Evaluation of a microarray for genotyping polymorphisms related to xenobiotic metabolism and DNA repair

Stefano Landi[1,2], Federica Gemignani[1,2], Lydie Gioia-Patricola[1], Amélie Chabrier[1], and Federico Canzian[1]

BioTechniques 35:816-827 (October 2003)

We present an oligonucleotide microarray ("MetaboChip") based on the arrayed primer extension (APEX) technique, allowing genotyping of single nucleotide polymorphisms (SNPs) in genes of interest for cancer susceptibility and pharmacogenetics. APEX consists of a sequencing reaction primed by an oligonucleotide anchored with its 5′ end to a glass slide and terminating one nucleotide before the polymorphic site. The extension with one fluorescently labeled dideoxynucleotide complementary to the template reveals the polymorphism. Ninety-three SNPs in 42 genes were selected among those resequenced in the context of the SNP500 project, using a set of 102 reference DNA samples from the Coriell Biorepository. Selected SNPs belong to the following genes: ADH1B, ALDH2, APEX, CDKN2A, COMT, CYP1A1, CYP1A2, CYP1B1, CYP2A6, CYP2C19, CYP2C9, CYP2E1, CYP3A4, DRD2, DRD4, EPHX1, ERCC1, ERCC2, ERCC4, ERCC5, GRPR, GSTA4, GSTM3, GSTP1, GSTT2, LIG3, MDM2, MGMT, MPO, NAT1, NAT2, NQO1, OGG1, PCNA, POLB, SLC6A3, SOD2, TP53, XRCC1, XRCC2, XRCC3, *and* XRCC9. *We assessed the performance of APEX by comparing the results obtained with MetaboChip against those reported by the SNP500. Among 88 SNPs that yielded signals, 6 showed less than 99% of concordance, whereas 82 performed accurately, showing that APEX is a reliable and sensitive genotyping method.*

INTRODUCTION

Exposure to xenobiotics is thought to be one of the most important causes of cancer in humans. However, human genetic variability may affect the individual risk of contracting the disease. For example, among environmental carcinogens, exposure to cigarette smoke is considered one of the main causes of cancer (1,2), but only 15% of lifelong smokers develop a lung cancer by the age of 75 years (3). Genetic variability governing metabolism of xenobiotics as well as DNA repair, cell cycle control, apoptosis, inflammation, and other processes are probably responsible for these inter-individual differences (4). It is thought that the genetic factors underlying multifactorial diseases could be dissected by the use of suitable markers such as single nucleotide polymorphisms (SNPs), and by association studies comparing the differences of SNP frequencies between cases and controls. For this purpose, new methods of rapid and high-throughput genotyping analysis have been developed in the last decade (5).

We present here an oligonucleotide microarray that allows parallel genotyping of SNPs in genes involved in xenobiotic metabolism, cell cycle control, DNA repair, and nicotine addiction (known candidates for conferring susceptibility to a broad range of diseases such as cancer) and of potential interest in pharmacogenetics. We call this microarray "MetaboChip." The MetaboChip is based on the arrayed primer extension (APEX) technology (6,7). APEX consists of a sequencing reaction primed by an oligonucleotide anchored with its 5′ end to a glass slide and terminating just one nucleotide before the polymorphic site. A DNA polymerase extends the oligonucleotide by adding one fluorescently labeled dideoxy nucleoside (5′) triphosphate (ddNTP) complementary to the variant base. Reading the incorporated fluorescence identifies the base in the target sequence. This method is suitable not only for SNPs but also for small insertion/deletion polymorphisms (8).

Previously, APEX produced valuable results when 17 mutations of the β-globin and *G6PD* genes (7,9), 40 among the major known mutations underlying the diseases belonging to the "Finnish disease heritage" (10), and 25 Y-chromosomal SNPs in a unique collection of samples representing five Finno-Ugric populations (11) were genotyped. It was also extensively used for resequencing 1.2 kb of the *TP53* gene (12). All these studies showed that the method is reliable, but also that each position performs differently depending on the flanking sequence, in a way that is not easily predicted in silico.

To better assess the reliability of

[1] International Agency for Research on Cancer, Lyon, France and [2] University of Pisa, Pisa, Italy

Table 1. Genotype Frequencies Obtained with MetaboChip and Compared with SNP500

Gene Name	Trivial SNP Name	SNP ID[a]	Homozygote Allele 1	Heterozygote	Homozygote Allele 2	Concordant Genotypes vs. SNP500 (%)
ADH1B	R48H	rs1229984	0.80	0.13	0.07	97
	R369C	rs2066702	0.91	0.07	0.02	100
ALDH2	348T>C	rs440	0.69	0.30	0.01	100
	483T>C	rs 441	0.69	0.28	0.03	100
	E504K	rs671	0.95	0.04	0.01	100
APEX/APE1	Q51H	rs1048945	0.96	0.04	0	100
CDKN2A	A148T	E0284_302	0.97	0.03	0	100
COMT	V158M	rs4680	0.49	0.35	0.16	100
	186C>G	rs4818	0.59	0.29	0.12	100
	A72S	rs6267	0.98	0.02	0	100
	rs4633	rs4633	0.40	0.44	0.16	100
CYP1A1	I462V	rs1048943	0.83	0.16	0.01	100
	T461N	rs1799814	0.99	0.01	0	100
	rs2606345	rs2606345	0.45	0.14	0.41	100
	rs4134577	rs4134577	0.53	0.32	0.16	100
CYP1A2	-3858G>A	rs2069514	0.66	0.30	0.04	100
	-164C>A	rs762551	0.58	0.36	0.08	100
	1545T>C	rs2470890	0.56	0.36	0.28	100
CYP1B1	-13C>T	rs2617266	0.47	0.47	0.06	100
	V432L	rs1056836	0.35	0.35	0.30	100
	rs1056837	rs1056837	0.41	0.32	0.27	100
	4390A>G	rs1800440	0.79	0.16	0.05	100
CYP2A6	L160H	rs1801272	0.92	0.08	0	100
CYP2C19	-681G>A	rs4134593	0.78	0.16	0.06	99
CYP2C9	R430C	rs1799853	0.87	0.12	0.01	99
	I359L	rs1057910	0.94	0.06	0	100
CYP2E1	9893C>G	rs2070676	0.48	0.37	0.15	100
	-1293G>C	rs3813867	0.49	0.36	0.15	98
	-1053C>T	rs2031920	0.94	0.06	0	100
	-333T>A	rs2070673	0.33	0.40	0.27	99
	-71G>T	rs4134606	0.88	0.10	0.02	100
CYP3A4	20230G>A	rs2242480	0.53	0.23	0.24	100
DRD2	32806C>T	rs1800497	0.52	0.44	0.04	100
	1412A>G	rs6276	0.33	0.38	0.29	100
	960C>G	rs1801028	0.98	0.02	0	100
	rs6277	rs6277	0.57	0.26	0.17	100
	7423A>G	rs1079597	0.72	0.21	0.07	99
	3208G>T	rs1076560	0.76	0.20	0.04	100
DRD4	-809G>A	rs936461	0.33	0.48	0.19	100
EPHX1	14622C>T	rs2234697	0.99	0.01	0	100
	17540T>C	rs2234698	0.96	0.04	0	100
	Y113H	rs1051740	0.42	0.44	0.14	100

Table 1. Continued.

Gene Name	Trivial SNP Name	SNP ID[a]	Homozygote Allele 1	Heterozygote	Homozygote Allele 2	Concordant Genotypes vs. SNP500 (%)
	H139R	rs2234922	0.60	0.38	0.02	100
ERCC1	19716G>C	rs3212948	0.41	0.44	0.15	100
	17677A>C	rs3212961	0.65	0.31	0.04	100
	8092C>A	rs3212986	0.42	0.44	0.14	100
ERCC2	L751Q	rs1052559	0.63	0.33	0.04	100
ERCC4	P379S	rs1799802	0.99	0.01	0	100
	R415Q	rs1800067	0.94	0.06	0	100
ERCC5	335T>C	rs1047768	0.40	0.44	0.16	99
GRPR	450C>T	rs4134652	0.48	0.18	0.34	100
	661C>T	rs4134653	0.42	0.23	0.35	100
GSTA4	Q117Q	rs1802061	0.93	0.07	0	100
GSTM3	*B	rs1799735	0.73	0.20	0.07	100
GSTP1	313A>G	rs947894	0.42	0.39	0.19	100
	341C>T	rs1799811	0.90	0.06	0.04	100
GSTT2	M139I	rs1622002	0.88	0.08	0.04	96
LIG3	R780H	rs3136025	0.95	0.05	0	100
MDM2	E345E	rs769412	0.81	0.18	0.01	100
MGMT	171C>T	rs1803965	0.75	0.21	0.04	100
	262C>T	rs12917	0.77	0.19	0.04	100
	427A>G	rs2308321	0.87	0.10	0.03	99
MPO	-463G>A	rs2333227	-	-	-	Low intensity
NAT1	-344C>T	rs4134716	0.98	0.02	0	100
	-40A>T	rs4134717	0.99	0.01	0	100
	445G>A	rs4151100	0.99	0.01	0	100
	560G>A	rs4986782	0.99	0.01	0	100
	1088T>A	rs1057126	0.70	0.19	0.11	92
	1095A>C	rs15561	-	-	-	Low intensity
NAT2	191G>A	rs1801279	0.92	0.08	0	100
	282C>T	rs1041983	0.43	0.32	0.25	96
	341T>C	rs1801280	0.58	0.30	0.12	100
	590G>A	rs1799930	0.55	0.34	0.11	100
	803A>G	rs1208	0.43	0.32	0.25	100
	857G>A	rs1799931	0.93	0.07	0	100
	481C>T	rs1799929	0.48	0.42	0.10	97
NQO1	R139W	rs4134728	0.95	0.05	0	100
	P187S	rs1800566	-	-	-	Low intensity
OGG1	S326C	rs1052133	0.67	0.31	0.02	100
PCNA	2232T>C	rs25406	0.41	0.36	0.23	100
	3890A>C	rs17352	-	-	-	Low intensity
POLB	P242R	rs3136797	0.97	0.03	0	100
SLC6A3	-1476T>G	rs4134767	0.95	0.05	0	100
SOD2	V16A	rs1799725	0.36	0.43	0.21	100
TP53	R72P	rs1042522	0.32	0.39	0.29	100
XRCC1	R194W	rs1799782	0.80	0.19	0.01	100

Table 1. Continued.

Gene Name	Trivial SNP Name	SNP ID[a]	Homozygote Allele 1	Heterozygote	Homozygote Allele 2	Concordant Genotypes vs. SNP500 (%)
	R280H	rs25489	0.88	0.11	0.01	100
	R399Q	rs25487	0.50	0.37	0.13	100
XRCC2	R188H	rs3218536	0.92	0.07	0.01	100
XRCC3	T241M	rs861539	0.60	0.30	0.10	100
XRCC9	P330S	rs4986940	0.99	0.01	0	100
	V464F	E0343_302	-	-	-	Low intensity
	T297I	rs2237857	0.96	0.04	0	100

[a]Reference SNP (rs) numbers from dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) or E0_numbers from SNP500 project (http://snp500cancer.nci.nih.gov/home.cfm) for single nucleotide polymorphisms (SNPs) not yet posted in dbSNP.
-, not applicable.

the method, we prepared the MetaboChip with 93 SNPs (belonging to 42 genes) characterized in the context of the SNP500 cancer project (http://snp500cancer.nci.nih.gov/home.cfm). Then we used the microarray to genotype the 102 reference DNA samples of the Coriell Biorepository (Camden, NJ, USA), resequenced with direct sequencing, in the context of the SNP500 project. In this way, we could assess the performance of each SNP genotyped with APEX by comparing the results obtained with MetaboChip against those obtained by the SNP500 project by direct sequencing, considered the "gold standard." Table 1 contains detailed information about every SNP.

MATERIALS AND METHODS

DNA Samples

DNAs from the SNP500 project were kindly donated by Dr. Stephen Chanock (National Cancer Institute, Bethesda, MD, USA). The SNP500 project studies the genomes of 102 individuals of self-described heritage, culled from different geographic and ethnic populations: African, Caucasian, Hispanic, and Pacific Rim heritage. The purpose of the SNP500 project is to resequence (by direct sequencing) reference DNAs using anonymous samples from the Coriell Biorepository. The project seeks to validate known or newly discovered SNPs and other important classes of genetic variants of potential importance to molecular epidemiology studies of cancer and other diseases.

SNP Selection

We selected 93 SNPs in genes related to xenobiotic metabolism, DNA repair, cell cycle control, and nicotine addiction for which the genotypes of the reference DNAs were known by direct sequencing from the SNP500 cancer project. Because there was no special priority in doing the selection, the selected SNPs have to be considered a random sample of SNPs taken for validation purposes.

PCR Protocol

PCR amplifications of all DNA fragments containing the 93 selected polymorphisms were designed as multiplex to minimize the total number of reactions and to work under the same conditions. Briefly, multiplex PCRs were performed in a final volume of 20 µL containing 1× Platinum® *Taq* buffer (Invitrogen, Carlsbad, CA, USA), 1.5 mM $MgCl_2$, 10 pmol of each forward and reverse primer, 1.25 U Platinum *Taq* polymerase, and 20 ng of template DNA. PCRs were performed with 50 µM dUTP, 150 µM dTTP, and 200 µM of dATP, dCTP, and dGTP, to allow PCR product fragmentation. Cycling parameters were as follows: initial denaturation at 94°C for 10 min; 20 touchdown cycles of 94°C for 30 s, 68°C for 30 s -1°C/cycle, 72°C 30 s; 20 cycles of 94°C for 30 s, 51°C for 30 s, 72°C for 30 s; and final elongation at 72°C for 10 min. The preparation of PCRs was performed by robotically plating primer mixes in a well-specific fashion and dispensing DNA and the reagent mixture on top of the primers. More details on primer sequences and grouping of multiplex PCRs are reported in Table 2 and elsewhere (http://www-gan.iarc.fr/MetaboChip.html and Supplementary Table 1, List of Genes and SNPs with Their Relative APEX Oligonucleotides, at http://www.BioTechniques.com/Oct03/Canziansupplementary.html) (13).

APEX Protocol

Since both sense and antisense strands are sequenced, two oligonucleotides were designed for each polymorphism. The complete sequence of APEX oligonucleotides is available on the Web (http://www-gan.iarc.fr/MetaboChip.html). 5′ (C-12) aminolinker oligonucleotides were synthesized by Sigma Genosys (Cambridge, UK) and spotted onto silanized slides as reported elsewhere (8,14). PCR products were pooled, purified, and concentrated using Microcon YM-30 columns (Millipore, Billerica, MA, USA). To allow better hybridization with the arrayed oligonucleotides, the PCR products were reduced in size by fragmentation. To achieve this,

Table 2. Full Gene Name, Trivial SNP Name, Sense and Antisense Primers, PCR Size, and Group of Multiplex to Amplify Selected Regions of SNPs

Gene Name	Amino Acidic Change/ Trivial Name	Forward PCR Primer	Reverse PCR Primer	PCR Size (bp)	Multiplex PCR Group
ADH1B	R48H	5′-AATCTTTTCTGAATCTGAACAG-3′	5′-GAAGGGGGGTCACCAGGTTG -3′	107	1
	R369C	5′-TGCTCCTTGGACTCTCACAACA -3′	5′-TCACTTGAATTTTAAATTTTCCTGAA -3′	288	2
ALDH2	348T>C ; 483T> C	5′-AAATATTGCTCTAGGCCAGGC -3′	5′-TGGGAATTCTAAATGGGACGG -3′	401	3
	E504K	5′-CGTTTCAAATTACAGGGTCAACTGCT-3′	5′-AGCCCCCAACAGACCCCAATC-3′	227	24
APEX/APE1	Q51H	5′-CTCATTTTATAGAGCCAGAGGCCAAGAAGA-3′	5′-CACTCACATCTAATCCTTTCTTCTTA -3′	207	25
CDKN2A	A148T	5′-ATGCCTGGGGCCGTCTGCCCGTGGACC-3′	5′-TTTCTGTGCTGGAAAATGAATGCTCTGAG-3′	210	4
COMT	V158M; 186C>G	5′-TACTGTGGCTACTCAGCTGT -3′	5′-TGAACGTGGTGTGAACACCT -3′	236	5
	A72S; rs4633	5′-GCCCATCCACAACCTGCTC -3′	5′-CAACCTTTCTTGTCGCCCA -3′	162	6
CYP1A1	I462V; T461N	5′-CTGTCTCCCTCTGGTTACAGGAAGC -3′	5′-TTCCACCCGTTGCAGCAGGATAGCC -3′	204	7
	rs2606345	5′-GCTTTCTGATGACAGGGGCT-3′	5′-CAAGGGAGATGTGACTGGTGAG-3′	232	18
	rs4134577	5′-CTGGGATTGAGCAGCAGAGAACT-3′	5′-CCAAGGTCACACAGCCAGAAAGT-3′	546	31
CYP1A2	-3858G>A	5′-AACACATGATCGAGCTATAC -3′	5′-GTGGTCTCTTCACTGTAAAGTTA -3′	596	25
	-164C>A	5′-CCCAGAAGTGGAAACTGAGA -3′	5′-GGGTTGAGATGGAGACATTC -3′	243	8
	1545T>C	5′-CGGTGTATCGGGGAAGTCCTG -3′	5′-GGAAGAGAAACAAGGGCTGAGTCC -3′	256	26
CYP1B1	-13C>T	5′-TCTCCAGAGAGTCAGCTCCG -3′	5′-GGGTCGTCGTGGCTGTAG-3′	786	33
	V432L; rs1056837; 4390A>G	5′-TCCCAGAAATATTAATTTAGTCACTG -3′	5′-TATGGAGCACACCTCACCTG -3′	885	21
CYP2A6	L160H	5′-TTCACCTCCCCAGGCGTGGTA-3′	5′-TGGCGCCTGCGGGTATGGC-3′	403	9
CYP2C9	R430C	5′-CAATGGAAAGAAATGGAAGGAGAT -3′	5′-GCTAACAACCAGGACTCATAATGAAAGA -3′	252	10
	I359L	5′-GTCCAGGAAGAGATTGAACGTGTGA -3′	5′-AAATGATACTATGAATTTGGGGACTTCG -3′	238	6
CYP2C19	-681G>A	5′-TTATAATTACAACCAGAGCTTGGC-3′	5′-TATCACTTTCCATAAAAGCAAG-3′	168	20
CYP2E1	9893C>G	5′-GTGGGATACTGCATCTCCAG -3′	5′-GAAGGTACTGCCTCTGATCT -3′	526	11
	-1293G>C; -1053C>T	5′-CCAGTCGAGTCTACATTGTCA-3′	5′-TTCATTCTGTCTTCTAACTGG -3′	413	12
	-333T>A; -71G>T	5′-GTGGCTGGAGTTCCCCGTTG -3′	5′-TGCTGCCAGCCCGGGAGACA -3′	361	26
CYP3A4	20230G>A	5′-CCCAGTGTACCTCTGAATTGC -3′	5′-TTTCAGAGCCTTCCTACATAGAG -3′	430	27
DRD2	32806C>T	5′-ACCACGGCTGGCCAAGTTGTCTAA -3′	5′-CACCTTCCTGAGTGTCATCAACCTC -3′	309	3
	1412A>G	5′-GCCGTGCCTCCCCGGCTCTG-3′	5′-GGCAGTGAGGAGCATGGAGCCAAG -3′	405	17
	960 C>G; rs6277	5′-AGGACAGGGGCAATCCTGCAGGGGC -3′	5′-GGAAGGACATGGCAGGGAATGGGAC -3′	592	7
	7423A>G	5′-GATACCCACTTCAGGAAGTC -3′	5′-GATGTGTAGGAATTAGCCAGG -3′	459	13
	3208G>T	5′-AGGACAGGGGCAATCCTGCAGGGGC -3′	5′-GGAAGGACATGGCAGGGAATGGGAC -3′	592	27
DRD4	-809G>A	5′-CAGGTCACAGGTCACCCCTCTT -3′	5′-TTGCTCATCTTGGAATTTTGCG -3′	792	34
EPHX1	14622C>T	5′-GGTCTTCCCCTCATCTTGC -3′	5′-CCCGGCCCAAGGTGCCTT -3′	229	28
	17540T>C; Y113H	5′-GGGGTCCTGAATTTTGCTCC -3′	5′-TGGCTGGCGTTTTGCAAACAT -3′	226	13
	H139R	5′-ACATCCACTTCATCCACGT -3′	5′-ATGCCTCTGAGAAGCCAT -3′	210	14

Table 2. Continued.

Gene Name	Amino Acidic Change/ Trivial Name	Forward PCR Primer	Reverse PCR Primer	PCR Size (bp)	Multiplex PCR Group
ERCC1	19716G>C	5′-TCTCAACCCACACACTGCTG-3′	5′-AGGGAGACGGAGAAGTGAGG-3′	157	28
	17677A>C	5′-GCCCTTAGTATTCCAGTGAG -3′	5′-GGACTAATTGAAGGGGATGT -3′	402	11
	8092C>A	5′-TAGTTCCTCAGTTTCCCG -3′	5′-TGAGCCAATTCAGCCACT -3′	255	15
ERCC2/XPD	L751Q	5′-CAGTGCCCCCCTCTCCCTTTCCTCTGTTC-3′	5′-GGGAACCAGGGCCAGGCAAGACTCAGGAGT-3′	193	16
ERCC4/XPF	P379S	5′-GCCTTTGGAAGACTTTATGGGTAAATAT-3′	5′-CATCTCCTTTTTTCCTACCTGGACCACCAA-3′	193	10
	R415Q	5′-TAAGGGGGCACAGGGAAACTAGGAGGACAA -3′	5′-CAGCTTTGCTATCCTTCTCAAAGGTTTTCC-3′	195	29
ERCC5/XPG	335T>C	5′-GGGTAACAAGAGTTCAACTAAAAG -3′	5′-ACTAAAAACTATACGACACTTACCA -3′	382	2
GRPR	450C>T; 661C>T	5′-TTCTTTCCTGCCCTAGATACA -3′	5′-CACCTACCTGCTTCTTGACA -3′	375	4
GSTA4	Q117Q	5′-TGGAAACTGATGGCTTCAAAGA -3′	5′-TGTGGTTTGGTTGCTCTGTGTA -3′	244	29
GSTM3	*B	5′-AAGCCTCAGTACTTGGAAGAGCT-3′	5′-CACATGAAAGCCTTCAGGTT-3′	273	15
GSTP1	313A>G	5′-TCCTTCCACGCACATCCTCT-3′	5′-AGCCCCTTTCTTTGTTCAGC-3′	290	23
	341C>T	5′-TGGACAGGATTTGGTACTAGCCT -3′	5′-CACCTGGTCTCCCACAATGAAGG-3′	249	1
GSTT2	M139I	5′-CCCCATCAGGTGTTGGGGCCA-3′	5′-CTGGGCAAAGTTCTGGCTTTCTCCAGT-3′	279	16
LIG3	R780H	5′-GGAGAAGGCAGACTTCACTGTAGTGGCTGG-3′	5′-TCACCATACCAGCCCCCTCCCTATCCTT -3′	205	17
MDM2	E354E	5′-TGAAGGACTATTGGAAATGCAC-3′	5′-GAGAATAGTCTTCACTTTCTTGTGC-3′	290	35
MGMT	171C>T; 262C>T	5′-CTAAGCCCCTGTTCTCACTTTT -3′	5′-ACACCGCAGATGGCTTAGTTAC -3′	200	5
	427A>G	5′-AAGACCTCGTTGTCCAGATCCCT -3′	5′-CGCTCAAACATCCATCCTACTGC -3′	293	17
MPO	-463G>A	5′-TGGTAGTGCTAAATTCAAAGGC-3′	5′-AGGCAAGAAGCTAATTTTTGTAT-3′	172	18
NAT1	-344C>T; -40A>T	5′-CTTTGTATAAGGCTCAGCTAAAAGG-3′	5′-TTCAATGTCCATGATCCCC -3′	452	30
	445G>A; 560G>A	5′-CCTACTCAAATCCAAGT -3′	5′-AAGGAACAAAATGATTTAC-3′	741	19
	1088T>A; 1095A>C	5′-TCTAGACCAAATCAGAA-3′	5′-CTAGCATAAATCACCAA -3′	663	23
NAT2	191G>A; 282C>T; 341T>C	5′-TATAACCATTGTGTTTTTACGTATT -3′	5′-TCTCTCCTGATTTGGTCCAG -3′	610	19
	590G>A; 803A>G; 857G>A; 481C>T	5′-GGAACAAATTGGACTTGGAAAC -3′	5′-GAGAGGATATCTGATAGCACATA-3′	907	30
NQO1	R139W	5′-TTGGAGTCCCTGCCATTCTG-3′	5′-CCTGCATCAGGACAGACCACC-3′	159	5
	P187S	5′-GTTGACTTACCTCTCTGTGCTTT-3′	5′-GCGTTTCTTCCATCCTTCCA-3′	149	12
OGG1	S326C	5′-ACTGTCACTAGTCTCACCAG -3′	5′-GGAAGGTGCTTGGGGAAT -3′	207	20
PCNA	2232T>C	5′-TCATTGGCTGGCGTGGGCATC -3′	5′-TTAGAAGGGGTTACCACTCTC -3′	566	30
	3890A>C	5′-AGTGAGGGTGCCAAATCATT -3′	5′-GCCTAACTTCCCAATCCTACC -3′	536	31
POLB	P242R	5′-AATGGCCTTGTGTTTTACTTGATTAAAATT-3′	5′-GGAGAAAACGAGACAAGTTTGGAAAAGTTA-3′	205	31
SLC6A3	-1476T>G	5′-ACAGCTTCGAGGTGGCAC -3′	5′-CTGTGTCTGGTGAGGGCC -3′	245	21
SOD2	V16A	5′-CAGCCTGCGTAGACGGTCC -3′	5′-TTGATGTGAGGTTCCAGGGC -3′	164	8

Table 2. Continued.

Gene Name	Amino Acidic Change/ Trivial Name	Forward PCR Primer	Reverse PCR Primer	PCR Size (bp)	Multiplex PCR Group
TP53	R72P	5′-ATCTACAGTCCCCCTTGCCG-3′	5′-GCAACTGACCGTGCAAGTCA-3′	296	22
XRCC1	R194W	5′-GACCTTAGAAGGTGACAGTGACCAA-3′	5′-CCTCAGACCCACGAGTCTAGGTCTCAAC-3′	183	31
	R280H	5′-TGGTGCTAACCTAATCTACTCTT-3′	5′-GGTGCCTTCTCCTCGGGGTTTGCC-3′	124	16
	R399Q	5′-ACAGCCAGGTCCTAGGCCTGGGA-3′	5′-GGCAGGCCCCAGTCTGACTCCCC-3′	175	24
XRCC2	R188H	5′-TAGACCGCGTCAATGGAGGAGA-3′	5′-CCACATCACACAGTCGTCGAGAGGCATG -3′	189	9
XRCC3	T241M	5′-GGCAGCCCCATTCCGCTG -3′	5′-GGGTGCAACCCTGCCTTGG -3′	183	32
XRCC9	P330S	5′-ATTCATCCCCCAAGTCACAA	5′-AGCAGAGCTGGAGAGTCTGG -3′	423	14
	V464F	5′-AGATGTCCCGGCTGTGGGAAGA	5′-TCATCCCTCCACACCCCCTCTAG -3′	201	32
	T297I	5′-CAGAGAGCACTGTTGTACTTGGTTG	5′-AGTCACCCCATCACAAGCACCTCAGGAA -3′	202	22

15 µL of eluate were collected and treated with 1 U uracil *N*-glycosylase (UNG; EPICENTRE, Madison, WI, USA) and 1 U shrimp alkaline phosphatase (sAP; Amersham Biosciences, Milwaukee, WI, USA). The mixture was incubated at 37°C for 1.5 h and at 95°C for 30 min. DNA with abasic sites is labile and is denatured and fragmented at 95°C, whereas UNG and sAP are inactivated. Then, 9 µL of the mixture were added to a reaction mixture containing fluorescently labeled ddNTPs (4 × 50 pM), 10× reaction buffer (260 mM Tris-HCl, pH 9.5, and 65 mM $MgCl_2$), and 4 U of Thermo Sequenase (Amersham Biosciences), diluted in the provided dilution buffer, to give a final volume of 20 µL. The mixture was quickly placed onto the spotted slide and incubated at 58°C for 25 min. Slides were washed and a droplet of SlowFade® Light Antifade Reagent (Molecular Probes, Eugene, OR, USA) was added to limit the bleaching of fluorescein. Slides were imaged by a Genorama™-003 four-color detector equipped with Genorama image analysis software (Asper Biotech, Tartu, Estonia). Four images were analyzed, each corresponding to a fluorochrome (i.e., a base). Fluorescence intensities at each position were converted automatically into base calls by the software. In the case of more than one signal present on a given position, only the main signal was considered, when the intensity of the weaker signal was lower than 10% of the main signal.

RESULTS AND DISCUSSION

Five out of 93 SNPs did not give signals intense enough to allow a clear discrimination among genotypes and were

discarded. Among the others, we obtained 76 SNPs (86%) with a complete concordance with genotypes reported by the SNP500 database, and 6 with a concordance of 99%. Overall, 82 out of 88 SNPs had a concordance equal to or greater than 99%. The remaining 6 had a concordance between 92% and 98.5%. The results are summarized in Table 1.

The hybridization and extension steps are conducted under the same conditions, which explains why a number of SNPs did not elicit any signal. The fact that all SNPs are processed in the same tube after pooling of PCRs is one of the advantages of the APEX approach, from the point of view of manipulation and cost. However, it also makes tuning of the conditions for individual SNPs impossible. The five failures (5/93) could be the consequence of insufficient hybridization between the oligonucleotides anchored onto the glass slide and the PCR product. Oligonucleotide design was performed by using an algorithm established for oligonucleotide in solution (NetPrimer: http://www.premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html), to allow a theoretical hybridization at 58°C. However, it has been shown that other variables should be included in the calculation when the oligonucleotide is anchored with its 5′ end on the glass, including the electric potential of the dielectric surface and the surface conditions (15). Moreover, secondary structures are sometimes unavoidable, because there is little freedom in choosing the position where to place the APEX oligonucleotides. Thus, we should consider the 5% failure rate as normal for this technique. Among the 88 SNPs that gave signals, the results are promising with an overall discordance rate of 0.26%. Only 6 out of 88 SNPs had a performance lower than 99% and should not be recommended for use with sensitive applications. These discrepancies may be explained by several different aspects of the technique. Often the SNP500 project used PCR primers different from those used in our laboratory. Thus, if there is an unknown polymorphism in linkage disequilibrium with the SNP under study, positioned in the region where PCR primers anneal, there could be a skewed amplification. An asymmetrical amplification of the two homologue chromosomes could lead to errors in the calls of the SNP under study, causing a misclassification of the heterozygotes as homozygotes. Alternatively, a polymorphism in linkage disequilibrium with the one of interest could reside in the region where APEX primers anneal. In this case one of the two alleles could be biased during the APEX step. We should also consider that, although we used the direct sequencing as the gold standard to test MetaboChip, this method can be affected by possible biases. These biases include skewed amplifications of the two homologues and/or an unequal incorporation of the terminators, leading to an unbalanced height of the peaks and resulting in a potential underestimation of heterozygotes when the sequences are read by the software.

However, the overall numbers of discordances are in agreement with previous studies where APEX was shown to be a genotyping technique with high specificity and sensitivity (7,9,12,16).

It is important to recognize that throughput of the APEX technology is lower than that of other approaches, such as TaqMan® (Applied Biosystems, Foster City, CA, USA) (17) or Invader® (Third Wave Technologies, Madison, WI, USA) (18) assays, in terms of number of samples processed. In practice, it is difficult to process more than a few hundred samples with the MetaboChip in a reasonable time span. On the other hand, it is important to note that the sheer cost of SNP-specific probes for TaqMan or Invader assays makes these approaches less attractive when more than a handful of SNPs are studied. APEX, however, offers the advantage that SNP-specific reagents are more than one order of magnitude less expensive than TaqMan. In this study we selected only polymorphisms of interest for our studies that were represented in the SNP500 database, to provide a rigorous validation. However, it is possible to produce APEX microarrays with several hundred SNPs using the same off-the-shelf technology we used for the present study (19).

In summary, we screened a relatively large number of polymorphisms by APEX, representative of biochemical pathways involved in metabolism of xenobiotics and DNA repair. After the testing phase, we showed that 82 out of 88 SNPs are highly reliable. Thus, the low overall error rate combined with the low operating costs could make the MetaboChip an efficient tool for future studies of genetic epidemiology and pharmacogenetics.

ACKNOWLEDGMENTS

SNP500 samples were kindly provided by Dr. Stephen Chanock from the Core Genotyping Facility, Advanced Technology Center, National Cancer Institute, Bethesda, MD, USA. The authors wish to thank Dr. John Cheney (International Agency for Research on Cancer) for revision of the English language. S.L. is a recipient of a Marie Curie fellowship (IHP-MCIF-99-1) of the European Commission. F.G. is recipient of a fellowship of the International Association for the Study on Lung Cancer (IASLC), part of the Cancer Research Foundation of America (CRFA). This work was partially funded by a donation from the Institute for Scientific Interchange, Turin, Italy. S. Landi and F. Gemignani contributed equally to this manuscript.

REFERENCES

1. **Pisani, P., D.M. Parkin, and J. Ferlay.** 1993. Estimates of the worldwide mortality from eighteen major cancers in 1985. Implications for prevention and projections of future burden. Int. J. Cancer. *55*: 891-903.
2. **Parkin, D.M., P. Pisani, A.D. Lopez, and E. Masuyer.** 1994. At least one in seven cases of cancer is caused by smoking. Global estimates for 1985. Int. J. Cancer. *59*:494-504.
3. **Peto, R., S. Darby, H. Deo, P. Silcocks, E. Whitley, and R. Doll.** 2000. Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. BMJ *321*:323-329.
4. **Shields, P.G.** 1999. Molecular epidemiology of lung cancer. Ann. Oncol. (*10* Suppl.) 5: S7-S11.
5. **Syvänen, A.C.** 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat. Rev. Genet. *2*:930-942.
6. **Shumaker, J.M., A. Metspalu, and C.T. Caskey.** 1996. Mutation detection by solid phase primer extension. Hum. Mutat. *7*:346-354.
7. **Kurg, A., N. Tõnisson, I. Georgiou, J. Shu-**

maker, J. Tollett, and A. Metspalu. 2000. Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology. Genet. Test. *4*:1-7.

8. **Metspalu, A. and J.M. Shumaker.** 1999. DNA resequencing, mutation detection and gene expression analysis by oligonucleotide microchips. Biomethods *10*:371-397.
9. **Gemignani, F., S. Landi, C. Perra, F. Canzian, A. Kurg, N. Tõnisson, R. Galanello, A. Cao, et al.** 2002. Reliable detection of β-thalassemia and G6PD mutations by a DNA microarray. Clin. Chem. *48*: 2051-2054.
10. **Pastinen, T., M. Raitio, K. Lindroos, P. Tainola, L. Peltonen, and A.C. Syvanen.** 2000. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. Genome Res. *10*:1031-1042.
11. **Raitio, M., K. Lindroos, M. Laukkanen, T. Pastinen, P. Sistonen, A. Sajantila, and A.C. Syvanen.** 2001. Y-chromosomal SNPs in Finno-Ugric-speaking populations analyzed by minisequencing on microarrays. Genome Res. *11*:471-482.
12. **Tõnisson, N., J. Zernant, A. Kurg, H. Pavel, G. Slavin, H. Roomere, A. Meiel, P. Hainaut, and A. Metspalu.** 2002. Evaluating the arrayed primer extension resequencing assay of TP53 tumor suppressor gene. Proc. Natl. Acad. Sci. USA *99*:5503-5508.
13. **Gemignani, F., S. Landi, F. Vivant, S. Zienolddiny, P. Brennan, and F. Canzian.** 2002. A catalogue of polymorphisms related to xenobiotic metabolism and cancer susceptibility. Pharmacogenetics *12*:459-463.
14. **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.
15. **Vainrub, A. and B.M. Pettitt.** 2003. Surface electrostatic effects in oligonucleotide microarrays: control and optimization of binding thermodynamics. Biopolymers *68*:265-270.
16. **Tõnisson, N., A. Kurg, K. Kaasik, E. Lohmussaar, and A. Metspalu.** 2000. Unravelling genetic data by arrayed primer extension. Clin. Chem. Lab. Med. *38*:165-170.
17. **Morris, T., B. Robertson, and M. Gallagher.** 1996. Rapid reverse transcription-PCR detection of hepatitis C virus RNA in serum by using the TaqMan fluorogenic detection system. J. Clin. Microbiol. *34*:2933-2936.
18. **Ryan, D., B. Nuccie, and D. Arvan.** 1999. Non-PCR-dependent detection of the factor V Leiden mutation from genomic DNA using a homogeneous invader microtiter plate assay. Mol. Diagn. *4*:135-144.
19. **Dawson, E., G.R. Abecasis, S. Bumpstead, Y. Chen, S. Hunt, D.M. Beare, J. Pabial, T. Dibling, et al.** 2002. A first-generation linkage disequilibrium map of human chromosome 22. Nature *418*:544-548.

Received 11 June 2003; accepted 25 June 2003.

Address correspondence to:

Federico Canzian
Genome Analysis (GAN) Group
International Agency for Research on Cancer
150, cours Albert-Thomas
F-69372 Lyon, France
e-mail: canzian@iarc.fr