of mutated DNA was titrated by mixing PCR products obtained from the mutant (Arg248Trp and Arg273His) TP53 cDNA clones at different ratios (Fig. 3). The mutations are located in different exons, and the clones were therefore used as a competitor fraction of normal DNA for each other. The signal patterns were different from the wild type, and both mutations were detected even if the sample contained as little as 5% of mutant DNA. The samples with zero percent of mutated DNA were matching the wild-type reference DNA (Arg273His with zero distance and Arg248Trp with a distance value of 3). The mixed samples with 5% of mutated DNA, on the contrary, did not match the signal pattern of the reference wild-type sample (average distances, 19 for Arg273His and 58 for Arg248Trp). In fact, 5% of the mutated DNA allowed identification from the analysis software window (Fig. 1*B*) by eye. APEX sensitivity to detect deletions was titrated with del 13–19 TP53 cDNA clone. The first base after deletion is detected instead of the first deleted base (10, 11). The deletion was detected with sensitivity equal to a point mutation

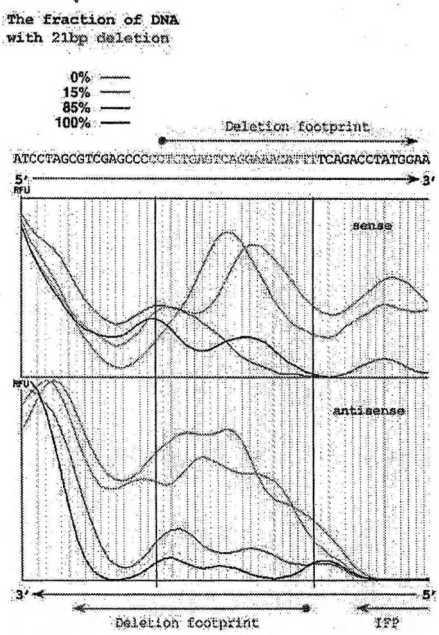


Fig. 4. APEX sensitivity for detecting deletions. Detailed patterns of signal intensities in the deletion area. The actual footprint (with weaker or missing signals) exceeds the deleted sequence by 13 to 15 bases in the 3′ direction from either strand because of partial annealing with the target sequence. The first base after deletion is detected instead of the first deleted base. Because of cDNA used as the template, intron 3 footprint (IFP) is detectable in the antisense strand.

by analyzing first base after the deletion. The complementing algorithm, based on detection of decreased signal intensities and deletion-specific footprint was less sensitive and required at least 15% mutant sequence for detection (Fig. 4). The actual footprint (with missing or weaker signals) exceeds the deletion by 13–15 bases at 3′ direction from either strand because of the partial annealing with the target sequence (Fig. 4).

Blind Test with Tumor Samples. The tumors tested were from a series of squamous cell carcinomas of the esophagus collected in Iran between 1992 and 1998. These cancers often contain TP53 mutations and are very good examples of a type of cancer in which TP53 mutation analysis may have a strong impact in clinical and epidemiological applications. Eleven samples, with a total of 12 point mutations and a 2-bp insertion in TP53, previously identified by TTGE plus manual or automated dideoxy sequencing of the extracted heteroduplex band, were used in a blind test for sensitivity and accuracy of APEX. Sequencing of the heteroduplex band has superior sensitivity to direct sequencing but requires gel purification of the PCR product. The total number of mutations determined was similar in both techniques. Two silent, six missense, one splice site mutation, and an insertion were concordantly identified (Table 1). A missense mutation at the codon 290 was found by APEX instead of a silent point mutation as identified by TTGE plus

dideoxy sequencing. One missense mutation not previously identified by TTGE plus sequencing was *de novo* identified by APEX. Only wild-type APEX signals were present in two samples, where missense mutations were previously determined.

Discussion

A practical approach to TP53 mutation screening has to combine affordable cost, high throughput, high specificity, and high sensitivity. So far, the most advanced, current alternative to dideoxy sequencing is the GENECHIP p53 assay (Affymetrix, www.affymetrix.com), which has been recently evaluated (12–14). The Affymetrix chip has good overall performance but a limited ability to detect deletions and insertions. Promising efforts have been made to couple the oligonucleotide array technology to single-base extension reaction by the DNA polymerase (10). Another recent approach, pyrosequencing, has shown accurate results for detection of mutations in a few exons of TP53 (15).

The currently described APEX-based sequencing approach by comparing a sample with the wild-type reference by the distance measure is comparable with the GENECHIP p53 assay where a score from a mixture of variables between the wild-type reference and a given sample is calculated. The higher the score for a probe set contributing to a given base, the higher the likelihood for the base being mutated (12, 14). In the GENECHIP p53 assay, the single cutoff level for calling mutations has been reported to be unsatisfactory (14). The same situation could apply to the TP53 APEX-based sequencing assay, but further studies are needed to evaluate the possible benefit of approaching each base as a separate entity. The applicable cutoff value for base calling also depends on whether the sample is analyzed for germ-line or somatic mutations. In the current work, prescan of the sequence was made with a general cutoff distance. The positions exceeding the threshold distance from the wild-type signal pattern were visually verified. Just one APEX oligonucleotide per each sequenced base and the general low noise makes possible the fast visual inspection at positions where the software is giving ambiguous results.

The results from the 100 healthy individuals analyzed are encouraging for applying APEX in large-scale TP53 studies, whereas single-nucleotide polymorphism data can have an impact on the analysis of individual risks or of cancer outcome. The identified Arg72Pro polymorphism has recently been proposed to play a role in tumorigenesis. Controversial evidence exists that the Arg-72 allele might be more sensitive to degradation induced by the oncoproteins of human papilloma viruses, suggesting that this polymorphism may predispose to cervical cancer (16). On the other hand, recent studies have shown that the cellular interactions of mutant p53 protein may be different depending of the allelotype of codon 72 (17). The fact that our assay can simultaneously perform mutation detection and correct identification of codon 72 status adds further weight to its usefulness as a one-step assay in clinical or epidemiological studies.

The TP53 detection limit for known alleles was identified as low as in 5%. The actual limit could sometimes be even less than 5%, but in real life the possible alleles are mostly unknown and reliable control and comparison with results obtained with standard methods can be technically difficult because of their own error rates. The dideoxy chain termination sequencing (ref. 18; Fig. 5) and the pyrosequencing (15) are operating at a 30% detection limit of mutation-specific signals. Heteroduplex analysis techniques like TTGE have 10^{-2} sensitivity under optimized conditions (19), but the most commonly used screening method, single-strand conformational polymorphism, has been shown to produce also 5% false-positive results (20). A potential explanation is misincorporation of bases in PCR. Therefore, the fraction of mutated DNA was not further diluted, and the TP53 APEX-based sequencing was evaluated with tumor samples in a blind test.

The total number of mutations determined in 11 esophageal cancer samples was similar by both APEX and TTGE plus

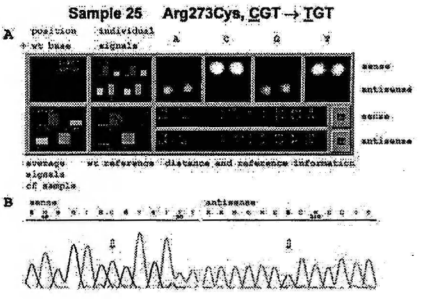


Fig. 5. Missense mutation Arg273Cys CGT→TGT, difficult to identify by automated dideoxy sequencing. (*A*) First base of TP53 codon 273, analyzed by APEX-based sequencing. The signals corresponding to T in the sense strand and A in the antisense strand are indicative for mutation. (*B*) Automated dideoxy sequencing images corresponding to the Arg273Cys mutation from both DNA strands. The indicated mutation-specific peaks are in the range of the background noise and can be easily missed by visual analysis.

sequencing. Two silent, six missense, one splice-site mutation, and an insertion were concordantly identified (Table 1). A missense mutation at the codon 290 was found by APEX instead of a silent-point mutation as identified by TTGE plus dideoxy sequencing. One missense mutation not previously identified by TTGE plus sequencing was *de novo* identified by APEX. Two samples with missense mutations escaped identification by APEX. However, in these specimens, identification of the mutation was possible only by dideoxy sequencing of a PCR product generated from excised TTGE bands with abnormal migration patterns, indicating that mutant DNA was present only in a tiny fraction of the tumor. The latter results suggest that performance and sensitivity of the APEX-based sequencing could be enhanced and all of the mutations possibly identified by use of enrichment techniques such as microdissection of tumor cells from the sample.

In conclusion, we have developed and evaluated an APEX-based sequencing test at the scale of the almost complete TP53-coding sequence, providing an accurate and cost-efficient tool for DNA sequence analysis of this frequently mutated gene. Novel analysis algorithms were developed enabling automatic sequencing. The evaluation test with tumor samples showed performance comparable with one of the most sensitive and also laborious technologies available, dideoxy sequencing of heteroduplex band obtained by TTGE. However, due to the reduced number of steps in template preparation and the possibility of performing automated analysis, APEX is much more suitable for developing tests for high-throughput in clinical diagnostics and large scale epidemiological studies.

Note Added in Proof. When this manuscript was in process, a paper describing the resequencing of exon 7 in the TP53 gene was published (21).

We thank Dr. A. Kristjuhan and V. Jaks, who kindly provided the mutant TP53 cDNA clones; E. Haamer, L. Land, I. Valvas, and V. Soo for technical assistance; Dr. K. Kask for critical reading of the manuscript; and Mrs. Krista Liiv for superior organizing work throughout the whole project. This work was supported in part by research grants from the European Community (IC15-CT98-0309), the Estonian Science Foundation (4479), and the Estonian Ministry of Education (Core Grant 0181518s98). A. Metspalu received the Visiting Scientist Award of the International Agency for Research on Cancer, Lyon, France.

1. Thorlacius, S., Borresen, A. L. & Eyfjord, J. E. (1993) *Cancer Res.* **53,** 1637–1641.
2. Wen, W. H., Reles, A., Runnebaum, I. B., Sullivan-Halley, J., Bernstein, L., Jones, L. A., Felix, J. C., Kreienberg, R., el-Naggar, A. & Press, M. F. (1999) *Int. J. Gynecol. Pathol.* **18,** 29–41.
3. Aas, T., Borresen, A. L., Geisler, S., Smith-Sorensen, B., Johnsen, H., Varhaug, J. E., Akslen, L. A. & Lonning, P. E. (1996) *Nat. Med.* **2,** 811–814.
4. Cabelguenne, A., Blons, H., de Waziers, I., Carnot, F., Houllier, A. M., Soussi, T., Brasnu, D., Beaune, P., Laccourreye, O. & Laurent-Puig, P. (2000) *J. Clin. Oncol.* **18,** 1465–1473.
5. Kurg, A., Tönisson, N., Georgiou, I., Shumaker, J., Tollett, J. & Metspalu, A. (2000) *Genet. Test* **4,** 1–7.
6. Hernandez-Boussard, T., Rodriguez-Tome, P., Montesano, R. & Hainaut, P. (1999) *Hum. Mutat.* **14,** 1–8.
7. Cronin, M. T., Fucini, R. V., Kim, S. M., Masino, R. S., Wespi, R. M. & Miyada, C. G. (1996) *Hum. Mutat.* **7,** 244–255.
8. Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R. & Smith, L. M. (1994) *Nucleic Acids Res.* **22,** 5456–5465.
9. Felix, C. A., Brown, D. L., Mitsudomi, T., Ikagaki, N., Wong, A., Wasserman, R., Womer, R. B. & Biegel, J. A. (1994) *Oncogene* **9,** 327–328.
10. Head, S. R., Rogers, Y. H., Parikh, K., Lan, G., Anderson, S., Goelet, P. & Boyce-Jacino, M. T. (1997) *Nucleic Acids Res.* **25,** 5065–5071.
11. Tönisson, N., Kurg, A., Löhmussaar, E. & Metspalu, A. (2000) in *Microarray Biochip Technology*, ed. Schena, M. (Eaton Publishing, Natick, MA), pp. 247–263.
12. Ahrendt, S. A., Halachmi, S., Chow, J. T., Wu, L., Halachmi, N., Yang, S. C., Wehage, S., Jen, J. & Sidransky, D. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 7382–7387.
13. Wen, W. H., Bernstein, L., Lescallett, J., Beazer-Barclay, Y., Sullivan-Halley, J., White, M. & Press, M. F. (2000) *Cancer Res.* **60,** 2716–2722.
14. Wikman, F. P., Lu, M. L., Thykjaer, T., Olesen, S. H., Andersen, L. D., Cordon-Cardo, C. & Orntoft, T. F. (2000) *Clin. Chem.* **46,** 1555–1561.
15. Garcia, C. A., Ahmadian, A., Gharizadeh, B., Lundeberg, J., Ronaghi, M. & Nyren, P. (2000) *Gene* **253,** 249–257.
16. Storey, A., Thomas, M., Kalita, A., Harwood, C., Gardiol, D., Mantovani, F., Breuer, J., Leigh, I. M., Matlashewski, G. & Banks, L. (1998) *Nature (London)* **393,** 229–234.
17. Tada, M., Furuuchi, K., Kaneda, M., Matsumoto, J., Takahashi, M., Hirai, A., Mitsumoto, Y., Iggo, R. D. & Moriuchi, T. (2001) *Carcinogenesis* **22,** 515–517.
18. Rosenblum, B. B., Lee, L. G., Spurgeon, S. L., Khan, S. H., Menchen, S. M., Heiner, C. R. & Chen, S. M. (1997) *Nucleic Acids Res.* **25,** 4500–4504.
19. Bjorheim, J., Lystad, S., Lindblom, A., Kressner, U., Westring, S., Wahlberg, S., Lindmark, G., Gaudernack, G., Ekstrom, P., Roe, J., *et al.* (1998) *Mutat. Res.* **403,** 103–112.
20. Yuan, B., Thomas, J. P., von Kodolitsch, Y. & Pyeritz, R. E. (1999) *Hum. Mutat.* **14,** 440–446.
21. Shumaker, J. M., Tollet, J. J., Filbin, K. J., Montague-Smith, M. P. & Pirrung, M. C. (2001) *Bioorg. Med. Chem.* **9,** 2269–2278.

V

Gemignani F, Perra C, Landi S, Canzian F, Kurg A, **Tõnisson N**, Galanello R, Cao A, Metspalu A, Romeo G.
Reliable detection of ß-thalassemia and G6PD mutations by a DNA microarray.
Clinical Chemistry 2002; 48(11):2051–4

could lead to their irreversible retention on the LC column, or elution as "ghosts", causing signal suppression in subsequent patient analyses. Urine collected at acidic pH is necessary to suppress the in situ ionization of the carboxyl moiety and facilitates retention of 5-HIAA on the reversed-phase SPE stationary matrix. Preparation of the internal standard (HIAA-d_6) in 10 mL/L acetic acid ensures that every specimen has an adequately acidic pH, eliminating the need to verify and adjust the pH of urine specimens before HPLC analysis.

The rate of repeat analyses for the HPLC method was 15%. Implementation of the LC-MS/MS method with an extended calibration range (0–150 mg/L) reduced the number of repeat analyses from 4% to <1%. The remaining 11% of HPLC repeat analyses were performed as mostly unsuccessful attempts to resolve chromatographic interferences caused by unidentified compounds. Under these circumstances, the specimen was deemed unsatisfactory, and a repeat urine collection was requested. As described for the method comparison, with the LC-MS/MS method, we were able to provide interference-free analytical results for 100% of specimens that had to be excluded because of chromatographic interference in the HPLC analysis.

In summary, we have developed a method for the routine determination of urinary 5-HIAA that overcomes the major limitations of an existing HPLC procedure. In particular, sample preparation is fully automated and analytical time is considerably shorter (2 vs 13 min/sample) with virtually no need for repeat analyses because of chromatographic interference or dilutions.

References

1. Deacon AC. The measurement of 5-hydroxyindoleacetic acid in urine [Review]. Ann Clin Biochem 1994; 31:215–32.
2. Davidson DF, Williamson J. Improved assay for urinary 5-hydroxyindoleacetic acid by HPLC with electrochemical detection [Technical Brief]. Clin Chem 1988;34:768–9.
3. Hyland K. Presentation, diagnosis, and treatment of disorders of monoamine neurotransmitter metabolism. Semin Perinatol 1999; 23:194–203.
4. Feldman J, Lee E. Serotonin content of foods: effect on urinary excretion of 5-hydroxyindoleacetic acid. Am J Clin Nutr 1985;42:639–43.
5. Mashige F, Matsushima Y, Kanazawa H, Sakuma I, Takai N, Bessho F, et al. Acidic catecholamine metabolites and 5-hydroxyindoleacetic acid in urine the influence of diet. Ann Clin Biochem 1996;33:43–9.
6. Magera MJ, Lacey JM, Casetta B, Rinaldo P. Method for the determination of total homocysteine in plasma and urine by stable isotope dilution and electrospray tandem mass spectrometry. Clin Chem 1999;45:1517–22.
7. Magera MJ, Helgeson JK, Matern D, Rinaldo P. Methylmalonic acid measured in plasma and urine by stable-isotope dilution and electrospray tandem mass spectrometry. Clin Chem 2000;46:1804–10.
8. Magera MJ, Stoor A, Helgeson JK, Matern D, Rinaldo P. Determination of homovanillic acid in urine by stable isotope dilution and electrospray tandem mass spectrometry. Clin Chim Acta 2001;306:35–41.
9. Lacey JM, Bergen HR, Magera MJ, Naylor S, O'Brien JF. Rapid determination of transferrin isoforms by immunoaffinity liquid chromatography and electrospray mass spectrometry. Clin Chem 2001;47:513–8.
10. Ford RE, Magera MJ, Kloke KM, Chezick PA, Fauq A, McConnell JP. Quantitative measurement of porphobilinogen in urine by stable-isotope dilution liquid chromatography-tandem mass spectrometry. Clin Chem 2001;47:1627–32.
11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.

Reliable Detection of β-Thalassemia and *G6PD* Mutations by a DNA Microarray, *Federica Gemignani,[1*] Chiara Perra,[3*] Stefano Landi,[1,2*] Federico Canzian,[1] Ants Kurg,[4] Neeme Tõnisson,[4,6] Renzo Galanello,[3] Antonio Cao,[3] Andres Metspalu,[4] and Giovanni Romeo[1,5†]* ([1] IARC, International Agency for Research on Cancer, 150, Cours Albert Thomas, Lyon 69372, France; [2] University of Pisa, Dipartimento di Scienze dell'Uomo e dell'Ambiente, Via S. Giuseppe 22, 56100 Pisa, Italy; [3] Dipartimento di Scienze Biomediche e Biotecnologie, Universita' di Cagliari, Ospedale Regionale Microcitemie, Via Jenner, 09121 Cagliari, Italy; [4] Institute of Molecular and Cell Biology, Estonian Biocentre, University of Tartu, 23 Riia Street, 51010 Tartu, Estonia; [5] Dipartimento di Medicina Interna, Cardioangiologia ed Epatologia, Università di Bologna, Policlinico S. Orsola-Malpighi, via Massarenti 9, 40125 Bologna, Italy; [6] Asper Biotech, Ltd., 3 Oru St., 51014 Tartu, Estonia; * these authors contributed equally to the project; † address correspondence to this author at: Dipartimento di Medicina Interna, Cardioangiologia ed Epatologia, Università di Bologna, Policlinico S. Orsola-Malpighi, via Massarenti 9, 40125 Bologna, Italy; fax 39-051-30-61-71, e-mail romeo@iarc.fr)

β-Thalassemia is an autosomal recessive disorder caused by the absence or reduction of β-globin chain synthesis. There are >400 million β-thalassemia carriers worldwide, and >160 β-thalassemia mutations have been described (*1*). Different populations exhibit a specific subset of mutations, as in Sardinia, where carriers are ~11% of the population and 95% of them present the β^0 39 mutation (*1–3*). In those populations, glucose 6-phosphate dehydrogenase (G6PD) deficiency is also common (*4–6*). For the *G6PD* gene, ~130 mutations or combinations of mutations have been described (*7*), and early detection might reduce the risk of hemolytic crisis in childhood. A program of screening newborns would be desirable in those populations. The molecular diagnosis of β-globin and *G6PD* mutations currently involves a combination of classic methodologies such as restriction fragment length polymorphism analysis, allele-specific oligonucleotide (ASO) hybridization, reverse dot blots, amplification refractory mutation system (ARMS), and direct sequencing (*2, 8–11*). These methods are laborious for large-scale screening.

We set up a microarray-based assay for parallel one-shot detection of 17 mutations commonly found in the Mediterranean population: β^+ −101(C→T); β^+ −87(C→G); β^0 codon 6 (-A); β^0 codon 39 (C→T); β^0-IVSI-1 (G→A); β^+-IVSI-6 (T→C); β^+-IVSI-110 (G→A); β^0-IVSII-1 (G→A); β^+-IVSII-745 (C→G); β^+-IVSII-844 (C→G); *G6PD* A^- variant (202G→A; 376A→G); Mediterranean variant (563C→T); Seattle variant (844G→C); Montalbano variant (854G→A); S. Antioco variant (1342A→G); and Maewo (1360C→T). We called this microarray "Thalassochip".

Thalassochip is based on the arrayed primer extension (APEX) technology (*12*) implemented with allele-specific primed extension (ASPEX) (*13*). APEX consists of a se-

quencing reaction primed by an oligonucleotide anchored to a glass slide (with its 5′ end) terminating just one nucleotide before the mutation site. DNA polymerase extends it by adding one fluorescently labeled dideoxynucleoside triphosphate complementary to the variant base. The reading of the incorporated fluorescence identifies the base in the target sequence. In ASPEX, the oligonucleotide ends with the variant base (ASO). Two pairs of oligonucleotides are needed for forward (F) and reverse (R) strands for both the mutant (M) and the wild-type (WT) alleles (ASO-M-F, ASO-M-R, ASO-WT-F, and ASO-WT-R, respectively). The extension occurs only when ASO completely matches to the target. Fig. 1 shows two examples. In this report we describe a validation study of Thalassochip.

We studied 117 individuals from Sardinia who were referred to the hematologic service of the Ospedale Regionale Microcitemie, Cagliari, Italy. We used classic routine methods considered "gold standards" [restriction fragment length polymorphism analysis, ARMS, reverse dot blots, and direct sequencing, *(3, 9, 14)*]. All participants gave informed consent in accordance with the Helsinki Declaration. Participants were selected to cover the mutations of Thalassochip for validation purposes. Blood samples were also analyzed for G6PD activity *(15)*.

In the present study, the *β-globin* gene was amplified from human genomic DNA in two fragments and the *G6PD* gene in five fragments as described elsewhere (box I in the supplemental data, available with the online version of this Technical Brief at http://www.clinchem.org/content/vol48/issue11/). One unique protocol and thermocycling profile was used for all seven PCRs. 5′ (C-12) aminolinker oligonucleotides were purchased from Sigma Genosys Ltd and spotted onto silanized slides as reported elsewhere [Refs. *(16–18)* and box III in the supplemental data]. A detailed list is given in Table 1 of the supplemental data. PCR products were pooled together, purified, and concentrated on Millipore Y30 columns, and 15 μL of eluate was collected. The PCR products were reduced in size by fragmentation to allow better hybridization with arrayed oligonucleotides. We treated 15 μL of purified PCR products (performed with 200 μM dATP, dCTP, and dGTP; 160 μM dTTP; and 40 μM dUTP) with 1 U of uracil *N*-glycosylase (Epicentre Technologies) and 1 U of shrimp alkaline phosphatase (Amersham Biosciences). The mixture was incubated at 37 °C for 3 h and then at 95 °C for 30 min to denature DNA with abasic sites and inactivate the uracil *N*-glycosylase and shrimp alkaline phosphatase. More details are provided in box III of the supplemental data. APEX and ASPEX work at the same time on the same slide with the reagent mixture and templates. Briefly, the 20-μL reaction mixture, containing fluorescently labeled dideoxynucleoside triphosphates (50 pmol of each), 10× buffer, fragmented PCRs (9 μL), and 4 U of Thermo Sequenase (Amersham Biosciences), was placed on the spotted slide and incubated for 25 min at 58 °C. Slides were washed, a droplet of SlowFade® Light Antifade Reagent (Molecular Probes) was added to limit the bleaching of fluorescein, and the slide was imaged on a Genorama-003 four-color detector equipped with Genorama image analysis software (Asper Biotech). More details about the APEX/ASPEX reaction as well as the detection of signals are given in box II in the supplemental data.

Each mutation is specified by a pattern of six different oligonucleotides. Two positions are for APEX, and four positions are for ASPEX. A simple algorithm has been established for combining the six signals to give the final genotype. Two examples are shown in Fig. 1. Shown in Fig. 1A is a WT homozygote for β IVS I-110. The WT homozygote gives two ASO-WT signals and two APEX signals for WT genotype. The four signals correspond to the WT genotype. No signal corresponds to the mutant allele. This condition is scored as 4/0. A heterozygote (carrier) for β IVS I-1 (shown in Fig. 1B) produces four signals from ASPEX and four signals from APEX. The score is 4/4. A mutant homozygote is identified by four signals from the mutant allele and will show a score of 0/4. Intermediate scores such as 4/2, 2/4, 3/2, and 2/3 were considered "heterozygotes"; 4/1, 3/1, 3/0, and 2/0 were considered WT homozygotes; and 1/4, 1/3, 0/3, and 0/2 were considered mutant homozygotes. Scores of 1/0, 0/1, 2/1, and 1/2 were considered inadequate for diagnosis (marked as N/A). We calculated the 95% confidence intervals, comparison proportions, sensitivities, and specificities by binomial distribution, considering the gold standard as 100%. Statistical analyses were carried out with Statgraphics plus 2.1 for Windows (Manugistic Inc.).

In our sample set, we were not able to analyze the mutant alleles β-IVSII-844 and G6PD854 because of their rarity in the Mediterranean area. In most cases, one or two of six oligonucleotides were enough to determine the genotype correctly. Twenty-seven oligonucleotides showed 100% correct extensions, and other 42 showed correct extensions in >90% of the samples. This means that, in the future, it will be possible to reduce the size of the microarray just picking the two to three best oligonucleotides within each mutation. Details about each specific oligonucleotide are reported in Table 2 of the supplemental data. Thalassochip and the gold standards are compared in Table 1. Of 1989 genotypes, only 9 were called incorrectly, making the overall error rate 0.45%. All nine mistakes were samples misclassified as heterozygotes (six WT and three homozygotes) and would have been retested with different methods. β-IVS II-745 demonstrated a lack of signal in most cases, which was reflected in a lower rate of detection (see Table 1). The region surrounding the β-IVS II-745 mutation is AT rich, and 58 °C was probably too high for hybridization. Further trials with longer APEX and ASPEX oligonucleotides seemed to solve this problem (data not shown). Redesigning of oligonucleotides was also helpful in other situations (data not shown), leading to the idea that melting temperature could be one of the critical factors for the assay. Alternatively, longer oligonucleotides could allow a more stable interaction between the target, the oligonucleotide, and the Thermo Sequenase.

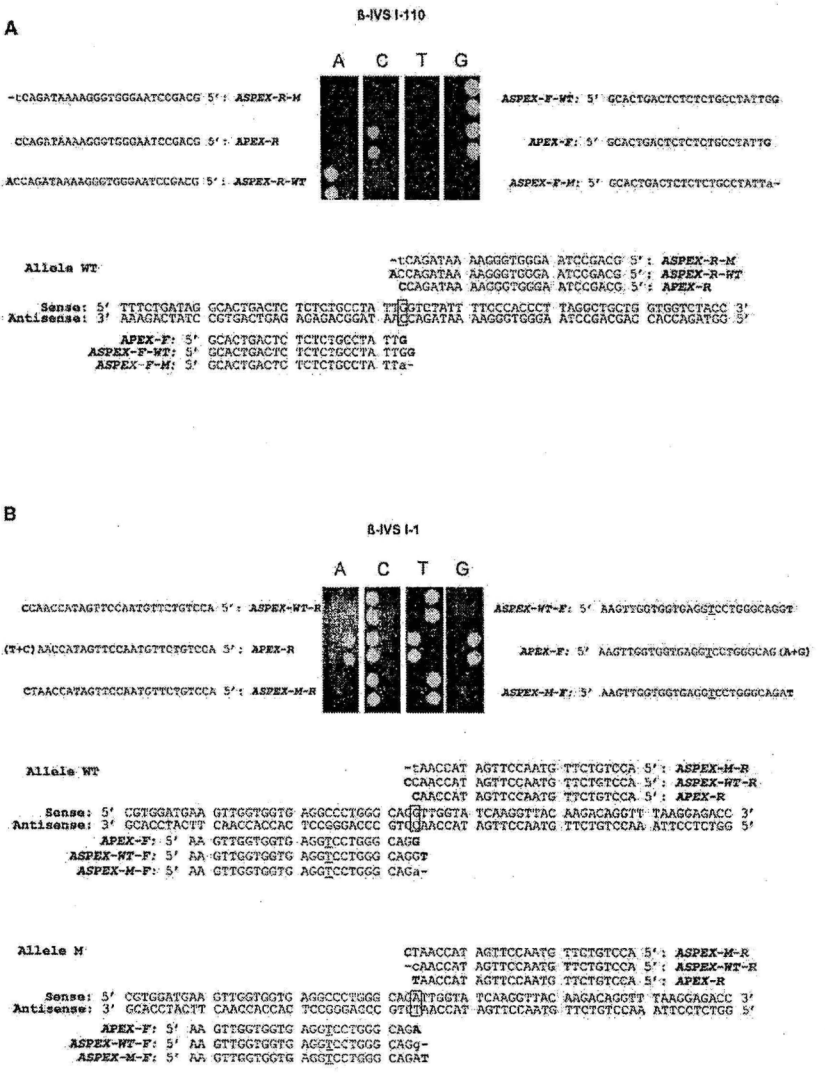


Fig. 1. Examples of ASO with complete matching.

(*A*), WT homozygote for β-IVS I-110. The WT homozygote produces two ASO-WT signals and two APEX signals. (*B*), heterozygote (carrier) for β-IVS I-1. The heterozygote produces four signals from ASPEX and four signals from APEX. *Boxed nucleotides* indicate the variant position; *underlined nucleotides* indicate mismatched bases for weakening secondary structures of nucleotides; *bolded nucleotides* indicate the bases extended by Thermo Sequenase.

Table 1. Comparison between analysis with standard techniques and Thalassochip.[a]

Mutation	Results from gold standard: WT homozygous	Heterozygous	Mutant homozygous
β-*globin* −101	109 (1)	7	0
β-*globin* −87	107 (4)	6	0
β-*globin* codon 6	101	10	6[b]
β-*globin* IVSI-1	107 (2)	6	2
β-*globin* IVSI-6	101 (1)	12	3
β-*globin* IVSI-110	107 (1)	6	3
β-*globin* codon 39	84[b]	24	9
β-*globin* IVSII-1	115	2	0
β-*globin* IVSII-745	108 (21)	9 (3)	0
β-*globin* IVSII-844	117	0	0
G6PD-202	115 (6)	1	1
G6PD-376	113	3	1
G6PD-563	97[c] (4)	10	10
G6PD-844	109[d] (2)	3	5
G6PD-854	117 (1)	0	0
G6PD-1342	105[d] (1)	2	10
G6PD-1360	115	0	2

[a] Absolute numbers of patients tested are reported. Results were obtained after evaluating the whole panel of six oligonucleotides (ASO-WT-F, ASO-WT-R, ASO-M-F, ASO-M-R, APEX-F, and APEX-R) according to the criteria specified in the text. The number of people with no results is indicated in parentheses.

[b–d] No. of volunteers misclassified as heterozygotes: [b] two; [c] three; [d] one.

The specificity of Thalassochip was 100% for 13 mutations, with the lowest sensitivity being 98.5% for G6PD563. Sensitivity was 100% for 14 mutations and 90.9% for β-codon 6. However, two carriers and two homozygotes for hemoglobin S (HbS) were correctly detected by APEX codon 6 oligonucleotides (data not shown), despite the fact that oligonucleotides for codon 6 were not specifically designed to detect HbS. More detailed information about the sensitivity and specificity is provided in Table 3 of the supplemental data.

Redundancy is important in microarray technology, both from a statistical point and for the correct interpretation of signals. Many different technologies require a redundancy of targets, which leads to more accurate responses with increasing numbers of signals. In our hands, six oligonucleotides per mutation were highly reliable for analysis, and we could have reached the same results with a lower number of oligonucleotides. The redundancy allowed us to circumvent artifacts such as lack of signals and nonspecific extensions.

Our results, along with the low reagent costs (approximately US $0.50/genotype) and short processing time (2 days for 30 samples testing for 17 mutations) for the microarray, indicate a potential use of this technology for screening programs. This approach may be expanded to other common diseases, such as α-thalassemia, Wilson disease, or cystic fibrosis.

Dr. Landi is a recipient of a Marie Curie fellowship (IHP-MCIF-99-1) of the European Commission. Dr. Gemignani is recipient of a special training award (STA) by the International Agency for Research on Cancer. This work was supported in part by grants from the foundation "Fondazione Italiana Leonardo Giambrone per la Guarigione dalla Thalassemia". Andres Metspalu was supported by a core grant from the Estonian ministry of education (No. 0180518s98) and Estonian SF Grant 4479 and was the recipient of the IARC senior fellowship.

References

1. Thein SL. β-Thalassaemia. In: Higgs DR, Weatherall DJ, eds. Bailliere's clinical haematology: sickle cell disease and thalassemia, Vol. 11. London: Bailliere Tindall, 1998:91–126.
2. Cao A, Galanello R, Rosatelli MC. Prenatal diagnosis and screening of the hemoglobinopathies In: Higgs DR, Weatherall DJ, eds. Bailliere's clinical haematology: sickle cell disease and thalassemia, Vol. 11. London: Bailliere Tindall, 1998:215–38.
3. Rosatelli MC, Dozy A, Faa V, Meloni A, Sardu R, Saba L, et al. Molecular characterization of β-thalassemia in the Sardinian population. Am J Hum Genet 1992;50:422–6.
4. Mehta A, Mason PJ, Vulliamy TJ. Glucose-6-phosphate dehydrogenase deficiency. Best Pract Res Clin Haematol 2000;13:21–38.
5. Gloria-Bottini F, Falsi AM, Mortera J, Bottini E. The relations between G-6-PD deficiency, thalassemia and malaria. Further analysis of data from Sardinia and the Po Valley. Experientia 1980;36:541–3.
6. Ninfali P, Baronciani L, Ruzzo A, Fortini C, Amadori E, Dall'ara G, et al. Molecular analysis of G6PD variants in northern Italy: a study on the population from the Ferrara district. Hum Genet 1993;92:139–42.
7. Vulliamy T, Luzzato L, Hirono A, Beutler E. Hematologically important mutations: glucose-6-phosphate dehydrogenase. Blood Cells Mol Dis 1997; 23:302–13.
8. Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res 1989;17:2503–16.
9. Martinez di Montemuros F, Dotti C, Tavazzi D, Fiorelli G, Cappellini MD. Molecular heterogeneity of glucose-6-phosphate dehydrogenase (G6PD) variants in Italy. Hematologica 1997;82:440–5.
10. Old JM, Varawalla NY, Weatherall DJ. Rapid detection and prenatal diagnosis of β-thalassaemia: studies in Indian and Cypriot populations in the UK. Lancet 1990;336:834–7.
11. Lindeman R, Hu SP, Volpato F, Trent RJ. Polymerase chain reaction (PCR) mutagenesis enabling rapid non-radioactive detection of common β-thalassaemia mutations in Mediterraneans. Br J Haematol 1991;78:100–4.
12. Shumaker JM, Metspalu A, Caskey CT. Mutation detection by solid phase primer extension. Hum Mutat 1996;7:346–54.
13. Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, Syvanen AC. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. Genome Res 2000;10:1031–42.
14. Sanger F, Micklen S, Coulson AR. DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci U S A 1977;74:5463–7.
15. Mosca A, Paleari R, Rosti E, Luzzana M, Barella S, Sollaino C, et al. Simultaneous automated determination of glucose 6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase activities in whole blood. Eur J Clin Chem Clin Biochem 1996;34:431–8.
16. Guo Z, Guilfoyle RA, Thiel AJ, Wang R, Smith LM. Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. Nucleic Acids Res 1994;22:5456–65.
17. Metspalu A, Shumaker JM. DNA resequencing, mutation detection and gene expression analysis by oligonucleotide microchips. Biomethods 1999;10: 371–97.
18. Kurg A, Tonisson N, Georgiou I, Shumaker J, Tollett J, Metspalu A. Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology. Genet Test 2000;4:1–7.

CURRICULUM VITAE

Name	Neeme Tõnisson
Date of birth	17.02.1971
Position at work	Scientist
Address	Asper Biotech Ltd.
	3 Oru St., Tartu 51014, Estonia.
Phone:	+372 7 441 556
Fax:	+372 7 442 343
E-mail:	neemet@asperbio.com
Marital status	Married, two children
Citizenship	Estonian

Educational history

Tartu Secondary School No 15 — 1989
M.D., University of Tartu. Date of graduation — June 1995.
Visiting student at the Institute of Developmental Biology, University of Goettingen, Germany, fellowship from DAAD, September — October 1995.
Ph.D. student, University of Tartu, September 1995 — present.
European School of Medical Genetics, 12th course, March 1999.

Professional history

1995–1996	Assistant at the *In vitro* fertilisation programme of Women's Clinic, University of Tartu, Estonia.
1996–1999	Development of primer extension technology on oligonucleotide array for DNA sequence analysis at the Institute of Molecular and Cell Biology, University of Tartu, Estonia
1999–2001	Head of Research, Asper Ltd.
2001–2002	Director of Genotyping Applications, Asper Biotech Ltd.

Teaching

1997–1999	Instructor in practical courses of Molecular Biotechnology of Tartu University.
2001–2002	Instructor in ESF/EBC international course "APEX on DNA microarrays"

Scientific Organizations

Member of Estonian Society of Human Genetics

Research interests

Genetics of complex diseases
High-throughput genotyping technologies

LIST OF PUBLICATIONS

1. Metspalu A, Saulep H, Kurg A, Tõnisson N. Primer extension from two-dimensional oligonucleotide grids for DNA sequence analysis. Genomis: Commercial Opportunities from a Scientific Revolution (ed. by Dixon GK, Copping LG, Livingstone D), BIOS Scientific Publishers 1998: 217–9
2. Kurg A, Tõnisson N, Metspalu A, Berik E. Laser diagnostic system for rapid mutation identification. Medical and Biological Engineering and Computing 1999; 37suppl.1:311
3. Kurg A, Tõnisson N, Tollett J, Georgiou I, Shumaker J, Metspalu A. Arrayed Primer Extension: Solid phase four-color DNA resequencing and mutation detection technology. Genetic Testing 2000; 4(1):1–7
4. Tõnisson N, Kurg A, Lõhmussaar E, Metspalu A. Arrayed primer extension on the DNA chip — method and applications. Microarray Biochip Technologies (M. Schena, ed.) Biotechniques Books. Eaton Publishing, Natick MA 2000: 247–63
5. Tõnisson N, Kurg A, Kaasik K, Lõhmussaar E, Metspalu A. Unravelling genetic data by arrayed primer extension. Clinical Chemistry and Laboratory Medicine 2000; 38(2):165–70
6. Tõnisson N , Zernant J, Kurg A, Pavel H, Slavin G, Roomere H, Meiel A, Hainaut P and Metspalu A. Evaluating the arrayed primer extension resequencing assay of TP53 tumor suppressor gene. Proc Natl Acad Sci USA 2002; 99(8):5503–8
7. Dawson, E., Abecasis, G. R., Bumpstead, S., Chen, Y., Hunt, S., Beare, D. M., Pabial, J., Dibling, T., Tinsley, E., Kirby, S., Carter, D., Papaspyridonos, M., Livingstone, S., Ganske, R., Lohmussaar, E., Zernant, J., Tonisson, N., Remm, M., Magi, R., Puurand, T., Vilo, J., Kurg, A., Rice, K., Deloukas, P., Mott, R., Metspalu, A., Bentley, D. R., Cardon, L. R. & Dunham, I. A first-generation linkage disequilibrium map of human chromosome 22. Nature 2002;418;544–8
8. Gemignani F, Perra C, Landi S, Canzian F, Kurg A, Tõnisson N, Galanello R, Cao A, Metspalu A, Romeo G. Reliable detection of ß-thalassemia and G6PD mutations by a DNA microarray. Clin Chem 2002;48(11):2051–4

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida.* Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand**. The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995, 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996, 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996, 60 p.
19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996, 150 p.

20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996, 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu, 1996, 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996, 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997, 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997, 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997, 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997, 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997, 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997, 80 p.
29. **Arvo Tuvikene**. Assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro*. Tartu, 1997, 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997, 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997, 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997, 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu, 1998, 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998, 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998, 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998, 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous crassland plant species. Tartu, 1998, 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998, 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998, 98 p.
40. **Arnold Kristjuhan.** Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998, 86 p.
41. **Sulev Ingerpuu.** Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998, 163 p.

42. **Veljo Kisand.** Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998, 118 p.
43. **Kadri Põldmaa.** Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998, 178 p.
44. **Markus Vetemaa.** Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998, 117 p.
45. **Heli Talvik.** Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998, 104 p.
46. **Katrin Heinsoo.** Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1999, 133 p.
47. **Tarmo Annilo.** Studies on mammalian ribosomal protein S7. Tartu, 1998, 77 p.
48. **Indrek Ots.** Health state indicies of reproducing great tits (*Parus major*): sources of variation and connections with life-history traits. Tartu, 1999, 117 p.
49. **Juan Jose Cantero.** Plant community diversity and habitat relationships in central Argentina grasslands. Tartu, 1999, 161 p.
50. **Rein Kalamees.** Seed bank, seed rain and community regeneration in Estonian calcareous grasslands. Tartu, 1999, 107 p.
51. **Sulev Kõks.** Cholecystokinin (CCK) — induced anxiety in rats: influence of environmental stimuli and involvement of endopioid mechanisms and erotonin. Tartu, 1999, 123 p.
52. **Ebe Sild.** Impact of increasing concentrations of O_3 and CO_2 on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva.** Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna.** Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro.** Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane.** Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
57. **Tiina Tamm.** Cocksfoot mottle virus: the genome organisation and translational strategies. Tartu, 2000, 101 p.
58. **Reet Kurg.** Structure-function relationship of the bovine papilloma virus E2 protein. Tartu, 2000, 89 p.
59. **Toomas Kivisild.** The origins of Southern and Western Eurasian populations: an mtDNA study. Tartu, 2000, 121 p.
60. **Niilo Kaldalu.** Studies of the TOL plasmid transcription factor XylS. Tartu 2000. 88 p.
61. **Dina Lepik.** Modulation of viral DNA replication by tumor suppressor protein p53. Tartu 2000. 106 p.
62. **Kai Vellak.** Influence of different factors on the diversity of the bryophyte vegetation in forest and wooded meadow communities. Tartu 2000. 122 p.
63. **Jonne Kotta.** Impact of eutrophication and biological invasionas on the structure and functions of benthic macrofauna. Tartu 2000. 160 p.

64. **Georg Martin.** Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000. 139 p.
65. **Silvia Sepp.** Morphological and genetical variation of *Alchemilla L.* in Estonia. Tartu, 2000. 124 p.
66. **Jaan Liira.** On the determinants of structure and diversity in herbaceous plant communities. Tartu, 2000. 96 p.
67. **Priit Zingel.** The role of planktonic ciliates in lake ecosystems. Tartu 2001. 111 p.
68. **Tiit Teder.** Direct and indirect effects in Host-parasitoid interactions: ecological and evolutionary consequences. Tartu 2001. 122 p.
69. **Hannes Kollist.** Leaf apoplastic ascorbate as ozone scavenger and its transport across the plasma membrane. Tartu 2001. 80 p.
70. **Reet Marits.** Role of two-component regulator system PehR-PehS and extracellular protease PrtW in virulence of *Erwinia Carotovora* subsp. *Carotovora*. Tartu 2001. 112 p.
71. **Vallo Tilgar.** Effect of calcium supplementation on reproductive performance of the pied flycatcher *Ficedula hypoleuca* and the great tit *Parus major,* breeding in Nothern temperate forests. Tartu, 2002. 126 p.
72. **Rita Hõrak.** Regulation of transposition of transposon Tn*4652* in *Pseudomonas putida*. Tartu, 2002. 108 p.
73. **Liina Eek-Piirsoo.** The effect of fertilization, mowing and additional illumination on the structure of a species-rich grassland community. Tartu, 2002. 74 p.
74. **Krõõt Aasamaa.** Shoot hydraulic conductance and stomatal conductance of six temperate deciduous tree species. Tartu, 2002. 110 p.
75. **Nele Ingerpuu.** Bryophyte diversity and vascular plants. Tartu, 2002. 112 p.

ISSN 1424–6479
ISBN 9985–56–703–x