Human Identification by Genotyping Single Nucleotide Polymorphisms (SNPs) Using an APEX Microarray

Lisa D. White[1,2], John M. Shumaker[2], Jeffrey J. Tollett[2], and Rick W. Staub[1]
[1]Identigene, Inc., 7400 Fannin, Suite 1222, Houston, TX 77054
[2]Dept. Molecular & Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston TX 77030

ABSTRACT

Interest in utilizing single nucleotide polymorphisms (SNPs) for genetic research is rapidly increasing. Nuclear SNPs are typically biallelic loci that segregate in a Medelian fashion and are the most common type of genetic variation in humans. They are eminently suited to a broad range of applications that include genome mapping, pharmacogenomics, and genotyping for disease diagnostics and genetic identity. By applying Arrayed Primer Extension (APEX) technology to a suite of 24 SNP loci simultaneously, we have successfully developed a prototype identity chip for use in DNA paternity testing and forensic identification. These 24 SNP loci were chosen based upon reported human population allele frequencies of approximately 0.5 and produce a combined match probability of approximately 16 billion. Sense and antisense oligonucleotide chip primers were designed flanking each SNP site. The use of sense and antisense primers increases the accuracy of the chip by providing an internal quality control check (the sense SNP allele detected must be complementary to the antisense SNP allele). Amplification of the SNP loci is accomplished by modified multiplex PCR technology. A template-dependent DNA polymerase extends each primer in the array with a dye-labeled ddNTP terminator. Detection of the labeled SNP sites is performed by a multicolor (4-color) Total Internal Reflection Fluorescence (TIRF) reader. Future development of the identity microarray will concentrate on increasing the number of SNP Loci to 40-50 for paternity testing (exclusion power of 99.97-99.997%). This combination of microarray SNP scoring and efficient fluorescence detection has great potential for increasing sample throughput in human identity testing laboratories.

INTRODUCTION

Single nucleotide polymorphisms (SNPs) are single base variations that occur about every 1,000 bases in the human genome[1]. During the past year, interest in scoring SNPs for a multitude of different applications has increased dramatically. Genome mapping, gene expression, and genotyping for disease diagnostics (mutation detection and disease susceptibility) and genetic identity (paternity testing and forensic identification) are just a few of the applications currently under consideration. In addition to the vast array of applications possible when working with SNPs, the various methods of detection or scoring these single base variations are also increasing. Among them are DNA chip microarrays, mass spectrometry[2], DNA sequencing, molecular beacons[3,4], and PCR-based single strand conformational polymorphisms (SSCPs)[5].

DNA microarray technologies themselves are numerous. They range from hybridization methods such as cDNA expression chips[6], genomic hybridizations with BACs and other genomic clones, and re-sequencing for genotyping or mutation detection[1], to other methods such as Arrayed Primer Extension[7] (APEX) to extend a single base for mutation detection and genotyping.

We are interested in scoring SNPs in a microarray format for genotyping individuals for genetic identity. Of the methods listed above, we chose APEX because detection of the signal is by extension of a single dye-labeled terminator and does not wholly depend upon hybridization. By applying APEX to a suite of 24 SNP loci that span the human genome we have successfully developed a prototype identity chip for use in DNA paternity testing and forensic identification.

MATERIALS AND METHODS

After reviewing the literature we chose 24 SNP loci with reported allele frequencies of approximately 0.5 (Table 1). The SNP site in each locus was identified and primers (25-mer oligonucleotides, 5' -aminated) were designed from the known DNA sequence directly flanking the site in both the sense and antisense orientation. These APEX primers were then arrayed onto silane-treated coverglasses (50.0 µM oligo in 80.0 mM NaOH).

Amplification primers were designed so that the amplicons ranged in size from 50-300 bases and encompassed the SNP site. PCR was performed on the loci in three multiplex reactions [100.0 ng human genomic DNA (isolated from buccal swab samples using a standard Proteinase K/phenol/chloroform extraction protocol); 1X Platinum Taq reaction buffer without $MgCl_2$; 5.0 mM $MgCl_2$; 1.0 mM dNTPs (1:1:1:0.8:0.2 ratio of A:C:G:T:U respectively); 0.2 µM each amplification primer; 2.5 units Platinum Taq (Gibco-Life Technologies, Inc.)] with the following conditions: 95°C – 10 minutes; 30[95°C – 30 seconds; 55°C – 2 minutes; 65°C – 2 minutes]; 65°C – 5 minutes.

Following amplification, the multiplex reactions for each individual were pooled and concentrated with Microcon-30 DNA microconcentrators (Amicon, Inc.). Amplicons were then treated with uracil-DNA-glycosylase and shrimp alkaline phosphatase for 1 hour at 37°C and then heat-treated at 96°C for 10 minutes for enzyme inactivation.

The array was washed 10 minutes in 96°C H_2O and then APEX reactions were carried out on the 24-locus identity chip at 48°C for 4-8 minutes with the following: 0.75µg multiplexed amplicons (degraded and phosphatased); 1X Thermosequenase reaction buffer; 1.25µM each dye-labeled ddTNP terminator (Cy3-ddATP, Cy2-ddCTP, Cy5-ddTTP, and Texas Red-ddGTP); 4 units Thermosequenase (Amersham-Pharmacia Biotech) in a total of 45.0 µl. After the reaction, the chip was washed again in 96°C H_2O for 10 minutes and then imaged on a 4-color Total Internal Reflection Fluorescence (TIRF) detector (Figure 1).

RESULTS

Genomic DNA (isolated from buccal swab samples) from individuals in seven parentage cases previously tested with conventional Short Tandem Repeat (STR) typing was tested by APEX on the prototype 24 locus SNP identity chip.

Multiplex PCR: Three multiplex PCR reactions were performed on each DNA sample for each of the seven cases. Figure 2 is a representative ethidium bromide stained agarose gel with the multiplexed amplicons (MP1, MP2, and MP3) from a single case (#8866). Case #8866 is a TRIO (M: mother, C: child, AF: alleged father). Although many amplicons were detectable by eye on the gel, some of the PCR products were indistinguishable by electrophoresis because they were the same size as other products within the multiplex. Those undetected amplicons were tested by actually running APEX on the microarray. All 24 amplicons were accounted for either by electrophoresis or APEX. We are currently attempting to combine the amplifications into a single multiplex reaction.

APEX: APEX is a powerful method for scoring SNP sites throughout the genome. In the reaction, a template-dependent DNA polymerase finds the 3' end of each oligo in the microarray as it hybridizes to its specific SNP amplicon and then extends a single base (the polymorphic site) with fluorescently labeled ddNTP terminators (Figure 3). After the reaction, the labeled array is imaged with a CCD camera on the TIRF 4-color detector. Results from our prototype 24 locus array indicate that APEX provides an unambiguous way to score SNP loci.

One 24 locus chip will genotype a single individual at 24 different loci simultaneously. Four fluorescently-labeled ddNTPs (four different dyes) are used in the APEX reaction and then imaged on the TIRF detector (Figure 1). The TIRF detector illuminates the chip with four lasers. As each laser causes the bound fluorophores to fluoresce, a CCD camera images the array. The results from each microarray then consist of four images (one for each fluorescent dye). Figure 4 is a Cy5-ddTTP CCD image of our prototype chip. Figure 5 is a close-up of quadrant one of the chip. The four images are CCD camera images from one individual (Cy3-ddATP, Cy5-ddTTP, Texas Red-ddCTP and Cy2-ddGTP). Seven of the 24 loci on this chip are arrayed (3 replicates per locus) in this quadrant.

Each locus in the array is indicated by colored squares and there are two lines (sense and antisense), reading from bottom to top, per locus. The results from this chip allowed us to genotype a single individual at 24 loci simultaneously using the combination of APEX microarray and 4-color TIRF detector.

Analysis of SNP scoring and Paternity Index calculations: Individuals from seven parentage cases were genotyped using APEX on the prototype identity chip. Results from these cases are shown in Table 1. Of the seven cases tested by SNPs, two were inclusionary (the alleged father could not be excluded) and five were exclusionary. Exclusions are indicated by a PI of zero and highlighted in yellow. These data corroborate the results previously obtained by typing these seven cases with conventional STR typing.

Individual PIs are likelihood ratios which compare the alleged father' s likelihood of contributing the obligate paternal allele for any SNP locus to that of a composite randomly chosen man of his race (Caucasian in these examples). The formulae used for these calculations are shown in Table 2. Note that whenever the mother and child are both heterozygous, the PI is 1, since either allele can be the paternal allele and the alleged father is certain to contribute an obligate paternal allele, as is the randomly chosen man.

The data in Table 3 show the type of information gained when using the 24 locus SNP array. The genotypes of individuals from two parentage cases: a TRIO (8875) and a MNT (8879, mother not tested) are shown. The cumulative PI for each case was calculated by multiplying the individual PIs for each locus, assuming independence of the loci utilized.

DISCUSSION

Human identity testing utilizing DNA has undergone an evolution from highly polymorphic multilocus RFLP probes, to less polymorphic single locus RFLP probes, to even less polymorphic STR systems, mostly to achieve laboratory automation. Future identity typing laboratories will undoubtedly utilize simple test systems such as SNPs in order to more easily automate the laboratory process.

We have shown here that it will be possible to develop a SNP testing battery on a single APEX microarray. These arrays will be capable of quickly genotyping humans in a highly discriminatory manner, enabling efficient forensic and parentage analyses. Although a good deal of improvement is necessary to efficiently automate the actual laboratory process, we have demonstrated here a prototype 24 locus SNP microarray that works quite well as an adjunct test in the resolution of parentage disputes.

To determine how many SNP loci will be necessary for an APEX human identity microarray that can function as a stand-alone test to efficiently process paternity trios through a parentage-testing laboratory, we have used two different methods. In the first method, the average exclusion potential (ability to exclude non-fathers) of each SNP is determined by calculating the average RMNE (random man not excluded) value and subtracting from 1. The average RMNE is determined by calculating the probability of obtaining each of the 15 possible genotype combinations using Hardy-Weinberg formulae and multiplying these by the appropriate individual RMNE values. Using this method, one can calculate that a SNP with two alleles each occurring at a frequency of 0.5 will have an average exclusion potential of 0.1875. Of course, this is the maximum value for exclusion potential of a SNP. Figure 6 shows average exclusion potentials for SNPs with varying allele frequencies.

Average RMNE values can be multiplied together to obtain cumulative RMNE values which can then be subtracted from 1 to obtain the cumulative average exclusion potential for a battery of SNPs with a particular allele frequency. These values are shown in Figure 7 for allele frequencies of 0.5, 0.4, and 0.3. It can be seen that acceptable exclusion powers are reached in the range of 30-50 SNPs for these allele frequencies. These range from about 0.9955 to 0.9999 depending on the number of SNPs and their allele frequencies. To perform paternity testing, we presently utilize a battery of six exceptionally powerful STRs that have a combined exclusion power of approximately 0.9995 in Caucasians.

The second method for determining how many SNPs will be necessary to successfully operate a paternity testing laboratory is to determine the frequency of paternity trio cases that will not reach a certain minimum paternity index (PI) value. Since paternity statutes in many states require that a PI of 100 be reached for the alleged father to be presumed the biological father, we have used this value for calculation purposes. It can be seen from Table 2 that Combinations 1, 5, 6, and 8 possess homozygous alleged fathers. For biallelic SNP loci with equivalent allele frequencies, Combinations 1, 5 and 8 will produce PIs of 2 while Combination 6 produces a PI of 1. All combinations where the alleged father is heterozygous produce a PI of 1, so 3/8 of the combinations produce a PI of 2 while 5/8 of all combinations produce a PI of 1. For N biallelic systems, (allele frequencies of 0.5), one must have at least seven systems for which the PI = 2 to obtain a cumulative PI that is greater than 100. Thus, the binomial theorem can be used to compute the probability of obtaining a PI less than 100 as follows:

$$P(PI < 100) = \sum_{i=0}^{6} \frac{N!}{(N-i)!i!}(3/8)^{i}(5/8)^{(N-i)}$$

Where N = Total SNP loci in test battery

i = Number of SNP loci where PI = 2.

When this formula is used for SNP batteries with allele frequencies of 0.5, it can be seen (Figure 8) that not until 30 SNPs are used in a testing battery will the number of cases requiring extended testing approach a reasonable value (~3%). Less than 1 in 10,000 cases will require extended testing if a battery of 50 SNPs is used. Interestingly, since many loci will be used in the test battery, it will be particularly important that these loci have low mutation rates.

Future work on the identity microarray will concentrate on increasing the number of SNP loci in the array to 40-50 for paternity testing (power of exclusion 99.97-99.997%). Following completion of the microarray, we will begin using the identity chip to gather gene frequency data on different population groups (Asian, Black, Caucasian, and Hispanic).

REFERENCES

[1]Wang, D. G., et al. Large-scale identification, mapping, and genotyping of single-nucleotide-polymorphism in the human genome. *Science* (1998) 280: 1077-1082.

[2]Ross, P. L., K. Lee, and P. Belgrader. Discrimination of single-nucleotide polymorphisms in human DNA using peptide nucleic acid probes detected by MALDI-TOF mass spectrometry. *Anal. Chem.,* (1997) 69: 4197-4202.

[3]Piatek, A. S., S. Tyagi, A. C. Pol, A. Telenti, L. P. Miller, F. R. Kramer, and D. Alland. Molecular beacon sequence analysis for detecting drug resistance in Mycobacterium tuberculosis. (1998) *Nature Biotechnology* 16: 359-363.

[4]van Schie, R. C. A. A., M. van Geel, S. A. E. Marras, and P. J. de Jong. Multicolored molecular beacons for fast and accurate detection of the biallelic TUB4q gene polymorphism. *Am. J. Hum. Genetics Supplement* 63: A242 (abstract).

[5]Orita, M., H. Iwahana, K. Kanazawa, K. Hayashi, and T. Sekiya. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci., USA* (1989) 86: 2766-2770.

[6]Schena, M., D. Shalon, R. W. Davis, and P. O. Brown. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* (1995) 270: 467-470.

[7]Shumaker, J.M., A. Metspalu, and C. T. Caskey. Mutation detection by solid phase primer extension. *Human Mutation* (1996) 7: 346-354.

[8]Syvänen, A., A. Sajantilla, and M. Lukka. Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing. *Am. J. Hum. Genet*. (1993) 52: 46-59.

[9]Hoog, Jan-Olov, L. O. Heden, K. Larsson, H. Jornvall, and H. Bahr-Lindstrom. The Y_1 and Y_2 subunits of human liver alcohol dehydrogenase cDNA structures, two amino acid replacements, and compatibility with changes in the enzymatic properties. In: *European Journal of Biochemistry* (1986) 159: 215-218.

[10]Peters, C., B. Schmidt, W. Rommerskirch, K. Rupp, M. Zuhlsdorf, M. Vingron, H. E. Meyer, R. Rohlmann, and K. von Figura. Phylogenetic conservation of arylsulfatases. *Journal of Biological Chemistry* (1990) 265-6: 3374-3381.

[11]Leitersdorf, E. and H. H. Hobbs. Human LDL receptor gene: HincII polymorphism detected by gene amplification. Nuc. Acids Res. (1988) 16:7215.

[12]Horn, G.T., B. Richards, J. J. Merrill, and K. W. Klinger. Characterization and Rapid Diagnostic Analysis of DNA Polymorphisms Closely Linked to the Cystic Fibrosis Locus. In: *Clinical Chemistry* (1990) 36-9: 1614-1619.

[13]Ploos van Amstel H.K., P. H. Reitsma, P. E. van der Loft, and R. M. Bertina. Intron-exon organization of the active human protein S gene PSα and its pseudogene PSβ: duplication and silencing during primate evolution. *Biochemistry* (1990) 29: 7853-7861.

[14]Kretzschmar, H.A., L. E. Stowring, D. Westaway, W. H. Stubblebine, S. T. Prusiner, and S. J. Dearmond. Molecular cloning of a human prion protein cDNA. *DNA* (1986) 5:315-324.

[15]Lachance, Y., V. Luu-The, C. Labrie, J. Simard, M. Dumont, Y. de Launoit, S. Guerin, G. Leblanc, and F. Labrie. Characterization of human 3β-Hydroxysteroid dehydrogenase/Δ^5-Δ^4-isomerse gene and its expression in mammalian cells. *Journal of Biological Chemistry* (1990) 265-33: 20469-20475.

[16]Oka, K, G. T. Tkalcevic, J. Stocks, J. D. Galton, and W. V. Brown. Nucleotide sequence of PvuII polymorphic site at the human lipoprotein lipase gene locus. *Nuc. Acids Res.* (1989) 17:6752.

[17]Pagter-Holthuizen, P., M. Jansen, R. A. van der Kammen, F. M. A. van Schaik, and J. S. Sussenbach. Differential expression of the human insulin-like growth factor II gene. Characterization of the IGF-II mRNAs and a mRNA encoding a putative IGF-II-associated protein. *Biochimica et Biophysica Acta* (1988) 950: 282-295.

[18]Tsujimoto, Y., and C. M. Croce. Analysis of the structure, transcripts, and protein products of bcl-2, the gene involved in human follicular lymphoma. *Proc. Natl. Acad. Sci., USA* (1986) 83: 5214-5218.

[19]Kwok, P. Y., Q. Deng, H. Zakeri, S. L. Taylor, and D. A. Nickerson. Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* (1996) 31: 123-126.

[20]Yasuda, T., K. Kishi, Y. Yanagawa, and A. Yoshida. Structure of the human deoxyribonuclease I (Dnase I) gene: identification of the nucleotide substitution that generates its classical genetic polymorphism. *Ann. Hum. Genet*. (1995) 59: 1-15.

[21]McPherson, R., S. M. Grundy, R. Guerra, and J. C. Cohen. Allelic variation in the gene encoding the cholesteryl ester transfer protein is associated with variation in the plasma concentrations of cholesteryl ester transfer protein. *J. Lipid Res.* (1996) 37:1743-1748.

[22]Wagner, F.F., and W. A. Flegel. Polymorphism of the *h* allele and the population frequency of sporadic nonfunctional alleles. *Transfusion* (1997) 37: 284-290 & 674.

[23]Olsson, M. L., C. Hansson, N. D. Avent, I. E. Akesson, C. A. Green, and G. I. Daniels. A clinically applicable method for determining the three major alleles at the Duffy (FY) blood group locus using polymerase chain reaction with allele-specific primers. *Transfusion* (1998) 38: 168-173.

[24]Li, Y., G-R. Sun, Q. Zheng, D. H. Yoo, N. Bhardwaj, D. N. Posnett, M. K. Crow, and S. M. Friedman. Allelic variants of human TCR BV17S1 defined by restriction fragment length polymorphism, single strand conformation polymorphism, and amplification refractory mutation system analyses. *Human Immunology* (1996) 49:85-95

[25]Day, C. E., T. Zhao, and M. A. Robinson. Silent allelic variants of a T-cell receptor Vβ12 gene are present in diverse human populations. *Human Immunology* (1992) 34: 196-202.
[26]Taroni, F., E. Verderio, S. Fiorucci, P. Cavadini, G. Finocchiaro, G. Uziel, E. Lamantea, C. Gellera, and S. DiDonato. Molecular characterization of inherited carnitine palmitoyltransferase II deficiency. *Proc. Natl. Acad. Sci., USA* (1992) 89: 8429-8433.
[27]Ogami, K., M. Hadzopoulou-Cladaras, C. Cladaras, and V. I. Zannis. Promoter elements and factors required for hepatic and intestinal transcription of the human ApoCIII gene. *Journal of Biological Chemistry* (1990) 265: 9808-9815.
[28]Venta, P. J., and R. E. Tashian. PCR detection of a BstNI RSP in exon 6 of the human carbonic anhydrase II locus, CA2. *Nucleic Acids Res.* (1991) 19:4795.
[29]Tiller, G. E., M. A. Weis, P. A. Polumbo, H. E. Gruber, D. L. Rimoin, D. H. Cohn, and D. R. Eyre. An RNA-splicing mutation ($G^{+5IVS20}$) in the Type II collagen gene (COL2AI) in a family with spondyloepiphyseal dysplasia congenita. *Am. J. Hum. Genet.* (1995) 56:388-395.
[30]Brenner, C.H. A note on paternity computation in cases lacking a mother. *Transfusion* (1993) 33:51-54.

ACKNOWLEDGEMENT

This work is supported in part by the U.S. Department of Commerce Advanced Technology Program Tools for DNA Diagnostics Grant #70-NANB5H1112 to the APEX Consortium. APEX Consortium members: Amersham-Pharmacia Biotech, Baylor College of Medicine, Duke University, and Identigene, Inc.

Table 1: 24 Locus SNP Identity Chip and Calculated Paternity Indices

SNP	Ref.	Chromosome	Polym.	%*	Case1	Case2	Case3	Case4	Case5	Case6	Case7
ADH3	8,9	4	A-G	56	1.79	0.89	1.00	1.09	0.00	0.89	1.79
ARSB	8,10	5	A-G	67	3.03	1.13	1.00	1.00	1.52	1.00	1.49
LDLR	8,11	19	T-C	45	1.00	1.01	1.00	1.00	0.91	1.00	1.00
METH	8,12	7	T-C	58	1.00	2.38	0.86	1.00	1.72	2.38	1.00
PROS1	8,13	3	T-C	58	0.86	1.03	1.00	1.00	1.00	0.00	1.00
PRP	8,14	20	A-G	66	1.00	1.11	1.00	1.00	1.00	1.00	1.00
HSD3B	8,15	1	A-C	77	2.17	1.41	1.00	1.00	1.00	0.65	0.00
LPL	8,16	8	A-G	52	1.00	2.08	1.00	1.00	0.96	1.00	1.92
IGF2	8,17	11	A-G	20	2.50	1.56	1.00	1.00	2.50	2.50	0.00
BCL2	8,18	18	A-G	56	1.79	0.89	1.00	1.79	0.00	1.00	0.89
W1-1417	19	8	C-T	48	1.00	0.96	1.00	0.00	1.04	0.00	1.92
D3S2344	19	3	G-C	48	1.00	1.00	1.92	1.92	1.00	1.00	0.00
D2S1301	19	2	G-A	55	1.00	1.11	0.91	0.00	1.00	1.82	0.00
D7S1760	19	7	T-C	50	2.00	1.00	1.00	1.00	1.00	1.00	1.00
DNASE1	20	16	A-G	56	0.89	2.27	1.00	0.89	2.27	2.27	1.79
CETP-1	21	16	C-A	53	1.89	1.89	1.89	1.89	1.89	1.89	1.89
FUT1	22	19	A-T	50	2.00	2.00	0.00	0.00	1.00	1.00	1.00
DUF-1	23	1	A-G	41	2.44	1.22	2.44	0.00	1.69	2.44	0.85
TRCVB17	24	7	C-T	48	1.00	1.92	0.96	1.04	1.92	0.96	0.00
TCRVB12	25	7	C-T	53	2.13	2.13	1.06	2.13	2.13	1.06	1.00
CPT1	26	1	G-A	49	1.02	1.00	0.98	0.00	1.00	1.02	1.00
APOC3	27	11	T-C	58	0.86	1.72	1.19	0.86	0.86	2.38	1.19
CA2	28	8	T-C	50	1.00	2.00	0.00	0.00	1.00	2.00	1.00
COL2A1	29	12	C-T	48	2.08	1.00	1.04	1.00	1.04	0.00	1.00
Cumulative PI					2901	1802	0	0	0	0	0

Table 1. Table showing the 24 SNP loci arrayed on the prototype identity chip and the paternity indices (PIs) for each locus. The cumulative PI is calculated by multiplying the individual PIs for each locus. DNA samples from seven paternity cases were genotyped using the prototype 24 locus SNP identity chip. These seven cases were previously genotyped with conventional STR typing at Identigene, Inc. Case 1 (TRIO) and Case 2 (Mother Not Tested, MNT) were found to be inclusionary (i.e., the alleged father could not be excluded with 6 unlinked STR genetic loci and had paternity indices of 6,316,803 and 1,635, respectively). Cases 3-7 were exclusionary cases. In all five exclusionary cases the alleged father was excluded from paternity by 2 or more STR loci. Our results from scoring 24 SNP loci in a DNA microarray using APEX corroborate the results of STR testing on these seven cases. The two true fathers are included with a probability of paternity exceeding 99.9% (Prior Probability = 50%). The five non-fathers are excluded by two or more SNPs in all cases. Exclusionary SNPs are highlighted in yellow. *Allele frequencies reported in the literature.

Table 2: Formulae used to calculate paternity indices in biallelic SNP systems

Combination	M	C	AF	Paternity Index
1	X	X	X	1/x
2	X	X	XY	0.5/x
3	X	XY	XY	0.5/y
4	XY	XY	XY	1/(x+y) = 1
5	XY	X	X	1/x
6	XY	XY	X	1/(x+y) = 1
7	XY	X	XY	0.5/x
8	X	XY	Y	1/y
9	Not tested	X	X	1/x
10	Not tested	XY	X	0.5/x
11	Not tested	XY	XY	0.25(x+y)/xy = 0.25/xy
12	Not tested	X	XY	0.5/x

Table 2. PI formulae for the 8 possible combinations of SNP phenotypes in paternity trios and 4 possible combinations of SNP phenotypes in cases where the mother is not tested. Homozygosity is assumed when individuals only display a signal for one nucleotide at any SNP locus. X and Y represent the nucleotide signals obtained while x and y are the gene frequencies for the respective SNP alleles. It should be noted that the calculations for combinations 10 and 11 assume that the mother and alleged father are of the same race (as they are in all cases presented here). If they differ in race, the PI formula for Combination 10 becomes PI = $x_M/(x_{AF}y_M+x_My_{AF})$, where x_M = freq(X) from mother' s racial database, x_{AF} = freq(X) from AF' s racial database, y_M = freq(Y) from mother' s racial database, y_{AF} = freq(Y) from AF' s racial database. Likewise, the PI formula for Combination 11 becomes PI = $(x_M+y_M)/[2(x_{AF}y_M+x_My_{AF})]$. Formulae derived according to Brenner.[30]

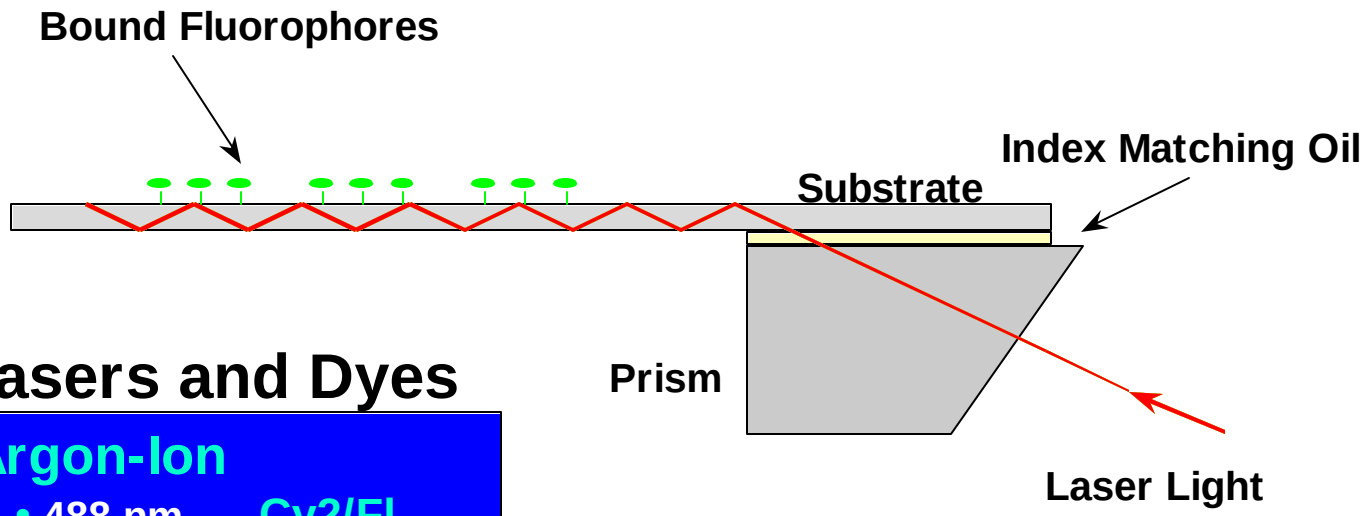
Table 3: Genotypes of Individuals Scored from 24 Loci on Identity Chip

SNP Locus	Polym	%	TRIO 8875M	8875C	8875AF	8875 PI calc.	8875 PI	MNT 8879C	8879AF	8879 PI calc.	8879 PI
ADH3	A-G	56	AA	AA	AA	1/0.56	1.79	AA	AG	.5/.56	0.89
ARSB	A-G	67	GG	GG	FF	1/0.33	3.03	AG	AG	1/(4*.67*.33)	1.13
LDLR	T-C	45	TC	TC	TC	1	1.00	TC	TC	1/(4*.45*.55)	1.01
METH	T-C	58	TC	TC	TT	1	1.00	CC	CC	1/.42	2.38
PROS1	T-C	58	TT	TT	TC	0.5/0.58	0.86	TC	TC	1/(4*.58*.42)	1.03
PRP	A-G	66	AG	AG	AA	1	1.00	AG	AG	1/(4*.66*.34)	1.11
HSD3B	A-C	77	CC	CC	AC	0.5/0.23	2.17	AC	AC	1/(4*.77*.23)	1.41
LPL	A-G	52	AG	AG	GG	1	1.00	GG	GG	1/.48	2.08
IGF2	A-G	20	AA	AA	AG	0.5/0.2	2.50	AG	AG	1/(4*.2*.8)	1.56
BCL2	A-G	56	AA	AA	AA	1/0.56	1.79	AG	AA	.5/.56	0.89
W1-1417	C-T	48	TC	TC	TT	1	1.00	TT	TC	.5/.52	0.96
D3S2344	G-C	48	GC	GC	GG	1	1.00	CG	CG	1/(4*.48*.52)	1.00
D2S1301	G-A	55	AG	AG	GG	1	1.00	AG	AA	.5/.45	1.11
D7S1760	T-C	50	TT	TC	CC	1/0.5	2.00	TT	TC	.5/.5	1.00
DNASE1	A-G	56	AA	AA	AG	0.5/0.56	0.89	GG	GG	1/.44	2.27
CETP-1	C-A	53	CC	CC	CC	1/0.53	1.89	CC	CC	1/.53	1.89
FUT1	A-T	50	AA	AA	AA	1/0.5	2.00	AA	AA	1/.5	2.00
DUF-1	A-G	41	AG	AA	AA	1/0.41	2.44	AA	AG	.5/.41	1.22
TRCVB17	C-T	48	TC	TC	TT	1	1.00	TT	TT	1/.52	1.92
TCRVB12	C-T	53	TT	TT	TT	1/0.47	2.13	TT	TT	1/.47	2.13
CPT1	G-A	49	AG	GG	AG	0.5/0.49	1.02	AG	AG	1/(4*.49*.51)	1.00
APOC3	T-C	58	TT	TT	TC	0.5/0.58	0.86	TT	TT	1/.58	1.72
CA2	T-C	50	CC	CC	TC	0.5/0.5	1.00	CC	CC	1/.5	2.00
COL2A1	C-T	48	TC	CC	CC	1/0.48	2.08	TC	TC	1/(4*.48*.52)	1.00
Cumulative PI							2901.30				1802.14

Table 3. Genotypes for the 24 loci on our prototype identity chip scored for two parentage cases (#8875 and #8879). Case #8875 is a TRIO (mother, child, alleged father) and case #8879 is a MNT (mother not tested). Both cases are inclusionary (i.e., the alleged father could not be excluded from paternity with 24 genetic loci). Paternity indices (PIs) for each locus were calculated (see Table 2) and the cumulative PI for each case is calculated as the product of the individual PIs. % = frequency of first-listed nucleotide from Polym. (polymorphism) column (taken from literature).

TIRF System

(Total Internal Reflection Fluorescence)



Lasers and Dyes

- Argon-Ion
  - 488 nm Cy2/Fl
- DPSS
  - 532 nm Cy3
- HeNe
  - 594 nm Texas Red
- Diode Laser
  - 635 nm Cy5

Figure 1. Layout of the TIRF detector. Lasers [Argon-Ion, DPSS (Diode Pumped Solid State), HeNe (Helium-Neon), Diode] are focused through a prism into the end of a glass slide. The light is then trapped within the glass, reflecting off of the inner surfaces as it travels the length of the slide before eventually exiting from the end. This internal reflection allows nearly instantaneous detection of the entire array and speeds the processing of arrays to approximately one minute/slide/4 colors.

Multiplex PCR for APEX reaction

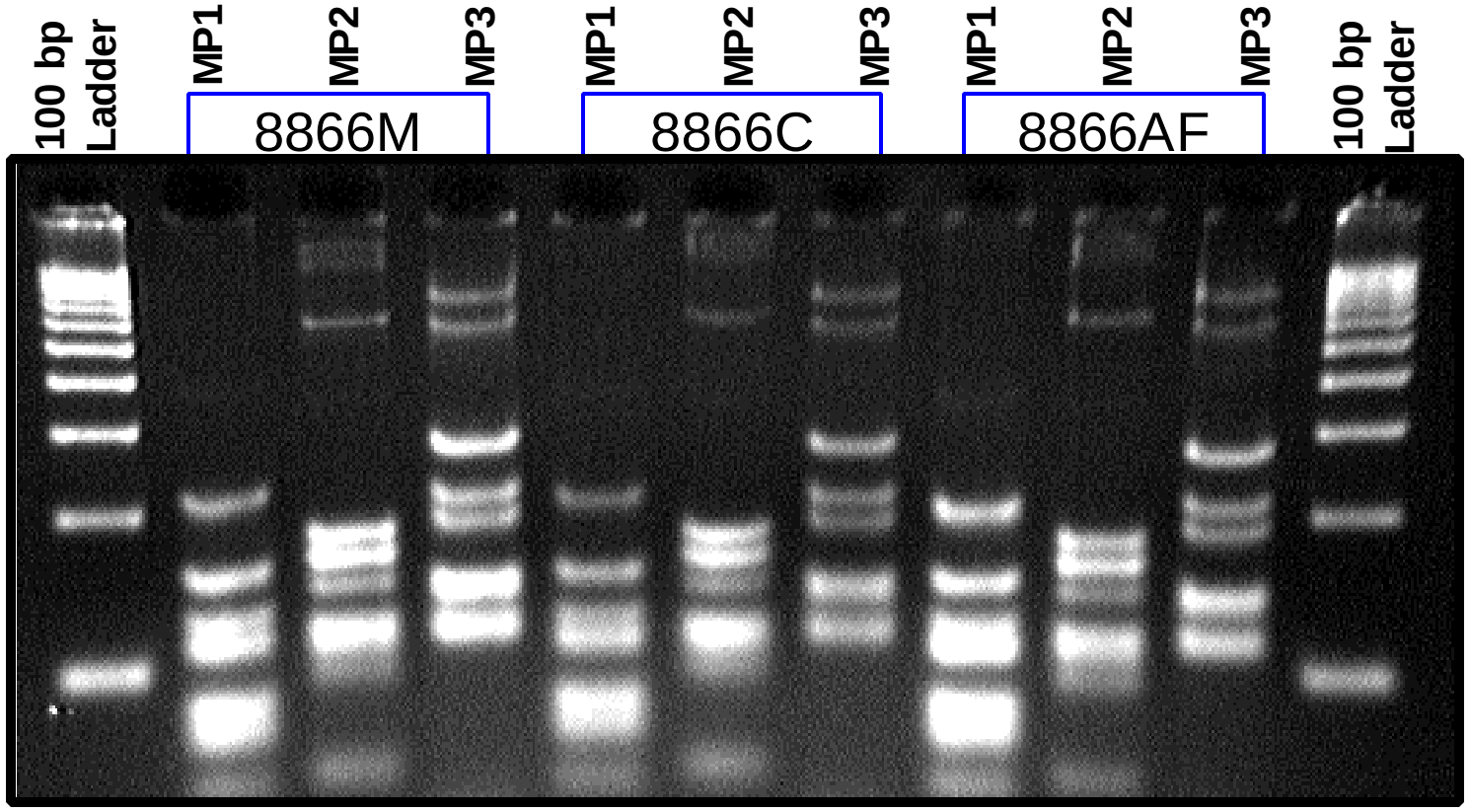
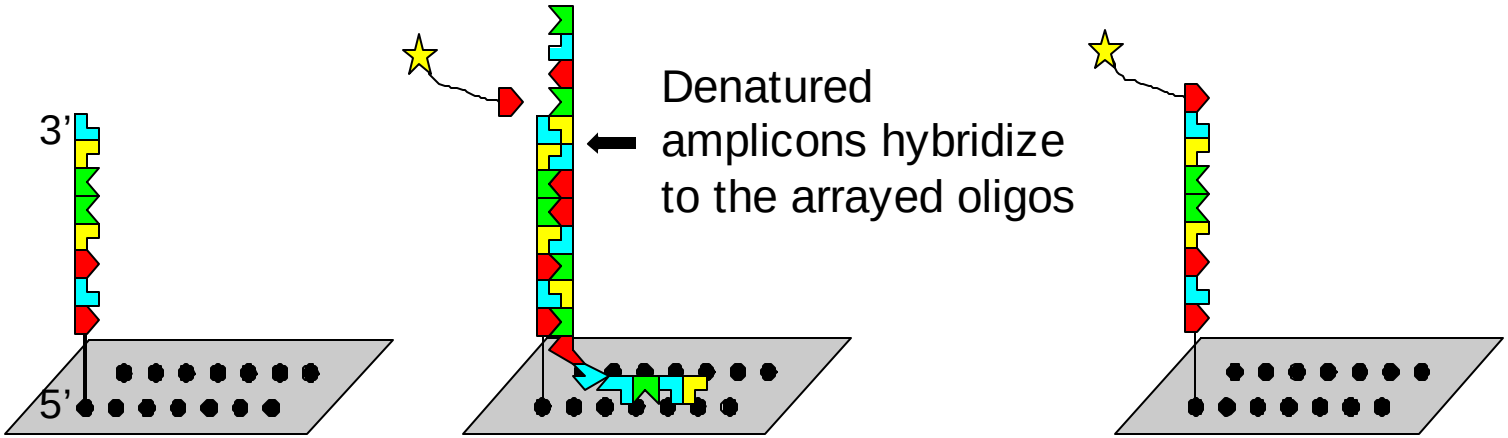


Figure 2. Image of ethidium bromide stained 4% agarose gel with multiplex PCR amplicons from a single case (#8866 TRIO). Three multiplex reactions were performed on genomic DNA from each individual in the case. MP1: multiplex 1 (ADH3, ARSB, LDLR, METH, PROS1, PRP, HSD3B, LPL, IGF2, BCL2); MP2: multiplex2 (WI-1417, D3S2344, D2S1301, D7S1760, FUT1, DUF-1, TCRVB17); MP3: multiplex 3 (TCRVB12, DNASE1, CETP-1, CPT1, APOC3, CA2, COL2A1). 8866M, mother; 8866C, child; 8866AF, alleged father.

APEX METHOD



Known oligos are arrayed in a site specific manner

Template-dependent ddNTP extension adds a single fluorescent labeled base to the oligo

Signal detection of fluorescent oligos with the TIRF detector

Figure 3. Sense and antisense oligonucleotide primers flanking each SNP site were aminated at the 5' end and arrayed onto silanized glass slides. Multiplexed PCR amplicons were denatured and pooled over the array with fluorescent labeled ddNTP terminators. The arrayed oligos were extended with the labeled terminators using Thermosequenase (Amersham-Pharmacia Biotech) in a template specific manner. The array was washed and the fluorescent signals were detected on a 4-color Total Internal Reflection Fluorescence (TIRF) detector and imaged with a CCD camera.

Human Identity Chip Array

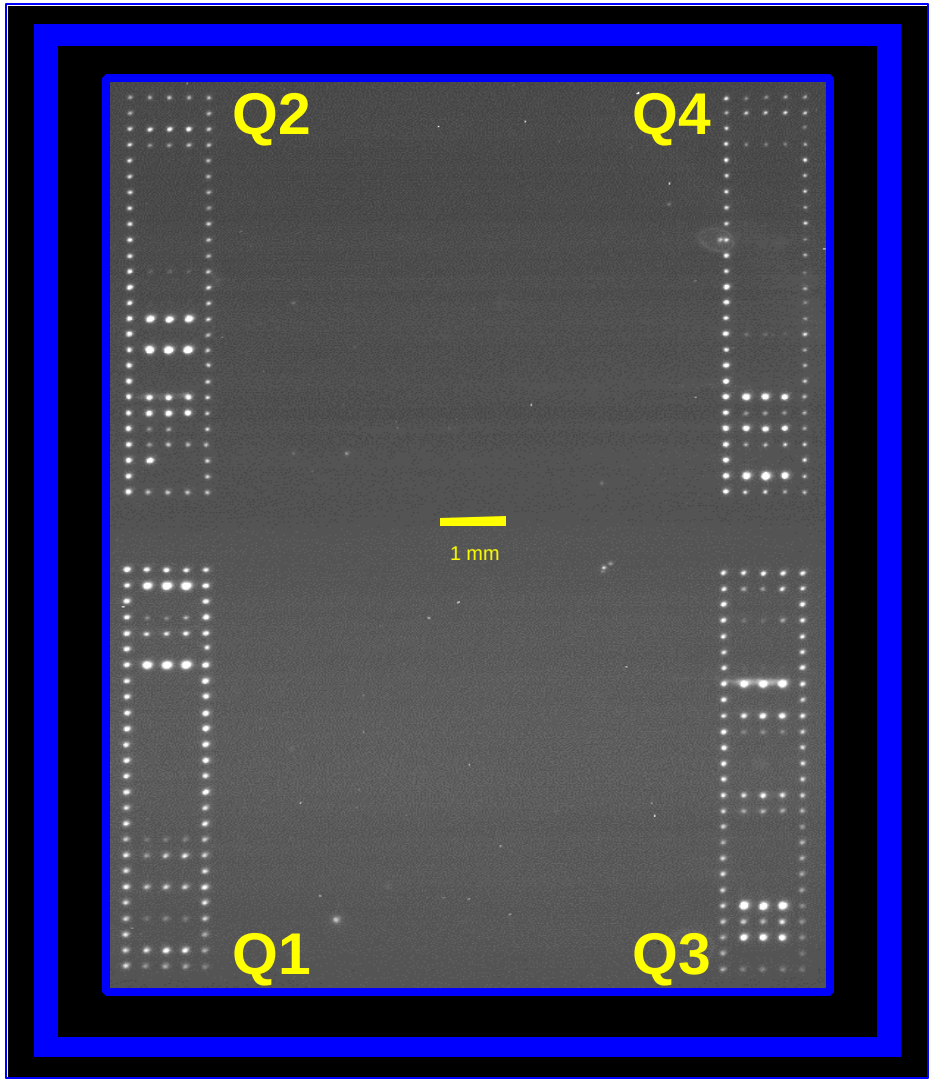


Figure 4. Cy5-ddTTP CCD image of prototype identity chip. The 24 loci are arrayed into four quadrants (Q1-Q4) with self-extending oligos as markers. Orientation is provided by a grid of self-extending "N"
oligonucleotides surrounding each quadrant in the array.

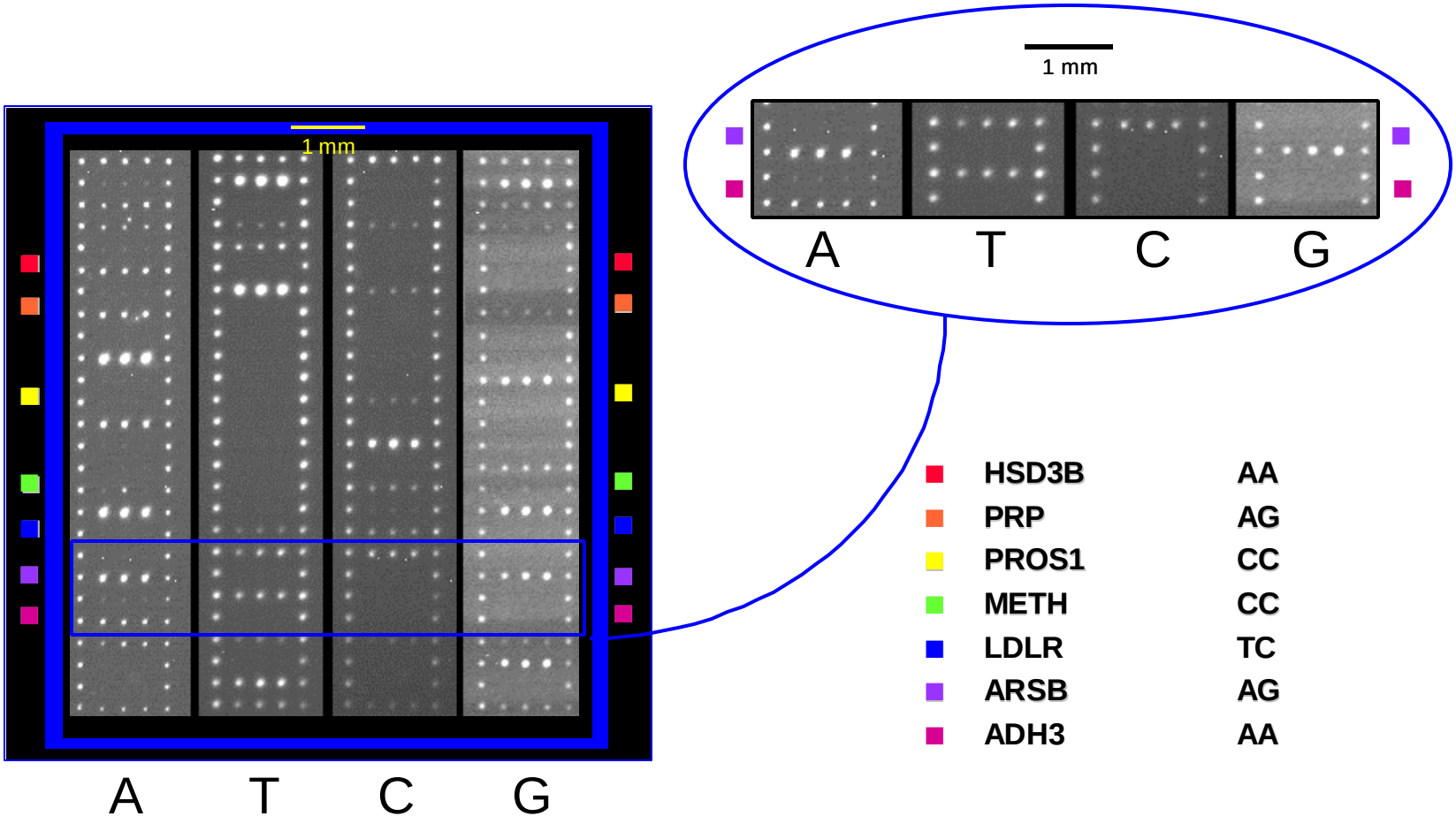

1 mm
A T C G
1 mm
A T C G
HSD3B AA
PRP AG
PROS1 CC
METH CC
LDLR TC
ARSB AG
ADH3 AA

Figure 5 . TIRF detector CCD camera image of quadrant 1 (Q1) of the SNP identity chip. This panel shows seven SNP loci arrayed (3 replicates per locus). Lane **A**: chip image of signals from fluorescent ddATP labeled oligos. Lane **T**: chip image of signals from fluorescent ddTTP labeled oligos. Lane **C**: chip image of signals from fluorescent ddCTP labeled oligos. Lane **G**: chip image of signals from fluorescent ddGTP labeled oligos. Reading from the bottom of the array toward the top, the array is: **N**, **T**, **G**, **N**, ADH3 sense (AA), ADH3 antisense (TT), ARSB sense (AG) , ARSB antisense (TC), LDLR sense (TC), LDLR antisense (AG), METH sense (CC), METH antisense (GG), **C**, **A**, PROS1 sense (CC), PROS1 antisense (GG), **A**, (space), PRP sense (AG), PRP antisense (TC), HSD3B sense (AA), HSD3B antisense (TT). The close up of the two loci ADH3 and ARSB (upper right) shows that (reading from the bottom to the top) ADH3 sense = A; ADH3 antisense =T; ARSB sense = AG, and ARSB antisense =TC. These results indicate that the individual tested on the chip is homozygous (AA) for ADH3 and heterozygous (AG) for ARSB. The genotypes of the other five loci are listed on the lower right.

Exclusion potential of SNPs

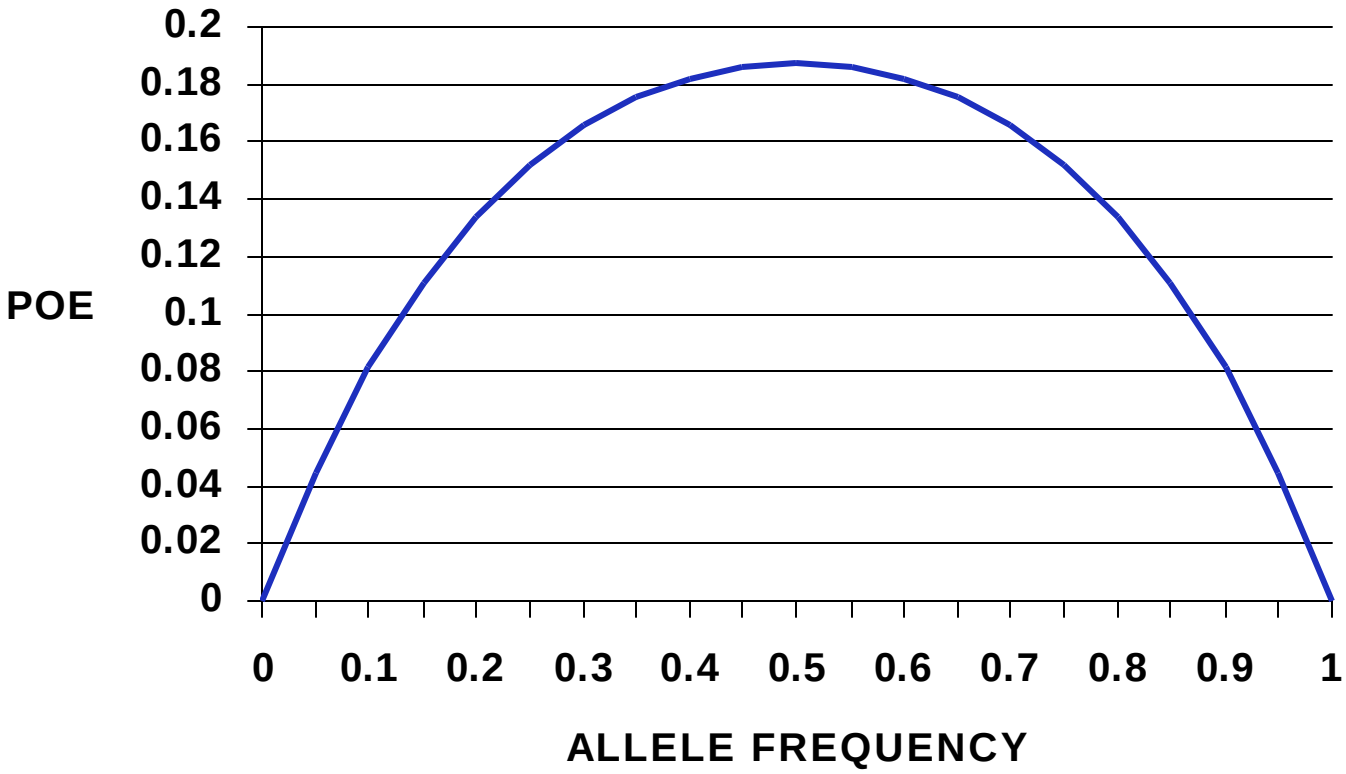


Figure 6. Plot of exclusion potentials vs. allele frequencies for single nucleotide polymorphisms. Exclusion potentials were calculated by subtracting the average RMNE (random man not excluded) value from 1. The average RMNE is determined by calculating the probability of obtaining each of the 15 possible genotype combinations using Hardy-Weinberg formulae and multiplying these by the appropriate individual RMNE values. The exclusion potential is maximized when the frequency of each allele is 0.5.

SNP exclusion power related to system number and allele frequency

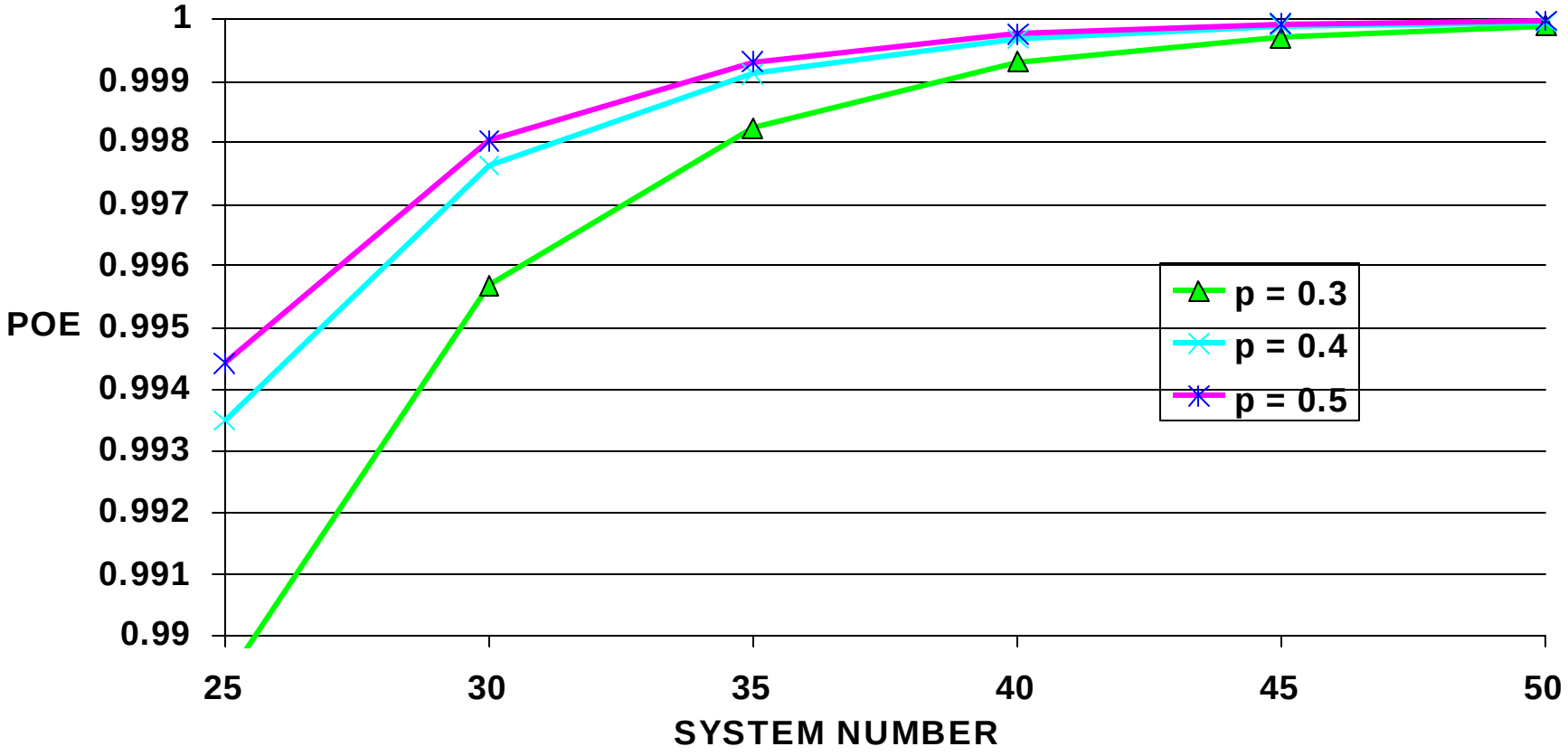


Figure 7. Proportion of non-fathers excluded with test batteries containing from 25 to 50 SNPs. Theoretically, 40-50 SNPs with allele frequencies in the range of 0.3-0.5 should have an average exclusion potential in excess of 0.999.

Resolution of paternity cases with SNPs (p=q=0.5)

SYSTEM #	P(PI < 100)
6	1.00000
10	0.96160
15	0.68522
20	0.32835
25	0.11560
30	0.03264
35	0.00785
40	0.00167
45	0.00032
50	0.00006

Figure 8. Using the binomial theorem, one can calculate the probability of obtaining a paternity index (PI) of less than 100 for test batteries with different numbers of SNPs that have allele frequencies of 0.5. If all cases must have a PI of at least 100, then one can determine how many cases would be extended for further testing when using various batteries. From this analysis, it appears that using a test battery containing at least thirty SNPs would be advisable to eliminate the need for frequent extended testing.