Microarray-based genetic analyses for studying susceptibility to arterial and venous thrombotic disorders

Eleanor S. Pollak*§, Lana Feng**, Halleh Ahadian**, Paolo Fortina§

*Department of Pathology and Laboratory Medicine, §Department of Pediatrics, University of Pennsylvania School of Medicine and The Children's Hospital of Philadelphia, Philadelphia, PA, **Nanogen, Inc., San Diego, CA, USA

Key words:
Microarray; Molecular screening; Single nucleotide polymorphism; Thrombosis.

We describe the potential of microarray technology for parallel, high-throughput approaches to molecular detection of DNA variations associated with progression to arterial and venous thrombotic diseases. The use of the newly commercialized NanoChip™ platform and a framework for genetic screening with a list of potential genetic targets useful to the evaluation of cardiovascular risk are presented. Implementation in clinical setting of analytical silicon and glass microarrays will facilitate early diagnosis, help direct specific advice to high-risk individuals and evaluate treatment strategies.

(Ital Heart J 2001; 2 (8): 568-572)

Address:

Paolo Fortina, MD

Department of Pediatrics
The Children's Hospital of Philadelphia
310-C Abramson Pediatric Research Center
34th Street and Civic Center Boulevard
Philadelphia, PA 19104
USA
E-mail: fortina@mail.med.upenn.edu

Medical progress has allowed us to understand a vast number of diseases that can be diagnosed on the basis of mutations in one gene causing mutations in one protein that result in a single disease. However, a majority of diseases are not inherited in a simple one-gene Mendelian pattern. We have now advanced to where one must analyze multiple genes and combine this information for diagnosis and therapy for diseases involving multiple genetic contributions. Microarray-based genetic analysis affords us the potential to apply such information to study a patient's disease susceptibility. A prime target for such analysis is cardiovascular disease. This brief report will provide an overview of what microarray-based technology can readily offer us in the year 2001. Specifically, we will address the following issues: 1) human sequence variation; 2) use of microarrays in detection of sequence variation; 3) sequence variation in genes involved in thrombotic disorders.

Human sequence variation

The human genome consists of 3 billion nucleotides, and differences in a single nucleotide (single nucleotide polymorphism-SNP) occur approximately once every 1000 nucleotides[1]. These SNPs serve as useful markers for identifying mutations causing specific diseases. However, SNPs are more frequently used as markers linked to the risk of developing a specific disease, but not themselves contributing to a disease causing genetic change. Eventually, it will be possible to screen people to discover their unique disease susceptibilities[2]. Therefore, each SNP can act as a landmark allowing us to pinpoint a region of interest for investigating whether a given region is associated with a specific disease. DNA microarrays, also called DNA chips, have become powerful tools in navigating through these DNA landmarks in both clinical genetics and biomedical research[3]. They can also be used in a wide variety of applications including monitoring gene expression, diagnosing genetic mutation and analyzing loss of heterozygosity. Microchip devices provide a relatively simple means of performing tens to hundreds of thousands of analytical reactions using a relatively small test instrument. Researchers using these devices in genomic studies are assigning function to the estimated 35 000 or so human genes. In addition, data from the microarray technology is being used to expand our knowledge and understanding of genomic DNA sequence variation[4].

Use of microarrays in the detection of sequence variation

A variety of DNA chips capable of distinguishing single base variations has been designed and several platforms are now commercially available. The two common denominators in SNP chip-based technolo-

gy are the hybridization of labeled targets and polymerase extension of arrayed oligoprobes. More specifically, DNA chips contain surface-bound probes that are used in hybridization reactions in a highly parallel fashion. The different methods for producing arrays comprise two broad categories: 1) deposition of pre-synthesized probes (antibodies, oligonucleotides, cDNAs, polymerase chain reaction-PCR products), and 2) *in situ* synthesis of peptide and oligonucleotide probes at individual locations on the array. Alternative protocols have also been described that overcome ambiguities in distinguishing homozygotes from heterozygotes[5]. It is anticipated that a catalog characterizing human sequence variations associated with variable phenotypes will be one of the most valuable resources of the human genome project. However, it is of utmost importance to design DNA microarrays that are easily customizable so that target samples can be analyzed at multiple loci to meet today's demands of genetic analyses of unprecedented volumes and diversity. In addition, it is critical that the chips be easy to manufacture and economically advantageous compared to current established technology. When compared to current laboratory methods, DNA microarray technology already demonstrates advantages for reduction in cost, analysis time and reagent consumption.

An example of SNP/mutation detection employing an electronic dot-blot like assay is shown in figure 1. Briefly, DNA is isolated from peripheral blood leukocytes. Following PCR amplification with a forward and a biotinylated reverse primer, PCR products of a region encompassing 3 single-base changes corresponding to positions -654 (C/T), -641 (A/G) and -476 (A/T) in the promoter region of the protein C gene are analyzed using the NanoChip™ platform (Nanogen Inc., San Diego, CA, USA). The system includes a chip reader (array processor and scanner) and chip loader, and is based on electronic dot-blot assays which are performed on silicon-based chips manufactured on oxidized silicon wafers using standard microfabrication technology. Each chip holds 100 microelectrodes arranged in a 10 × 10 array, and a thin hydrogel permeation layer containing streptavidin allows binding of the biotinylated PCR product strand[6]. Stabilizers and either Cy5- or Cy3-tagged oligonucleotide reporters for each locus are hybridized and the chip is then washed and imaged. As shown in figure 1, fluorescence signal ratios of reporter A to B clearly allows the discrimination between homozygote and heterozygote samples for the three SNPs in the promoter region of the protein C gene.

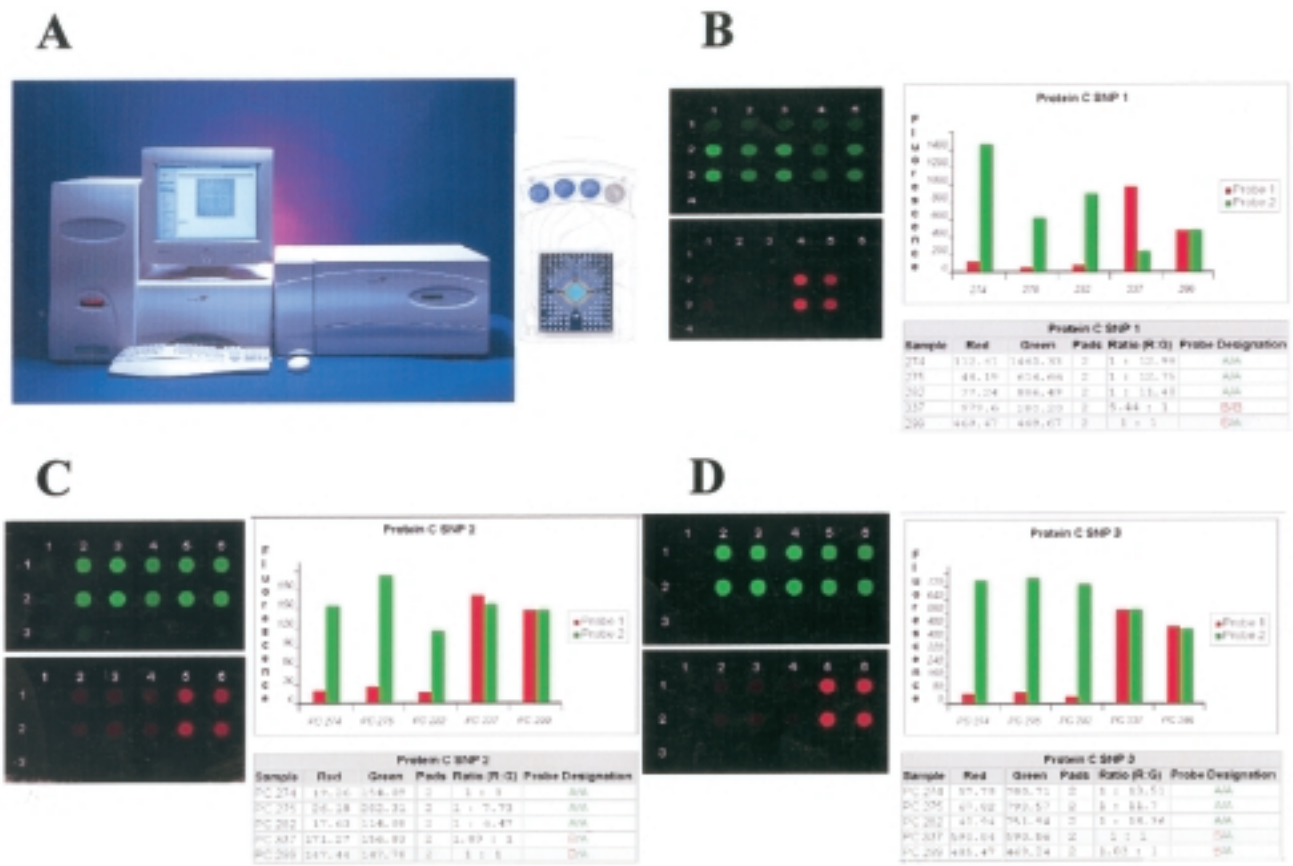

Figure 1. *Detection of three single nucleotide polymorphisms (SNPs) in the promoter region of the protein C gene.* Panel A *shows the Nanogen Research System and a NanoChip™ (Nanogen Inc., San Diego, CA, USA).* Panels B, C *and* D *show results of 5 samples (labeled 264, 275, 282, 337, and 299) run in duplicate (2 pads) using the NanoChip™ platform. A and B alleles are shown in* green *and* red *channels, respectively.* On the left *is shown an electronic image of the array. Results are represented either by bar graphs or numerically in tabular formats. The table includes fluorescent intensity of red and green channels, the ratio of intensities, and the allele assignment (as indicated by the column entitled Probe Designation).*

Sequence variation in genes involved in thrombotic disorders

Arterial and venous thrombotic diseases have long been recognized as the major cause of morbidity and mortality in the western world. However, only recently has vascular disease been characterized at a genetic level. The genetic basis of vascular disease exhibits complexity with variable penetrance and strong environmental influence. Vascular disease is clearly not simply the result of a single Mendelian-inherited gene. Until recently, the capability of routinely performing numerous genetic analyses was too cumbersome for general screening purposes. However, with the capabilities of microarrays, a parallel high-throughput approach is now feasible for combining information currently acquired from serially performed genetic tests. Such an approach could be used for evaluating the risk of thromboembolic disease in high-risk families, in patients who have already experienced a thromboembolic complication, in women being placed on medication containing estrogens, and ultimately for general population screening similar to what is currently done for monitoring cholesterol levels.

Although severe inherited hypercoagulability is most commonly due to a deficiency of one of the anticoagulant proteins such as protein C, protein S or antithrombin III, it is becoming increasingly clear that additional genetic mutations and polymorphisms lead to an increased risk of arterial and/or venous thrombotic phenomena. Many hypercoagulability-related genetic variations have been described in the past 6 years including polymorphisms/mutations in the genes encoding prothrombin, factor V, factor VII, platelet glycoprotein IIIa and methylenetetrahydrofolate reductase[7-11].

A clear example of a genetic risk factor is a mutation at nucleotide 1691 encoding amino acid 506 in the heavy chain of factor V (factor V Leiden)[7]. This mutation occurs in up to 15% of individuals in selected northern European populations but in less than 1% of individuals in most African and Asian populations. It would be particularly attractive to screen for multiple polymorphisms/mutations, which, in combination, enhance the pathogenicity of common mutations such as factor V Leiden. Additionally, as we become more aware of the interplay between genes and environment, screening may help direct specific advice and therapy to high-risk individuals whose risk of thrombotic disease may be selectively enhanced by stimulating factors such as surgery, the *peri-partum* period, smoking, oral contraceptive pills and diet. As the population ages, thrombosis due to numerous polymorphisms and environmental stimuli becomes a rapidly increasing problem as advanced age results in increased disruption of the smooth antithrombotic vascular endothelial surface.

It is crucial in analyzing polymorphisms specific to a given ethnicity to have an appropriate population control. A genetic screen could include genetic tests for genes that are definitively associated with hypercoagulability and atherosclerotic disease, those which have been proposed to be associated, and those which are highly controversial. In addition, two main strategies could be planned: 1) a case-finding strategy in which persons are screened who are at increased risk of premature arterial or venous thrombotic disease on the basis of personal or family history, and 2) a population-based strategy where people are screened irrespective of symptomatology, similar to the current screening for cholesterol levels. Because the presence of high-risk DNA polymorphisms is common, the benefit of screening using a population-based strategy will be great, especially once appropriate therapeutic guidelines have been established for the various risks. Below we include a list of potential genetic targets for evaluating cardiovascular risk in the arterial and venous circulation.

Relevant genes within the arterial system

Fibrinogen. Fibrinogen, the molecule whose hydrolysis generates a fibrin clot, is a dimeric protein composed of three polypeptide chains (α, β and γ) which are encoded in tandem for three genes on the long arm of chromosome 4. Increased levels of fibrinogen have been associated with coronary artery disease and venous disease. A common variation of a single-base pair polymorphism in the promoter region of the β fibrinogen gene at position -455 has been shown to influence fibrinogen levels[12,13].

Factor VII. Increased levels of factor VII, which combines with the membrane protein, tissue factor, to initiate clotting, have been associated with an increased incidence of coronary artery disease. Two common variations, a single-base pair polymorphism in the coding region of factor VII at position 10976 in amino acid 353 (Arg to Glu), and the presence of a decanucleotide insert CCTATACTTC at -323 in the promoter have been associated with decreased levels of factor VII. Three other polymorphisms at -402, -401 and -122 have also been described in addition to a 37 nucleotide variable number of tandem repeats within intron 7[9,14].

PlA1 and PlA2. PlA is an allele of the platelet glycoprotein IIIa, a platelet integral membrane protein of the integrin family that binds to fibrinogen. Although highly controversial, some studies have shown that the presence of a single-base change correlates with the presence of coronary artery disease. Persons positive for PlA1 have a leucine at position 33 of the mature glycoprotein IIIa. Persons positive for PlA2 have a proline at this position, which is the result of a substitution of cytosine for thymidine at position 1565 in exon 2 of the glycoprotein IIIa gene. A test for this polymorphism

could potentially be useful in directing therapy as numerous gpIIb/IIIa inhibitor pharmaceuticals are currently being evaluated. Also, the test could be useful in blood banking where it is sometimes necessary to test for the PlA1/PlA2 phenotypes in mothers with thrombocytopenic newborns (for neonatal alloimmune thrombocytopenia), and in patients refractory to platelet transfusions (anti-PlA1 antibodies due to previous exposure)[10].

5,10-methylenetetrahydrofolate reductase. 5,10-methylenetetrahydrofolate reductase is an enzyme leading to the degradation of homocysteine, a molecule shown to greatly influence the presence of coronary artery disease. A common single-base change corresponding to position 677 (Ala to Val) in the coding region causes heat lability of 5,10-methylenetetrahydrofolate reductase and may be associated with increased incidence of venous thromboembolic disease[11,15].

Apolipoprotein E. The apoE4 allele of apoE is associated with an increased risk of cardiovascular disease. Molecular biological tests to identify the presence of two polymorphisms in codons 112 and 158 aid in establishing the apoE genotype[13,16,17].

Relevant genes within the venous system

Factor V Leiden. Factor V is a procoagulant protein essential for blood clot formation. A common mutation at bp 1691 in exon 10 leading to a mutation in amino acid 506 in factor V, one of the sights at which factor V is normally degraded by activated protein C and protein S, causes resistance to activated anticoagulant protein C, and is highly correlated with an increased risk of thrombotic disease[7].

Prothrombin 3' untranslated region. Thrombin, the activated form of the precursor prothrombin, is a vitamin K-dependent procoagulant enzyme responsible for catalyzing the formation of a fibrin clot. A single-base change (G to A) at position 20210 present in the 3'-untranslated region of prothrombin is associated with prothrombin plasma levels and increased incidence of both venous and arterial disease. Its presence is highly associated with the presence of factor V Leiden in some reports of patients with venous thrombotic disease[8].

Protein C promoter polymorphisms. Protein C is a vitamin K-dependent anticoagulant protein that catalyzes breakdown of the procoagulant factors Va and VIIIa. Three single-base pair polymorphisms in the promoter region at -654 (C to T), -641 (A to G) and -476 (A to T) of the protein C gene are associated with plasma protein C levels which have been shown to influence incidence of venous thrombotic disease. Although 27 genotypes of these 3 SNPs are theoretically possible, only 9 have been biologically observed, 5 of which occurred in > 95% of individuals[18].

Plasminogen activator inhibitor-1 promoter polymorphism. The plasminogen activator inhibitor-1 (PAI-1) gene influences the rate of clot lysis and thus plays a role in the development of thrombotic disease, particularly related to triglyceride levels and diabetes. A common single-base insertion in the promoter region of the PAI-1 gene is associated with a decrease in plasma PAI-1 levels[19].

Lipoprotein(a). Lipoprotein(a) acts as an anti-fibrinolytic factor by inhibiting clot lysis. Three single-base pair polymorphisms in the 5'-flanking region of the apo(a) gene at -771, +93, and +121 are associated with lipoprotein(a) plasma levels, high levels of which are a risk factor for thromboembolic disease[20].

Thrombomodulin. Thrombomodulin, an integral membrane protein normally expressed on the lumenal surface of the endothelium, transforms thrombin from a procoagulant to an anticoagulant protein. Two polymorphisms in the promoter region may constitute a risk factor for coronary artery disease[21].

In conclusion, the microarray age has brought with it a tremendous capacity for high-throughput, parallel genetic analysis of numerous genomic sequence that will enable us to not only diagnose current disease but to predict risk of future disease. However, in so doing we must develop intelligent schemes for applying the new technology in an appropriate medical scenario.

References

1. Nickerson DA, Taylor SL, Weiss KM, et al. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nature 1998; 19: 233-40.
2. Roses AD. Pharmacogenetics and the practice of medicine. Nature 2000; 405: 857-65.
3. The Chipping Forecast. Nat Genet 1999; 21 (Suppl 1): 1-60.
4. Hacia JG. Resequencing and mutational analysis using oligo-nucleotide microarrays. Nat Genet 1999; 21 (Suppl 1): 42-7.
5. Gerry NP, Witowski NE, Day J, Hammer RP, Barany G, Barany F. Universal DNA microarray method for multiplex detection of low abundance point mutations. J Mol Biol 1999; 292: 251-62.
6. Gilles PN, Wu DJ, Foster CB, Dillon PJ, Chanock SJ. Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips. Nat Biotechnol 1999; 17: 365-70.
7. Bertina RM, Koeleman BP, Koster T, et al. Mutation in blood coagulation factor V associated with resistance to activated protein C. Nature 1994; 369: 64-7.
8. Poort SR, Rosendaal FR, Reitsma PH, Bertina RM. A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. Blood 1996; 88: 3698-703.

9. Girelli D, Russo C, Ferraresi P, et al. Polymorphisms in the factor VII gene and the risk of myocardial infarction in patients with coronary artery disease. N Engl J Med 2000; 343: 774-80.
10. Weiss EJ, Bray PF, Tayback M, et al. A polymorphism of a platelet glycoprotein receptor as an inherited risk factor for coronary thrombosis. N Engl J Med 1996; 334: 1090-4.
11. Frost P, Blom HJ, Milos R, et al. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate. Nat Genet 1995; 10: 111-3.
12. Thomas AE, Green FR, Humphries SE. Association of genetic variation at the beta-fibrinogen gene locus and plasma fibrinogen levels; interaction between allele frequency of the G/A-455 polymorphism, age and smoking. Clin Genet 1996; 50: 184-90.
13. Kessler C, Spitzer C, Stauske D, et al. The apolipoprotein E and beta-fibrinogen G/A-455 gene polymorphisms are associated with ischemic stroke involving large-vessel disease. Arterioscler Thromb Vasc Biol 1997; 17: 2880-4.
14. Bernardi F, Marchetti G, Pinotti M, et al. Factor VII gene polymorphisms contribute about one third of the factor VII level variation in plasma. Arterioscler Thromb Vasc Biol 1996; 16: 72-6.
15. Clarke R, Daly L, Robinson K, et al. Hyperhomocysteinemia: an independent risk factor for vascular disease. N Engl J Med 1991; 324: 1149-55.
16. Lucotte G, Loirat F, Hazout S. Pattern of gradient of apolipoprotein E allele *4 frequencies in Western Europe. Hum Biol 1997; 69: 253-62.
17. Srinivasan SR, Ehnholm C, Wattigney WA, et al. The relation of apolipoprotein E polymorphism to multiple cardiovascular risk in children: the Bogalusa Heart Study. Atherosclerosis 1996; 123: 33-42.
18. Spek CA, Koster T, Rosendaal FR, et al. Genotypic variation in the promoter region of the protein C gene is associated with plasma protein C levels and thrombotic risk. Arterioscler Thromb Vasc Biol 1995; 15: 214-8.
19. Dawson S, Hamsten A, Wiman B, et al. Genetic variation at the plasminogen activator inhibitor-1 locus is associated with altered levels of plasma plasminogen activator inhibitor-1 activity. Arterioscler Thromb 1991; 11: 183-90.
20. Suzuki K, Kuriyama M, Saito T, et al. Plasma lipoprotein(a) levels and expression of the apolipoprotein(a) gene are dependent on the nucleotide polymorphisms in its 5'-flanking region. J Clin Invest 1997; 99: 1361-6.
21. Ireland H, Kunz G, Kyriakoulis K, et al. Thrombomodulin gene mutations associated with myocardial infarction. Circulation 1997; 96: 15-8.