(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 April 2008 (24.04.2008)

PCT

(10) International Publication Number
WO 2008/048355 A2

(51) International Patent Classification:
C07H 21/04 (2006.01) *C12Q 1/68* (2006.01)
C12M 1/00 (2006.01)

(21) International Application Number:
PCT/US2007/005121

(22) International Filing Date:
27 February 2007 (27.02.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/777,078 27 February 2006 (27.02.2006) US

(71) Applicant *(for all designated States except US)*: **GENOMICS USA** [US/US]; 1660 Brittany Lane, Hoffman Estates, IL 60195 (US).

(71) Applicants: **HOGAN, Michael, E.**; 3801 E. Calle de Soto, Tuscon, AZ 85716 (US). **JAYARAMAN, Krishna**; 1660 Brittany Lane, Hoffman Estates, IL 60195 (US). **MITRA, Rahul**; 3623 Avery Court, Pearland, TX 77584 (US). **EGGERS, Frederick, H.**; 1937 E. Campbell Terrace, Tuscon, AZ 85718 (US).

(74) Agent: **ADLER, Benjamin, A.**; Adler & Associates, 8011 Candle Lane, Houston, TX 77071 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

WO 2008/048355 A2

(54) Title: POPULATION SCALE HLA-TYPING AND USES THEREOF

(57) Abstract: The present invention provides a portable system for real-time population-scale HLA genotyping and/or allelotyping in a field environment and methods of such population-scale HLA genotyping. The individual components of the system are portable to and operable within a field environment thereby providing high throughput with real-time geno- or allelotyping. Also provided are HLA gene-specific primers and HLA allele-specific or single nucleotide polymorphism-specific hybridization probes. In addition the present invention provides a microarray comprising the hybridization probes. Further provided is a kit comprising the HLA gene-specific primers and the microarray.

POPULATION SCALE HLA-TYPING AND USES THEREOF

BACKGROUND OF THE INVENTION

Cross-Reference to Related Applications

This nonprovisional application claims benefit of provisional U.S. Serial No. 60/777,078, filed February 27, 2006, now abandoned.

Field of the Invention

The present invention relates generally to the fields of microarray technology and population genotyping. More specifically, the present invention relates to a portable system and method of real-time high throughput population-scale HLA genotyping in a field environment.

Description of the Related Art

Bioterrorism and military interests have compelled the Department of Homeland Defense to invest heavily in high speed, flexible and high capacity methods of vaccine development. Recent studies have begun to confirm what basic immunology had predicted, namely that, within a large exposed population, individual response to infection and individual response to vaccination may vary greatly as a function of HLA type (1-2). However, only a few such studies have been performed to date, in part because HLA typing has been too expensive to implement as part of the epidemiology of infectious disease or the clinical epidemiology of vaccine development. Moreover, from the viewpoint of Homeland Defense, even if a thorough knowledge of the relationship between HLA type and infection or

vaccine response were known, and even if "personalized" vaccines were available based on the HLA type, the current technologies for HLA-typing do not have rapid field response capability and are too expensive and too complicated to be implemented in the context of a population-scale emergency.

Human immunogenic response to pathogens and vaccinations is dependent on the HLA loci. The response to pathogens is due to two distinct classes of polymorphic cell surface glycoproteins that are encoded by the HLA loci (3). HLA class I molecules identify the endogenous antigen present in the cytoplasm due to infection by bacteria or viruses and present it to the CD8+ cytotoxic T lymphocytes which kill the infected cells. HLA class II molecules also tag the infected cells by displaying exogenously derived epitopes on the surface of antigen-presenting cells for CD4+ helper T cells which results in an immune response against an invading pathogen. A diverse range of specificities for the epitope-HLA-Binding interaction is dependent on the extensive polymorphisms at the HLA loci.

Polymorphisms at the HLA loci are brought about by recombination, gene conversion and mutation and their natural selection in response to pathogens and infectious diseases (4). Hence, a diversity of HLA alleles enhances human ability to respond to and resist infectious and pathogenic agents at the population scale. HLA polymorphisms have been associated with several diseases and most recently with resistance to AIDS virus (5). Since most of the viral vaccines are viral surface antigens in a low dose, one's ability to react to such a vaccination is dependent on the polymorphism at the HLA loci. For example, the haplotype HLA-B8, SC01, DR3 lacks a response gene for hepatitis B virus surface antigen (6). In order to develop a vaccination it is very vital to find out the HLA type and classify the vaccine response to a set of known haplotypes.

The traditional serological methods for HLA typing have been limited to the availability of the allele-specific sera to identify structural differences due to single nucleotide polymorphisms (7). The antibodies used in the conventional methods are specific to HLA surfaces. However, structural differences in the peptide binding groove of HLA heavy chain due to single or multiple nucleotide polymorphisms cannot be easily identified using the antibody-based methods.

Nucleic acid based methods utilize sequence specific oligonucleotide probes (SSOP) or sequence specific primers (SSP). The sequence specific oligonucleotide probe

method is based on the use of either individual DNA samples or sequence specific oligonucleotide probes to identify the polymorphism (8). Current methods of primer design rely on simple BLAST like alignments to identify the primers and do not always perform well to pick out the unique primer set. Individual primers identified as specific to the loci are used to amplify the whole locus and specific probes are used to identify the polymorphism.

These are tiered approaches where the resolution is low to medium, and high resolution can be achieved by further probing with specific probes. The two versions of this method are dot blot where the DNA sample is immobilized on a membrane support and a labeled sequence specific oligonucleotide probe is allowed to hybridize to identify the polymorphism in the immobilized sample or a reverse dot blot where the sequence specific oligonucleotide probe is immobilized and a labeled DNA sample is added to the sequence specific oligonucleotide probe to identify the polymorphism. Immobilization of sequence specific oligonucleotide probes allows the testing of several polymorphisms, where as the immobilization of the DNA sample allows the testing of several samples for a specific polymorphism.

The sequence specific primer method uses specific primers targeted to each of the polymorphism (9). The number of primers required for the analysis of a locus depends on the number of polymorphisms in that particular locus. Typically, a large number of PCR reactions are needed to complete the HLA typing. This is a PCR based method where the presence or absence of a polymorphism results in amplification of the product. Using conventional gel electrophoresis the presence or absence of the PCR product can be ascertained. The PCR reactions contain positive control primers that amplify conserved regions.

Other methods are structure based or utilize sequencing methods. A structure-based method to identify polymorphisms is based on the fact that mismatched heteroduplexes containing looped out regions migrated differently than a heteroduplex without any mismatched loops in a non-denaturing gel (10). With the automation of DNA sequencing, HLA typing has been done on sequencing machines (11-12). The methodology is dependent on the number of polymorphisms and the number of exons, for example, for HLA class II the polymorphisms are in exon 2 which has a few hundred bases. In contrast, for class I typing

the polymorphisms require several exons to be sequenced and hence become more complicated and can result in errors.

Single nucleotide polymorphisms in the HLA types are shared by the several subtypes of the alleles. This could result in ambiguities when the conventional methods are used. In order to overcome this problem due to cross hybridization, a combination of probes and primers combined with the knowledge of the polymorphisms is essential. Hence, a simple SSOP or sequence specific primer hybridization might not result in the assignment of the HLA type.

The accurate assignment of HLA types is then based on carefully sifting through the patterns of a combination of probes for several subtypes. A PCR based method or a dot blot method would require a high amount of sample and would turn out to be very costly. Thus, a miniaturized technique that requires less amount of sample and is economical is needed. Microarrays (13) in combination with pattern recognition software provide such a platform to generate a 2-dimensional barcode to unambiguously identify the HLA type.

Microarrays are suited ideally for the high-throughput requirements in HLA typing. They offer the convenience of miniaturization and the ability to perform thousands of hybridizations in a single experiment. This highly parallel nature of the microarrays and their unique format makes them ideally suited for field use. In spite of these potential benefits, microarrays have not been perfected for field use in HLA typing. Cost, quality, and portability are among the limiting factors and are dependent on the method of manufacture.

Current microarrays in the market use specific dyes and so a specific type of imager needs to be used. Ideally, an imager should be able to image any dye. Also, current imagers in the market are not portable. Additionally, current analysis packages are equally cumbersome to use and require some manual intervention to identify the patterns.

The first olignucleotide microarray for the detection of allelic variants was reported in 1989 (14). Sequence specific oligonucleotide probes were spotted onto nylon membranes and hybridized to biotinylated CR products of the DNA samples. Genotype of the alleles was identified using the color intensity of the spots. More recently another study reported the use of a 130 probe element DNA microarrays to identify the allelic variations of class II polymorphisms (15). While the applicability of the microarrays to obtain medium to

high resolution HLA typing is obvious, the technology in its current form still suffers from several limitations, both technical and economical.

Additionally, using conventional methods, e.g., sequence specific oligonucleotide probes, the DNA sample is double stranded and the probe is single stranded. The presence of a double stranded product reduces the efficiency of hybridization. T7 or T3 polymerase sequences have been used to create single-stranded target molecules by in vitro transcription. Labeling RNA is difficult and hence the amplification methods utilize an end-labeled primer with biotin or a fluorescent dye so that all of the product can be labeled. The presence of biotin could interfere with the amplification procedure.

Furthermore, a significant limitation to performing population-scale HLA typing is the collection of the samples. Traditional methods of sample collection have focused on a blood draw of 10-15 ml by invasive procedures. This form of collection leads to a degradation, contamination and inaccurate results. Blood samples collected in this way would require a large scale handling, storage, and transportation problems that enormously increase the cost and logistical complexity of HLA typing. In addition to the handling and collection problems with the blood draw methods, the storage of isolated DNA becomes an issue. Hence, any technology for population-scale HLA typing must have alternate methods for sample collection and archiving the extracted DNA.

There is a need in the art for improvements in systems and methods for population-scale genotyping. Specifically, the prior art is deficient in a low cost, mass-produced and field-ready portable microarray system using advanced methods of genome analysis for rapid-response HLA typing of large populations. The present invention fulfills this long-standing need and desire in the art.

SUMMARY OF THE INVENTION

The present invention is directed to primers for amplifying an HLA gene. The HLA gene may be HLA A, B or DRB1 or an exon therein. The primers comprise sequences shown in SEQ ID NOS: 14-37.

The present invention also is directed to hybridization probes for detecting single nucleotide polymorphisms (SNPs) in an HLA gene. The hybridization probes comprise about a 9-15mer oligonucleotide complementary to a region containing the SNP and 5' and 3' flanking sequences. The SNPs may be located in HLA-A exon 2 or exon 3 or HLA-B exon 2 or exon 3. The probes comprise sequences shown in SEQ ID NOS: 48-291.

The present invention is directed further to a microarray device comprising a substrate having a cationic surface; and a monolayer comprising one or more of the hybridization probes described herein adsorbed thereto. In a related invention the microarray may comprise an oligo-thymidine co-absorbed with the hybridization probes described herein. In another related invention the oligo-thymidine may comprise a fluorescent dye attached thereto. In yet another related invention the microarray device may comprise a capping agent.

The present invention is directed further still to a kit comprising gene-specific primers for amplifying an HLA gene and the microarray device, both described herein. In a related invention the kit may further comprise buffers and polymerases for a PCR reaction or a fluorescent dye or a combination thereof.

The present invention is directed further still to a system for real-time high throughput population-scale HLA allelotyping in a field environment. The system comprises the microarray device described herein, means for collecting and purifying DNA samples from individuals comprising a population, means for generating by PCR cRNA target amplicons of one or more HLA genes of interest from the collected DNA, and means for assigning an HLA allelotype to each individual HLA gene of interest; wherein individual means and devices comprising said system are portable and operable in real time within the field environment. The primers described herein are useful to generate the cRNA target amplicons.

The present invention is directed further still to a method for real time population-scale HLA allelotyping in a field environment. The method comprises collecting DNA from one or members of the population, purifying the DNA for analysis and generating a target amplicon from an HLA gene of interest comprising the DNA the using gene specific primers described herein. The hybridization probes comprising the microarray described herein are contacted with the target and the hybridization pattern formed after the contact is imaged where each HLA allelotype has a pattern associated therewith. The present invention

is directed to a related method comprising further assessing a risk of infection by a biological agent or weapon for each individual based on the assigned allelotype. The present invention is directed to another related method comprising further assessing a response to a particular vaccine against the biological agent or weapon by each individual.

Other and further aspects, features, and advantages of the present invention will be apparent from the following description of the presently preferred embodiments of the invention. These embodiments are given for the purpose of disclosure.

BRIEF DESCRIPTIONS OF THE DRAWINGS

So that the matter in which the above-recited features, advantages and objects of the invention, as well as others which will become clear, are attained and can be understood in detail, more particular descriptions of the invention briefly summarized above may be had by reference to certain embodiments thereof which are illustrated in the appended drawings. These drawings form a part of the specification. It is to be noted, however, that the appended drawings illustrate preferred embodiments of the invention and therefore are not to be considered limiting in their scope.

Figures 1A-1J show gels of amplified PCR products of HLA-A and HLA-B loci (**Figures 1A-1B**), HLA-A exons 2 and 3 (**Figures 1C-1D**), HLA-B exons 2 and 3 (**Figures 1E-1F**), HLA-DRB1 locus (**Figures 1G-1H**), and HLA-DRB1 exon 2 (**Figures 1I-1J**) using various primer pairs from Table 2. PCR and gel conditions are described in Example 2.

Figures 2A-2B demonstrates that a 558 bp amplicon amplified using PCR from DNA extracts of five buccal samples collected using the "mouthwash" method and stored on FTA paper (**Figure 2A**) is intact and similar to freshly extracted human DNA from blood samples (**Figure 2B**). DNA was extracted and stored on FTA paper cards, then eluted via a GenVault DNA elution product (GenVault, Carlsbad, CA). PCR amplification was performed using standard methods and analyzed on agarose gels. Each lane consists of PCR amplified product from a 10 ng starting material of DNA. A positive control with a DNA sample to generate a 558 bp amplicon is shown in lane 6 and no amplicon was added in lane 7. The 1 kb size marker is shown in lane 8.

Figures 3A-3B are gels of UCLA reference and volunteer samples amplicons. **Figure 3A** shows a gel of the 558 bp amplicons generated using Fitzco Dacron cheek swabs. Lanes 2-4 show the DNA recovered using Argylla prep particles and lanes 5-7 show the DNA recovered using Qiagen clean up columns. Lane 1 is a molecular weight marker. **Figure 3B** is a gel showing the PCR products using specific primers and 4 UCLA reference standards 59, 15, 20 and 45 and DNA extracted from buccal swabs of two volunteers: MH and BI. 5 ng DNA of each, as assessed by PicoGreen ($1/100^{th}$ -$1/200^{th}$ of samples), was used in nested PCR to amplify HLA-B exon 2 yielding a 281 bp amplicon. Gel has $1/10^{th}$ of PCR product per lane. Lane 1 in both gels is the molecular weight marker.

Figure 4A-4B illustrates the effectiveness of the software package ImageAnalyzer in advanced automated image analysis on microarray images. **Figure 4A** shows a partially damaged microarray image section. **Figure 4B** shows the same image in Figure 4A after filtering, background compensation and precise gridding by ImageAnalyzer. Spot gridding is marked by circles.

Figures 5A-5D show the microarray patterns depicting the results of hybridization using capture probes shown in Table 5 for K-ras 1, K-ras 2 and K-ras 7. **Figure 5A** shows the hybridization of all of the targets wildtype and mutants 2-5. **Figure 5B** shows the binding of the homozygous wildtype and **Figure 5C** shows the binding of the homozygous mutant 7. **Figure 5D** shows the binding profile of the heterozygous target sample containing wildtype and mutant 2. Pattern recognition was visual.

Figures 6A-6B show a comparison of a UCLA reference sequence, re-sequencing and primary performance data from the HLA-B chip. The grey cells are UCLA allele types. The cells with bold fonts indicate array assay for one or both alleles; the cells with an 'underline' represents that the array is discordant with UCLA allele type and the cells with fonts in italics represent that the sequencing is discordant with UCLA allele type. All other cells represent results with 100% concordance with UCLA allele type.

Figures 7A-7D show HLA microarray images for UCLA reference samples 72, 21, 27, 57 respectively and **Figures 7E-7F** show HLA microarray images of DNA from buccal swabs of two volunteers MD and BI respectively. **Figure 7G** illustrates the quantification of spot intensity within the arrays of the top panel for codon 9. Data in Figure 7G is presented as six clusters. The first four clusters correspond to hybridization data from

UCLA reference samples of known allelotype at codon 9. Within each cluster, two sets of probe type were tested: a "long probe", i.e., comprising the oligo-T flanking sequences, and a "short probe" where the sequence specific sequence at the center is shortened by one base, in order to determine if specificity can be enhanced. **Figure 7H** illustrates similar hybridization results for HLA-B codon 50 with long probes (with flanking segments) and short probes (without flanking probes).

Figure 8 illustrates the process of determining spot size. P is considered a part of the spot if $D \leq R$, where D = square root of $(A^2 + B^2)$.

Figure 9 illustrates a decision tree based approach towards allele calling. The left arrow indicates that the condition in the diamond above is true and the right arrow indicates that the condition in the diamond above is false.

Figure 10 is a chart listing the 210 HLA alleles identified by automated searching of the PubMed database. Yellow shading indicates the allele is pesent in the UCLA class I panel. Green shading indicates the the allele is present in the UCLA class II panel. Orange shading indicates that the allele is present in either the UCLA panel at a higher resolution or at a lower noncoding resolution. Blue shading indicates that the allele is not typed in either UCLA panel.

DETAILED DESCRIPTION OF THE INVENTION

In one embodiment of the present invention there are primers for amplifying an HLA gene. In this embodiment the HLA gene may be HLA A, B or DRB1 or an exon therein.

In one aspect of this embodiment the HLA-A primers may have sequences shown in SEQ ID NOS: 14-15, the HLA-A exon 2 primers may have sequences shown in SEQ ID NOS: 20-21 and the HLA-A exon 3 primers may have the sequences shown in SEQ ID NOS: 22-26. In another aspect the HLA-B primers may have sequences shown in SEQ ID NOS: 16-19, the HLA-B exon 2 primers may have the sequences shown in SEQ ID NOS: 27-28 and the HLA-B exon 3 primers may have the sequences shown in SEQ ID NOS: 29-31. In

yet another aspect the HLA-DRB1 primers may have sequences shown in SEQ ID NOS: 32-37 and the HLA-B exon 2 primers may have the sequences shown in SEQ ID NOS: 38-47.

In another embodiment of the present invention there are provided hybridization probes for detecting single nucleotide polymorphisms (SNPs) in an HLA gene, comprising about a 9-15 mer oligonucleotide complementary to a region containing the SNP; and 5' and 3' flanking sequences.

In all aspects of this embodiment the flanking sequences may be oligo-thymidines or an oligo-thymidine-like polyanionic polymer. Also in all aspects the SNPs may be located in HLA-A exon 2 or exon 3, HLA-B exon 2 or exon 3 or HLA-DRB1 exon 2. In one particular aspect the HLA-A exon 2 probes may have the sequences shown in SEQ ID NOS: 48-99 and the HLA-A exon 3 probes may have the sequences shown in SEQ ID NOS: 100-155. In another particular aspect the HLA-A exon 2 probes may have the sequences shown in SEQ ID NOS: 156-239 and the HLA-A exon 3 probes may have the sequences shown in SEQ ID NOS: 240-291.

In yet another embodiment there is provided a microarray device microarray device for allelotyping an HLA gene, comprising a substrate having a cationic surface; and a monolayer comprising one or more of the hybridization probes described supra adsorbed thereto. Further to this embodiment the microarray may comprise and an oligo-thymidine co-absorbed with the hybridization probes. The oligo-thymidine may have about 20 to about 40 thymidines. Further still the oligo-thymidine may comprise a fluorescent dye linked thereto. In another further embodiment the microarray device may comprise a capping agent. In all embodiments the cationic surface may comprise an aminosilane, a guanidinium, tin oxide, aluminum oxide or zirconium oxide or other equivalently charged moiety. Also in all embodiments the substrate may be glass, plastic or a metal.

In a related embodiment the present invention provides a kit for population-scale HLA genotyping, comprising gene-specific primers for amplifying an HLA gene; and the microarray device described supra. Further to this embodiment the kit may comprise buffers and polymerases for a PCR reaction or a fluorescent dye or a combination thereof. Gene-specific primers may the sequences shown in SEQ ID NOS: 14-47.

In yet another embodiment of the present invention there is provided a system for real-time high throughput population-scale HLA allelotyping in a field environment,

comprising the microarray device described supra; means for collecting and purifying DNA samples from individuals comprising a population; means for generating by PCR DNA target amplicons of one or more HLA genes of interest from the collected DNA; and means for assigning an HLA allelotype to each individual HLA gene of interest; wherein individual means and devices comprising said system are portable and operable in real time within the field environment.

In all aspects of this embodiment the HLA gene may be HLA-A, HLA-B or HLA-DRB1. Also, in all aspects real time high throughput allelotyping is about 200 to about 300 HLA allelotypes per hour per system operated. In one aspect of this embodiment the the means for collecting DNA samples may comprise a container suitable to receive a buccal wash sample, a buccal swab sample or a blood sample collected from the individuals. In another aspect the means for generating target amplicons may comprise HLA gene-specific primers for amplifying the HLA gene of interest. Examples of the gene-specific primers have sequences shown in SEQ ID NOS: 14-47. In yet another aspect the means for assigning an HLA-allelotype to each individual may comprise an imaging device adapted to detect hybridization patterns formed on the microarray device after hybridization of the target to the hybridization probes adsorbed thereto; and pattern recognition software comprising a set of algorithms adapted to recognize the imaged hybridization patterns as HLA allelotypes. Examples of the hybridization probes have sequences shown in SEQ ID NOS: 48-291.

In yet another embodiment of the present invention there is provided a method for for real time population-scale HLA allelotyping in a field environment, comprising collecting DNA from one or members of the population; purifying the DNA for analysis; generating a target amplicon from an HLA gene of interest comprising the DNA using gene specific primers; contacting the hybridization probes comprising the microarray described supra with the target; and imaging the hybridization pattern formed after the contact wherein each HLA allelotype has a pattern associated therewith.

Further to this embodiment the method may comprise storing the collected DNA. In another further embodiment the method may comprise assessing a risk of infection by a biological agent or weapon for each individual based on the assigned allelotype. In yet another further embodiment the method may comprise assessing a response to a particular vaccine against the biological agent or weapon by each individual.

In all embodiments the DNA may be collected from blood, with a buccal wash or with a buccal swab. Also, the gene-specific primers may have sequences shown in SEQ ID NOS: 14-47. In addition, the hybridization probes have sequences shown in SEQ ID NOS: 48-291.

As used herein, the term, "a" or "an" may mean one or more. As used herein in the claim(s), when used in conjunction with the word "comprising", the words "a" or "an" may mean one or more than one. As used herein "another" or "other" may mean at least a second or more.

Provided herein is a human leukocyte antigen (HLA) chip and microarray technology to enable population-scale HLA-typing in a simple, portable and field-ready environment. The HLA chip is designed to thoroughly analyze the human HLA-B genotype. It is contemplated that HLA chips may be designed to analyze the entire human HLA loci. This microarray technology is effective to HLA type a large exposed population, for example, although not limited to, 100,000 individuals per week using as few as 5-10 low cost, portable field laboratories.

It also is contemplated that data acquired through population typing can be used in real time to anticipate, at the HLA level, individual risk of infection by a biological weapon or to anticipate personalized response to vaccination against the same infectious agent. Additionally, variants of the HLA chip may be used to provide field-ready neonatal screening in a third world environment or for battlefield-ready personnel identification. Furthermore, the HLA typing technology may be used for civilian identification during or after a disaster or for forensic applications. Thus, population-scale HLA typing has applications in military, anti-bioterrorism or epidemiological contexts.

Real-time interpretation of microarray data can be made by non-experts in a field application environment. This may be implemented in the field through low cost, compact, highly portable microarray imagers. Complete HLA analysis using equipment readily portable in a vehicle may have an overall process time from sample collection to a final HLA identification of less than 4 hours. This duty cycle can be maintained with a manual-only throughput of about 20 samples per person per 4 hours. With minimal sample handling automation, throughput routinely can be scaled up to a steady state of about 200 to

about 300 complete HLA genotypes per hour per workstation in a mobile, field-ready environment.

Furthermore, pattern recognition software, such as ImageAnalyzer (16), provides barcode-like simplicity in the image analysis and conversion of microarray fluorescence patterns into an HLA allelotype. The algorithms comprising the pattern recognition software use traditional and novel statistical and data mining approaches, such as, but not limited to, Euclidian and mutual information based distances and Fourier and wavelet transformations. The microarray pattern recognition software is effective for recognizing the patterns of hybridization and for extracting automatically information of what genomes/species are present based on the possibly fuzzy patterns of hybridization. This information then is compared to an existing database of patterns for each HLA subtype based on the expected pattern of spots. Such a system may include a user-friendly GUI interphase that can function on a laptop computer. A "cam-corder" sized portable microarray imager is suitable to work in a highly portable data collection environment.

The microarray design and fabrication or microfabrication requires no chemical modification of the probe ends, i.e., no linker, to immobilize the oligonucleotide probe to the surface (17). The surface saturation by the oligonucleotide occurs at a fraction of the concentration of olignucleotides required for covalent attachment. Thus, all of the oligonucleotide delivered to the surface is immobilized via adsorptive association with a monolayer surface that bears a net positive charge and additionally may be hydrophobic or hydrophilic. The shape and morphology of the spot is dependent of the initial contact of the drop dispensed by the microarrayer. Since there is no covalent bond formation, spot to spot variations are minimized.

The present invention provides short oligonucleotide probes of about 9 to about 15mers to discriminate among single nucleotide polymorphisms within the target. These probes are flanked with oligo-thymidine (oligo-T) sequences. Preferably the flanked probes comprise about 30 nucleotides total. It is further contemplated that analogues of oligo-T may be used as flanking sequences. Without being limiting an oligo-thymidine-like polyanionic polymer flanking sequence, e.g., polysulfonate, may replace the oligo-T flanking sequences. It is contemplated that the probes are designed for all the clinically relevant HLA subtypes. The current number of alleles at the IMGT/HLA sequence database is 977 alleles for HLA

Class I and 652 alleles for HLA Class II (18). Useful probes provided herein are effective to discriminate within HLA-A exon 2 (SEQ ID NOS: 48-99) and exon 3 (SEQ ID NOS: 100-155), HLA-B exon 2 (SEQ ID NOS: 156-239) and exon 3 (SEQ ID NOS: 240-291) and HLA-DRB1 exon 2.

Fabrication of the microarrays used herein uses an extremely simple and reproducible method (17) employing adsorptive, noncovalent attachment of the short oligonucleotide probes to the positive or cationic surface. For example, the cationic surfaces may comprise or may be coated with an amine function such as, although not limited to, aminosilane, or may comprise a guanidinium group. Alternatively, the surface may comprise a cationic metal or metal oxide, such as tin oxide, zirconium oxide or aluminum oxides or other metal oxides with a net positive charge or other equivalently charged moiety. Such oxide coatings may be particulate in nature or may be smooth and placed on a glass, plastic or metal substrate.

Generally, the method requires deposition or printing of oligo-T flanked oligonucleotide probes dissolved in water onto the cationic or net positively charged surface of the substrate. Alternatively, the oligo-T flanked probes may be co-printed with a second, constant oligonucleotide probe. This probe is the same in all instances of printing and may comprise an oligo-T sequence with about T20 to about T40 bases. The oligo-T sequence may comprise a dye linked thereto. An example of a dye may be, but not limited to, Cy-5

It is contemplated that the oligo-T sequence is inert with respect to nucleic acid hybridization to human DNA for HLA typing. The oligo-T is introduced as a marker to identify where the nucleic acid probes have printed, either by the direct detection of the oligo-T coupled to a dye or by oligo-T hybridization to dye-labeled oligo-adenine (oligo-A), the Watson-Crick complement of oligo-T. The inclusion of oligo-T improves the ability to orient hybridization image data for analyis and is useful for quality control during microarray fabrication.

In addition, on an amine coated surface, probe deposition may be followed by drying and capping of those residual surface charges or moieties not involved in direct association with adsorbed probe molecules. For example, and as known and standard in the art, capping of an aminosilane surface can be performed by reacting unused amine groups with a capping agent such as the surfactant sodium dodecylsulfate. Alternatively, for ceramic

or metal oxide surfaces, capping may be performed by reacting the surface with boric acid, fluoride ion or phosphate. After drying and capping, the attached oligonucleotides cannot be removed from the surface under standard hybridization and washing conditions, including high salt, 5M NaCl and high pH treatments. Thus, within the fabricated microarray, even though the adsorbed oligonucleotide is bound, presumably via multiple contacts to the surface and, therefore, may have lost configurational freedom required to form a perfect double helix with its cognate target, the product of such adsorptive coupling, followed by judicious capping to neutralize excessive charges on the surface, displays specificity for duplex formation which is as high as that seen in a standard solution state hybridization reaction or for surface hybridization to probes linked covalently to the surface at a single point.

PCR primers for the HLA locus are designed using novel algorithms to compute the frequency of occurrence of short subsequences, i.e., *n*-mers of olgionucleotide sequences of length of n= 5-25+ nucleotides in any genome within a reasonable time, e.g., minutes (19-20). These algorithms are used to perform a comparative statistical analysis of the presence of all possible "n-mers" in genomes of more than 250 microbial, viral and multicellular organisms, including humans. The results show a remarkable similarity of presence/absence distributions for different n-mers in all genomes. It suggests that the presence/absence distribution of n-mers in all genomes considered, provided that the condition $M<<4n$ holds, where M is the total genome sequence length, can be treated as nearly random. The massive computational analysis of the presence/absence of short subsequences in more than one genome simultaneously was performed for all published, i.e., prior to May 2002, microbial and virus genomes and was repeated for the 1600+ genomes which were available by May 2003. This produces unique sequences that are not repeated, anywhere in a given genome.

These new algorithms and data structures, together with the collection of 1600+ complete genomes, make it possible to significantly improve the quality of PCR primers design process itself. Using these algorithms, it is possible to find primers which appear exactly once and differ from the rest of n-mers in the entire human genome, including known SNPs, by at least 2 or 3 mismatches. Furthermore, primers are excluded which are present in some bacterial/viral genomes. Such massively parallel primer design is particularly important, when considering buccal washes as a population-scale DNA source, in that such

samples may contain up to 50% of some other contamination. The efficacy of this approach is demonstrated in the Examples presented herein in that a set of primers reported in a PCR/microarray study (15) prime several other chromosomes in addition.

Thus, the present invention provides primers or primer pairs effective to amplify HLA class I and class II loci. For example, for primary PCR reaction primers with SEQ ID NOS: 14-15 are useful to amplify the class I HLA-A locus and primers with SEQ ID NOS: 16-19 are useful to amplify the HLA-B locus. In secondary PCR reactions HLA-A exons 2 and 3 may be amplified with SEQ ID NOS: 20-21 and SEQ ID NOS: 22-26, respectively. HLA-B exons 2 and 3 may be amplified with SEQ ID NOS: 27-28 and SEQ ID NOS: 29-31, respectively. In addition, the class II HLA-DRB1 locus may be amplified in primary PCR reactions using primers with SEQ ID NOS: 32-37. Secondary PCR is performed to amplify HLA-DRB1 exon 2 using primers with SEQ ID NOS: 38-47. These primers are effective to amplify all alleles comprising the gene, i.e., the primers are gene specific and allele blind. The allele fine structure of interest within the entire amplified gene is determined by hybridization of the PCR product to probes in the microarray as provided herein.

DNA samples may be obtained from a spot of dried blood, from buccal wash DNA, DNA from a single 10 μl finger prick or DNA from a paraffin-embedded thin section. Preferably, the DNA samples are buccal DNA samples collected via the "mouthwash" method or buccal swab sample collected on Fitzco Dacron swabs (21). DNA extracted from the samples may be stored or archived on FTA paper (GenVault, Carlsbad, CA). This method of immobilization of the sample onto the treated FTA paper provides for archiving and, subsequently, for complete recovery of the DNA without degradation and suitable for PCR reactions.

The targets for the short oligonucleotide probes used herein are single stranded DNA transcribed from the amplified DNA sample or denatured double stranded PCR products. The amplification methods described herein produce labeled dsDNA PCR products. The DNA can be directly chemically labeled using modified PCR primers or cis-platinum conjugated dyes in one simple reaction (22) and is hydrolyzed at the same time with controlled alkaline treatment. This produces uniformly labeled DNA for hybridization.

The present invention also provides a kit comprising one or more primer pairs of HLA allele forward and reverse primers suitable to amplify exactly one or several DNA regions in a DNA sample. Optionally, the kit may further comprise one or more microarrays with HLA probes as described herein. Particularly, these kits may have primer pairs and, optionally, microarrays designed to hybridize HLA-A, HLA-B or HLADRB1 loci and determine which SNPs are present in the sample, thereby genotyping an individual or a population. Furthermore, the kits may comprise buffers and polymerases for a PCR reaction.

Particularly, the present invention provides a method for genotyping and allelotyping that uses human DNA collected from blood, buccal swabs or buccal wash. Collected samples may be used immediately or may be stored in the dry state. A gene specific PCR using the primers disclosed herein amplifies an HLA gene of interest, e.g., HLA-A, -B or –C or HLA-DRB1 or other HLA genes, in the purified DNA without allele bias. The exact allele type in any particular human DNA sample is determined by measuring the pattern of hybridization to the sequence-specific probes on the microarray. The pattern of hybridization determines the allele.

The following example(s) are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion.

EXAMPLE 1

Primers for amplification of class I and II HLA loci

Validation of algorithms for design of SNP specific primers

Forward primer 5'GCTCCCACTCCATGAGGTAT3' (SEQ ID NO: 1) and reverse primer 5'ATACCTCATGGAGTGGGAGC3' (SEQ ID NO: 2) was used to amplify an exon 2 PCR product for Class I HLA-B type to generate a specific product of 456 bp (15). The algorithms presented herein are used to ascertain the uniqueness of these primers to see if they primed only the exon 2 the Class I HLA-B locus. It was determined that the forward primer could bind to 11 other locations within chromosome 6 and also bind to one other location on chromosome X. The reverse primer for exon 2 of the Class I HLA-B locus was found to bind to five other locations on chromosome 6 and one location on chromosome 4 and 13. Another primer 5'ACCCGGTTTACCCGGTTTCATTTG3' (SEQ ID NO: 6) for the amplification of exon 3 of Class I HLA-B was found to bind in eight locations on

chromosome 6 other than intron 2, position 164-184, and to several other chromosomes (Table 1). Table 1 shows the number of times the primer or substrings of the primer sequences, shown underlined, were found in locations other than the correct one. The algorithm (19-20) clearly shows that the design of the primers is very crucial for performing high resolution HLA typing. The issue is complicated further by the possibility of contamination from other bacterial and viral genomes present in human bodily fluids.

TABLE 1

Primer No.	Seq Id	Sequence (5'-3')	Chromosome	Occurrence
Forward	1	gctcccactccatgaggtat	6	12
Exon 2	1,3	gctcccactccatgaggtat	X	1
Reverse	2	atacctcatggagtgggagc	6	6
Exon 2	2,4	aaatgaaaccgggtaaac	4	1
	2,5	aaatgaaaccgggtaaac	13	1
Forward	6	acccggtttacccggttcatttg	6	9
Exon 3	6,7	acccgtttacccggttcatttg	9	1
	6,8	acccgtttacccggttcatttg	9,3,11	2
	6,9	acccgtttacccggttcatttg	9,13	1
	6,10	acccgtttacccggttcatttg	1,3,11	1
	6,11	acccgtttacccggttcatttg	11,2,3	2
	6,12	acccgtttacccggttcatttg	18	1
	6,13	acccgtttacccggttcatttg	4,10	1

Primer design for amplification of HLA-B and HLA gene cluster

A simple 500 bp long amplicon spanning the entire region of interest for each HLA gene and a pair of approximately 250 bp long products are generated. This is suitable

for a simple PCR assay for HLA-A, B or DRB1. However, the entire HLA gene cluster would require 14 or 28 primer pairs designed by this method.

For each of the 14 HLA loci, i.e., ten class I HLA loci and 4 class II loci, the hypervariable region of interest spans approximately 500 bp. Therefore the primer design problem is to find a set of forward and reverse primers that are thermodynamically similar, but distinct relative to the human genome and other genomes which would contaminate the buccal DNA. To initiate the process, an approximately 100 bp region is identified at each end of the 14 approximately 500 bp HLA loci. A set of all possible 18-20 mer primers is readily obtained for these pair-wide 100 bp domains using the calculational tools described herein. This set of approximately 1400 forward and 1400 reverse primers is then filtered to obtain a subset of similar calculated thermodynamic stabilities. It is then additionally filtered to remove those with Watson-Crick complementarity with others in the set and to eliminate capacity for hairpin formation.

The remaining set is then additionally filtered relative to the remainder of the human genome set of 18-20 mers to eliminate the capacity for false priming. Since the desired PCR products are relatively small the greatest weight is given to pair-wise possible forward primer/reverse primer associations elsewhere which would produce PCR products smaller than 2000 bp. This is based upon the practical observation that spurious reactions producing products that are greater than 2000 bp are highly inefficient.

It is contemplated that the above described filtering process will produce several possible forward/reverse pairings for each of the 14 HLA loci. These primer pairs are tested individually in a 96-well thermal cycler environment and the products analyzed by electrophoresis to obtain that subset that functions optimally in an experimental context. This final level of primer filtering is done with both pure human DNA obtained from blood and also with human-non-human mixtures obtained from the buccal wash method

Table 2 is a representative list of primary and secondary primers effective to amplify HLA A, B and DRB loci and the identified exons.

Table 2

HLA Primers	SEQ ID	Label	Sequence
Primary Class I			
A-LOC-FP1	14		GCCTCTGYGGGGAGAAGCAA
A-LOC-RP1	15		GTCCCAATTGTCTCCCCTCCTT
B-LOC-FP1	16		GGGAGGAGMGAGGGGACCGCAG
B-LOC-RP2	17		TTCTCCATTCAASGGAGGGCGACA
B-LOC-RP1	18		GGAGGCCATCCCGGGCGATCTAT
B-LOC-RP3	19		GGAGGCCATCCCCGGCGACCTAT
Secondary Class I			
AX2-FP-1	20	biotin	AGCCGCGCCKGGAGGAGGGTCG
AX2-RP-1	21	biotin	GCCCGTCCGTGGGGGATGAG
AX3-FP-1	22	biotin	CAAAAATCCCCCCRGGTTGGTCGG
AX3-RP-1	23	biotin	GGCCCCTGGTACCCGTGCGCTG
AX3-FP2	24	biotin	GTTTCATTTTCAGTTTAGGCCA
AX3-RP-2	25	biotin	GTGCGCTGCAGCGTCTCCTTCC
AX3-RP-2	26	biotin	GTGCGCTGCAGCGTCTCCTTCC
BX2-FP-2	27	biotin	GAGCCGCGCCGGKAGGAGGGTC
BX2-RP-2	28	biotin	GGTCACTCACCGKCCTCGCTCT
BX3-FP-1	29	biotin	GGGGCCAGGGTCTCACA
BX3-RP1	30	biotin	CCCACTGCCCCTGGTACC
BX2-RP-3	31	biotin	CGGGCCGTMCGTGGGGGATGG
Primary Class II			
DRB-LOC-FP1a	32		CTTGGAGGTCTCCAGAACAGG
DRB-LOC-FP1b	33		CTTAGAGGTCTCCAGAACCGG
DRB-LOC-RP1a	34		GCCCCCAGCACCCACCTCCCTT
DRB-LOC-RP1b	35		GCCCCCTGTACCCCCCTCCCAC
DRB-LOC-RP1c	36		GCTCCGTGCACCCACCTCCCTT
DRB-LOC-RP1d	37		GCCGCCCGCACCCACCTCCCTT
Secondary Class II			
DRB1-X2-FP1a	38	biotin	CACAGCACGTTTCTTGGAGG
DRB1-X2-FP1b	39	biotin	TCCCCACAGCACGTTTCTTGA
DRB1-X2-FP1c	40	biotin	TCCCCACAGCACGTTTCTTGTG
DRB1-X2-FP1d	41	biotin	CCCCACAGCACGTTTCCTGTG
DRB1-X2-FP1e	42	biotin	CAGCACGTTTCTTGGAGCAGGT
DRB1-X2-FP1f	43	biotin	TCCCACAGCAGGTTTCCTGTG
DRB1-X2-FP1g	44	biotin	CCCACAGCACGTTTCTTGGAGT
DRB1-X2-RP1a	45	biotin	CACACACACACACACACTCAGATTC
DRB1-X2-RP1b	46	biotin	CACACACACAACCACACTCAGATTC
DRB1-X2-RP1c	47	biotin	CACACACACACACAGAGTCAGATTC

EXAMPLE 2

Primary and secondary amplification of Class I and II HLA loci

HLA-A and B loci and exons 2 and 3

In primary PCR reactions primer pairs A-LOC-FP1/A-LOC-RP1 for HLA-A locus and B-LOC-FP1/B-LOC-RP1 for HLA-B locus are used to generate amplified products of various UCLA standards, a positive control and a negative control. The PCR protocol is for a 50 μL volume in 96 well plate: one pre-PCR denaturing cycle 94°C for 4 min, 35 PCR cycles at 98°C for 1 min, 71°C for 1 min, 72°C for 1 min; hold cycle 72°C for 7 min (polymerase: Roche Fast Start Taq). Amplified products (10 μL samples) are run on a 2% agarose gel at 150 volts for 35 min (Lane 1: C1-034, Lane 2: C1-035, Lane 3: C1-036 and Lane 4: CCR1), a positive control (Lane 5: Roche DNA) and a negative control (Lane 6: H_2O); the last lane has weight standards. The gels show products of 980 bp for HLA-A (Fig. 1A) and 1007 bp for HLA-B (Fig. 1B).

In secondary PCR reactions primer pairs A-X2--FP1/A-X2-RP1 and A-X3--FP1/A-X3-RP1 for HLA-Ax2 and HLA-Ax3 exons are used to generate amplified products. The PCR protocol is for a 50 μL volume in 96 well plate: one pre-PCR denaturing cycle 94°C for 4 min, 35 PCR cycles at 98°C for 1 min, one of 58.3, 60.7, 63.3, 66.0, 68.6, or 71.0 °C for 30 sec, and hold cycle 72°C for 7 min (polymerase: Lucigen EconoTaq). Amplified products (10 μL samples) are run on a 2% agarose gel at 150 volts for 45 min (Lane 1: Aex2 or Aex3 negative control, Lane 2: 58.3°C, Lane 3: 60.7°C, Lane 4: 63.3°C, Lane 5: 66.0°C, Lane 6: 68.6°C, Lane 7: 71.0°C; Figs. 1C-1D).

In secondary PCR reactions primer pairs B-X2-FP1/B-X2-RP1 and B-X3-FP1/B-X3-RP1 for HLA-Bx2 and HLA-Bx3 exons are used to generate amplified products. The PCR protocol is for a 50 μL volume in 96 well plate: one pre-PCR denaturing cycle 94°C for 4 min, 35 PCR cycles at 98°C for 1 min, one of 60.7, 63.3, 66.0, 68.6, 71.0, or 73.0°C for 30 sec, and hold cycle 72°C for 7 min (polymerase: Roche Fast Start Taq). Amplified products (10 μL samples) are run on a 2% agarose gel at 150 volts for 45 min (Lane 1: Aex2 or Aex3 negative control, Lane 2: 60.7°C, Lane 3: 63.3°C, Lane 4: 66.0°C, Lane 5: 68.6°C, Lane 6: 71.0°C, Lane 7: 73.0°C; Figs. 1E-1F). The gel shows product around 1018 bp.

HLA-DRB1 locus and exons 2 and 3

In a primary PCR reaction primer pair DRB-LOC-FP1a/ DRB-LOC-RP1 for HLA-A locus and B-LOC-FP1/B-LOC-RP1 for HLA-B locus are used to generate amplified products. The PCR protocol is for a 25 μL volume in 96 well plate: one pre-PCR denaturing cycle 94°C for 4 min, 35 PCR cycles at 98°C for 1 min, one of 55.2, 56.5, 58.3, 60.7, 63.3, 66.0, 68.6, 71.0, 73.0, 74.4, or 75.2°C for 30 sec, and hold cycle 72°C for 7 min (polymerase: Roche Fast Start Taq). Amplified products (10 μL samples) are run on a 2% agarose gel at 150 volts for 35 min (Lane 1: DRB1 negative control, Lane 2: 55.2°C, Lane 3: 56.5°C, Lane 4: 58.3°C, Lane 5: 60.7°C, Lane 6: 63.3°C, Lane 7: 66.0°C, Lane 8: 68.6°C, Lane 9: 71.0°C, Lane 10: 73.0°C, Lane 11: 74.4°C, Lane 12: 75.2°C; Fig. 1G).

Combinations of primer pairs are used to amplify HLA-DRB1. The PCR protocol is for a 25 μL volume in 96 well plate: one pre-PCR denaturing cycle 94°C for 4 min, 35 PCR cycles at 98°C for 1 min, 74°C for 1 min, 72°C for 1 min; hold cycle 72°C for 7 min (polymerase: Roche Fast Start Taq). Amplified products (10 μL samples) are run on a 2% agarose gel at 150 volts for 35 min (Lane 1: DRB-LOC-FP1a/DRB-LOC-RP1a, Lane 2: DRB-LOC-FP1a/DRB-LOC-RP1b, Lane 3: DRB-LOC-FP1a/DRB-LOC-RP1c, Lane 4: DRB-LOC-FP1a/DRB-LOC-RP1d, Lane 5: DRB-LOC-FP1a/DRB-LOC-RP1a and Lane 6: DRB-LOC-FP1b/DRB-LOC-FP1a); the last lane has weight standards (Fig. 1H).

Secondary PCR reactions are run to amplify a HLA-DRB1 exon 2 genomic DNA template using primer pair DRB-x2-FP1g/DRB-x2-RP1a under conditions as described for Figure 1G. Amplified products (10 μL samples) are run on a 2% agarose gel at 150 volts for 35 min (Lane 1: DRB1ex2 negative control, Lane 2: 55.2°C, Lane 3: 56.5°C, Lane 4: 58.3°C, Lane 5: 60.7°C, Lane 6: 63.3°C, Lane 7: 66.0°C, Lane 8: 68.6°C, Lane 9: 71.0°C, Lane 10: 73.0°C, Lane 11: 74.4°C, Lane 12: 75.2°C; Fig. 1I). Combinations of primer pairs are used to amplify a HLA-DRB1 exon 2 genomic DNA template under conditions as described for Figure 1H. Amplified products (10 μL samples) are run on a 2% agarose gel at 150 volts for 35 min (Lane 1: DRB-x2-FP1g/ DRB-x2-RP1a, Lane 2: DRB-x2-FP1g/DRB-x2-RP1b, Lane 3: DRB-x2-FP1g/DRB-x2-RP1c, Lane 4: DRB-x2-FP1a/DRB-x2-RP1a, Lane 5: DRB-x2-FP1b/DRB-x2-RP1a, Lane 6: DRB-x2-FP1c/DRB-x2-FP1a, Lane 7: DRB-x2-

FP1d/DRB-x2-FP1a, Lane 8: DRB-x2-FP1e/DRB-x2-FP1a, Lane 9: DRB-x2-FP1f/ DRB-x2-FP1a, Lane 10: DRB-x2-FP1g/DRB-x2-FP1a) and the last lane has weight standards (Fig. 1J)

EXAMPLE 3

Probe design for microarray

A suitably designed microarray is used to test the hybridization parameters in all possible nearest neighbor contexts for a given mismatch. This allows for emulation of any sort of single nucleotide polymorphism. For a triple sequence there are 64 combinations which are studied by the central base pair and mispair in their nearest neighbor contexts. The results from all these combinations, enables prediction of the binding properties of the single nucleotide polymorphism for any HLA subtype. This is a much better estimate of the binding characteristics than from thermal melting profiles.

Briefly, for the HLA-B model the known allelic diversity of HLA-B defines 137 polymorphisms of clinical or epidemiological value, thus requiring a set of 137 allele specific probes to provide for a full hybridization analysis. Using the adsorptive approach to microarray manufacture described herein produces excellent single nucleotide specificity with probes in the 12-15 base range. A set of all 11-16 base long candidate probes which include the polymorphisms near the calculated center position, i.e., position 3 to N-2, to avoid end artifacts are generated by calculation. This probe set then is filtered to obtain a subset with identical calculated thermodynamic stabilities in binding cognate, perfectly matched target. For that standard, a calculated Tm of 55°C in 0.1 M NaCl is set.

From the primary filtered subset of probe candidates, a second filtering step relative to the remainder of the human genes is performed to eliminate those probe candidates which also appear in repetition sequences elsewhere, or at any other site in the HLA locus. Since hybridization is to be performed on PCR amplified targets, additional stringency will not be required in principle. However, for completeness the secondarily filtered set also will be analyzed for similarity to the entirety of the 1600 genome library of microbial and non-human vertebrate sequences described herein. These probes are used to manufacture microarrays.

Specific probe sequences

Tables 3-6 identify probe sequences for HLA A and HLA B loci, exons 2-3, respectively. oligo-T flanking segments are introduced at the 3' and 5' end of each probe, such that the flanking segments adsorb to the array surface but have little or no affinity for the solution state target nucleic acid. Probe recognition size is about 9-15 bases and oligo-T segments were added to both the 3' and the 5' side for a final overall length of 30 bases for all microarray probes.

TABLE 3

Probe	Sequence	SEQ ID
AX2-C002-SE-V1-1	TTTTTTTTTGCTCCCACTCCACTTTTTTTT	48
AX2-C002-SE-V2-1	TTTTTTTTGCTCTCACTCCATTTTTTTTTT	49
AX2-C002-AS-V1-1	TTTTTTTTTTGGAGTGGGAGCTCTTTTTTT	50
AX2-C002-AS-V2-1	TTTTTTTCTATGGAGTGAGAGCTCTTTTTT	51
AX2-C009-SE-V1-1	TTTTTTTTGTATTTCTTCACATCTTTTTTT	52
AX2-C009-SE-V3-1	TTTTTTTTGTATTTCTCCACATTTTTTTTT	53
AX2-C009-AS-V1-1	TTTTTTTTATGTGAAGAAATACTCTTTTTT	54
AX2-C017-AS-V3-1	TTTTTTTTTGTGGAGAAATACTCTTTTTTT	55
AX2-C017-SE-V1-1	TTTTTTTTTCCGCGGGGAGCTTTTTTTTTT	56
AX2-C017-SE-V2-1	TTTTTTTTTCAGTGGAGAGCCCTTTTTTTT	57
AX2-C017-AS-V1-1	TTTTTTTTTCTCCCCGCGGCTTTTTTTTTT	58
AX2-C017-AS-V2-1	TTTTTTTTGCTCTCCACTGCCTTTTTTTT	59
AX2-C024-SE-V1-1	TTTTTTTTTTTCATCGCCGTGTTTTTTTTT	60
AX2-C024-SE-V2-1	TTTTTTTTTCTTCATCGCAGTGTTTTTTTT	61
AX2-C024-AS-V1-1	TTTTTTTTCCACGGCGATGAATTTTTTTTT	62
AX2-C024-AS-V2-1	TTTTTTTTCCACTGCGATGAAGTTTTTTTT	63
AX2-C036-SE-V1-1	TTTTTTTTCTCGGTTCGACAGCTTTTTTTT	64
AX2-C036-SE-V2-1	TTTTTTTCTCGGTTTGACAGCGTTTTTTTT	65
AX2-C036-AS-V1-1	TTTTTTTTTCTGTCGAACCGCTTTTTTTTT	66
AX2-C036-AS-V2-1	TTTTTTTCTGCTGTCAAACCGCTTTTTTTT	67
AX2-C044-SE-V1-1	TTTTTTTTTCCAGAGGATGGAGTTTTTTTT	68
AX2-C044-SE-V2-1	TTTTTTTTTCCAGAAGATGGAGTTTTTTTT	69

AX2-C044-AS-V1-1	TTTTTTTTTCCATCCTCTGGCCTTTTTTTT	70
AX2-C044-AS-V2-1	TTTTTTTTTCCATCTTCTGGCCTTTTTTTT	71
AX2-C049-SE-V1-1	TTTTTTTTTTGGGCGCCGTGTTTTTTTTTT	72
AX2-C049-SE-V2-1	TTTTTTTTTCGGGCACCGTGTTTTTTTTTT	73
AX2-C049-AS-V1-1	TTTTTTTTCTCACGGCGCCCTTTTTTTTTT	74
AX2-C049-AS-V2-1	TTTTTTTTTCCCACGGTGCCCTTTTTTTTT	75
AX2-C056-SE-V1-1	TTTTTTTTTAGGGGCCGGAGCTTTTTTTTT	76
AX2-C056-SE-V2-1	TTTTTTTTTGAGGGTCCGGAGCTTTTTTTT	77
AX2-C056-AS-V1-1	TTTTTTTTTTCCGGCCCCTCTCTTTTTTTT	78
AX2-C056-AS-V2-1	TTTTTTTCTCTCCGGACCCTCTCTTTTTTT	79
AX2-C062-SE-V1-1	TTTTTTTCTGGACCAGGAGACTTTTTTTTT	80
AX2-C062-SE-V4-1	TTTTTTTCTGGACGAGGAGACTTTTTTTTT	81
AX2-C062-AS-V1-1	TTTTTTTTCGTCTCCTGGTCCTTTTTTTTT	82
AX2-C062-AS-V4-1	TTTTTTTTCGTCTCCTCGTCCTTTTTTTTT	83
AX2-C066-SE-V1-1	TTTTTTTTGGAATGTGAAGGCTTTTTTTTT	84
AX2-C066-SE-V2-1	TTTTTTTTGGAAAGTGAAGGCTTTTTTTTT	85
AX2-C066-AS-V1-1	TTTTTTTTCCTTCACATTCCGTTCTTTTTT	86
AX2-C066-AS-V2-1	TTTTTTTTCCTTCACTTTCCGTTCTTTTTT	87
AX2-C070-SE-V1-1	TTTTTTTTGCCCACTCACAGAACTTTTTTT	88
AX2-C070-SE-V2-1	TTTTTTTTGCCCAGTCACAGAACTTTTTTT	89
AX2-C070-AS-V1-1	TTTTTTTTTCTGTGAGTGGGCTCTTTTTTT	90
AX2-C070-AS-V2-1	TTTTTTTTTCTGTGACTGGGCTCTTTTTTT	91
AX2-C076-SE-V2-1	TTTTTTTTCCGAGAGAACCTGTTTTTTTTT	92
AX2-C076-SE-V3-1	TTTTTTTTTCGAGCGAACCTGTTTTTTTTT	93
AX2-C076-AS-V2-1	TTTTTTTTCAGGTTCTCTCGGCTTTTTTTT	94
AX2-C076-AS-V3-1	TTTTTTTTCAGGTTCGCTCGTCTTTTTTT	95
AX2-C081-SE-V1-1	TTTTTTTTTGACCCTGCGCGGTCTTTTTTT	96
AX2-C081-SE-V2-1	TTTTTTTTTGATCGCGCTCCGTCTTTTTTT	97
AX2-C081-AS-V1-1	TTTTTTTTTCCGCGCAGGGTCTTTTTTTTT	98
AX2-C081-AS-V2-1	TTTTTTTTTCGGAGCGCGATCTTTTTTTTT	99

TABLE 4

Probe	Sequence	SEQ ID
AX3-C095-SE-V1-1	TTTTTTCTTCACACCATCCAGACTTTTTTT	100
AX3-C095-SE-V2-1	TTTTTTTTCCACACCGTCCAGACTTTTTTT	101
AX3-C095-AS-V1-1	TTTTTTATTCTGGATGGTGTCATTTTTTTT	102
AX3-C095-AS-V2-1	TTTTTTTATTCTGGACGGTGTGTTTTTTTT	103
AX3-C097-SE-V1-1	TTTTTTTTTCCAGAGGATGTATTTTTTTTT	104
AX3-C097-SE-V3-1	TTTTTTTTTCCAGATGATGTATGTTTTTTT	105
AX3-C097-AS-V1-1	TTTTTTTTATACATCCTCTGGAATTTTTTT	106
AX3-C097-AS-V3-1	TTTTTTTCATACATCATCTGGAATTTTTTT	107
AX3-C105-SE-V1-1	TTTTTTTTTGGTCGGACGGGTTTTTTTTTT	108
AX3- C105-SE-V2-1	TTTTTTTTTGGCCGGACGGGTTTTTTTTTT	109
AX3- C105-AS-V1-1	TTTTTTTTTCCCGTCCGACCTTTTTTTTTT	110
AX3- C105-AS-V2-1	TTTTTTTTTCCCGTCCGGCCTTTTTTTTTT	111
AX3- C109-SE-V1-1	TTTTTTTTCGGCGCTTCCTCCTTTTTTTTT	112
AX3- C109-SE-V2-1	TTTTTTTTCTGCGCCTCCTCCTTTTTTTTT	113
AX3- C109-AS-V1-1	TTTTTTTTTGGAGGAAGCGCCTTTTTTTTT	114
AX3- C109-AS-V2-1	TTTTTTTTTTGGAGGAGGCGCTTTTTTTTT	115
AX3- C114-SE-V2-1	TTTTTTTTTGTACCGGCAGGATTTTTTTTT	116
AX3- C114-SE-V3-1	TTTTTTTTGTACCAGCAGGACTTTTTTTTT	117
AX3- C114-AS-V2-1	TTTTTTTTTTCCTGCCGGTACTTTTTTTTT	118
AX3- C114-AS-V3-1	TTTTTTTTGTCCTGCTGGTACTTTTTTTTT	119
AX3-C126-SE-V1-1	TTTTTTTTCCCTGAAAGAGGATTTTTTTTT	120
AX3-C126-SE-V2-1	TTTTTTTCTGCCCTGAACGAGTTTTTTTTT	121
AX3-C126-AS-V1-1	TTTTTTTTTCCTCTTTCAGGGTTTTTTTTT	122
AX3-C126-AS-V2-1	TTTTTTTCTCTCGTTCAGGGCTTTTTTTTT	123
AX3-C139-SE-V1-1	TTTTTTTTTATGGCGGCTCAGCATTTTTTT	124
AX3-C139-SE-V2-1	TTTTTTTTCATGGCAGCTCAGCATTTTTTT	125
AX3-C139-AS-V1-1	TTTTTTTTCCTGAGCCGCCATTCTTTTTTT	126
AX3-C139-AS-V2-1	TTTTTTTCCTGAGCTGCCATGCTTTTTTTT	127

AX3-C142-SE-V1-1	TTTTTTTTCTCAGATCACCAAGTTTTTTTT	128
AX3-C142-SE-V3-1	TTTTTTTTTTCAGACCACCAAGTTTTTTTT	129
AX3-C142-AS-V1-1	TTTTTTCTCTTGGTGATCTGAGTATTTTTT	130
AX3-C142-AS-V3-1	TTTTTTTCTCTTGGTGGTCTGATTTTTTTT	131
AX3-C145-SE-V1-1	TTTTTTTAAACCAAGCGCAAGTTTTTTTTT	132
AX3-C145-SE-V3-1	TTTTTTTAAACCCAGCGCAAGCTTTTTTTT	133
AX3-C145-AS-V1-1	TTTTTTTTTACTTGCGCTTGGTTTTTTTTT	134
AX3-C145-AS-V3-1	TTTTTTTTTCTTGCGCTGGGTTTTTTTTTT	135
AX3-C152-SE-V1-1	TTTTTTTTTGCCCATGTGGCGTTTTTTTTT	136
AX3-C152-SE-V2-1	TTTTTTTTTGCCCATGAGGCGTTTTTTTTT	137
AX3-C152-AS-V1-1	TTTTTTTCTCGCCACATGGGCTTTTTTTTT	138
AX3-C152-AS-V2-1	TTTTTTTTCTCGCCTCATGGGCTTTTTTTT	139
AX3-C152-SE-V1-2	TTTTTTTTTCCCATGTGGCGGTTTTTTTTT	140
AX3-C152-SE-V2-2	TTTTTTTTTCCCATGAGGCGGTTTTTTTTT	141
AX3-C152-AS-V1-2	TTTTTTTTCCCGCCACATGGGTTTTTTTTT	142
AX3-C152-AS-V2-2	TTTTTTTTCCCGCCTCATGGGTTTTTTTTT	143
AX3-C156-SE-V2-1	TTTTTTTTAGCAGCAGAGAGCTCTTTTTTT	144
AX3-C156-SE-V3-1	TTTTTTTTAGCAGTGGAGAGCTCTTTTTTT	145
AX3-C156-AS-V2-1	TTTTTTTTTGCTCTCTGCTGCTTTTTTTTT	146
AX3-C156-SE-V3-1	TTTTTTTTTGCTCTCCACTGCTTTTTTTTT	147
AX3-C163-SE-V1-1	TTTTTTTTTAGGGCACGTGCGCTTTTTTTT	148
AX3-C163-SE-V2-1	TTTTTTTTTAGGGCCGGTGCTCTTTTTTTT	149
AX3-C163-AS-V1-1	TTTTTTTTTCGCACGTGCCCTTTTTTTTTT	150
AX3-C163-AS-V2-1	TTTTTTTTTGCACCGGCCCTTTTTTTTTT	151
AX3-C166-SE-V1-1	TTTTTTTTTGTGGAGTGGCTCTTTTTTTTT	152
AX3-C166-SE-V2-1	TTTTTTTTTGTGGACGGGCTCTTTTTTTTT	153
AX3-C166-AS-V1-1	TTTTTTTTTGAGCCACTCCACTTTTTTTTT	154
AX3-C166-AS-V2-1	TTTTTTTTTGAGCCCGTCCACTTTTTTTTT	155

TABLE 5

Probe	Sequence	SEQ ID
BX2-C009-AS-V1-1	TTTTTTTTGGTGTAGAAATACTCTTTTTTT	156
BX2-C009-AS-V2-1	TTTTTTTTTGTGTGGAAATACTCTTTTTTT	157
BX2-C009-AS-V3-1	TTTTTTTTTGTGTCGAAATACTCTTTTTTT	158
BX2-C011-SE-V1-1	TTTTTTTATCACCGCCATGTCTTTTTTTTT	159
BX2-C011-SE-V2-1	TTTTTTTATCACCTCCGTGTCTTTTTTTTT	160
BX2-C011-SE-V3-1	TTTTTTTATCACCGCCGTGTCTTTTTTTTT	161
BX2-C011-AS-V1-1	TTTTTTTTTGACATGGCGGTGCTTTTTTTT	162
BX2-C011-AS-V2-1	TTTTTTTTTGACACGGAGGTGCTTTTTTTT	163
BX2-C011-AS-V3-1	TTTTTTTTTGACACGGCGGTGCTTTTTTTT	164
BX2-C024-AS-V1-1	TTTTTTTTTCACTGCGATGAAGTTTTTTTT	165
BX2-C024-AS-V2-1	TTTTTTTTCCACTGAGATGAAGTTTTTTTT	166
BX2-C024-AS-V3-1	TTTTTTTTTCACGGTGATGAAGTTTTTTTT	167
BX2-C024-AS-V4-1	TTTTTTTTCCACTGCAATGAAGTTTTTTTT	168
BX2-C031-SE-V1-1	TTTTTTTTCGACACCCAGTTCTCTTTTTTT	169
BX2-C031-SE-V2-1	TTTTTTTTCGACACGCTGTTCTCTTTTTTT	170
BX2-C031-SE-V3-1	TTTTTTTTCGACACGCAGTTCTCTTTTTTT	171
BX2-C031-SE-V4-1	TTTTTTTCTGACGGCACCCAGCCTTTTTTT	172
BX2-C031-AS-V1-1	TTTTTTTCTACGAACTGGGTGTTTTTTTTT	173
BX2-C031-AS-V2-1	TTTTTTTCTACGAACAGCGTGTTTTTTTTT	174
BX2-C031-AS-V3-1	TTTTTTTCTACGAACTGCGTGTTTTTTTTT	175
BX2-C031-AS-V4-1	TTTTTTTTTCTGGGTGCCGTCTTTTTTTTT	176
BX2-C031-AS-V1-2	TTTTTTTCTACGAACTGGGTGCTTTTTTTT	177
BX2-C031-AS-V2-2	TTTTTTCTACGAACAGCGTGTCTCTTTTTT	178
BX2-C031-AS-V3-2	TTTTTTCTACGAACTGCGTGTCTCTTTTTT	179
BX2-C031-AS-V1-3	TTTTTTTTTCGAACTGGGTGTTTTTTTTTT	180
BX2-C031-AS-V2-3	TTTTTTTTGAACTGCGTGTCGTTTTTTTTT	181
BX2-C035-SE-V1-1	TTTTTTTCTTTCGTGAGGTTCGTTTTTTTT	182
BX2-C035-SE-V2-1	TTTTTTTTTTTCGTGCGGTTCGTTTTTTTT	183

BX2-C035-SE-V2-2	TTTTTTTTTGTTCGTGCGGTTCTTTTTTTT	184
BX2-C041-SE-V1-1	TTTTTTTTTACGCCGCGAGTCTTTTTTTTT	185
BX2-C041-SE-V2-1	TTTTTTTTTACGCCACGAGTCTTTTTTTTT	186
BX2-C041-SE-V1-2	TTTTTTTTTTCGCCGCGAGTCTTTTTTTTT	187
BX2-C045-AS-V1-2	TTTTTTTTTGCTCCTCTCTCGGTTTTTTTT	188
BX2-C045-AS-V2-2	TTTTTTTTTGCTCCGTCCTCGGTTTTTTTT	189
BX2-C045-AS-V3-2	TTTTTTTTTGCTCCTTCCTCGGTTTTTTTT	190
BX2-C045-AS-V4-2	TTTTTTTTTGCGCCATCCTCGGTTTTTTTT	191
BX2-C045-AS-V5-2	TTTTTTTTTGCTCCCCTCTCGGTTTTTTTT	192
BX2-C050-SE-V1-1	TTTTTTTTGCGCCATGGATAGTTTTTTTTT	193
BX2-C050-SE-V2-1	TTTTTTTTTCGCCGTGGATATTTTTTTTTT	194
BX2-C050-SE-V3-1	TTTTTTTTTTGCCGTGGGTGTTTTTTTTTT	195
BX2-C050-SE-V3-2	TTTTTTTTTCCGTGGGTGGATTTTTTTTTT	196
BX2-C050-AS-V2-2	TTTTTTCTCTCTATCCACGGCGCTTTTTTT	197
BX2-C054-AS-V1-2	TTTTTTTTCCTCCTGCTCCACCTTTTTTTT	198
BX2-C054-AS-V2-2	TTTTTTTTCCTCCTGCTCTATCTTTTTTTT	199
BX2-C054-AS-V3-2	TTTTTTTTCCCTCTTGCTCTATCTTTTTTT	200
BX2-C058-SE-V1-1	TTTTTTTTTCGGAGTATTGGGATTTTTTTT	201
BX2-C058-SE-V2-1	TTTTTTTTCCGGAATATTGGGATTTTTTTT	202
BX2-C063-AS-V1-2	TTTTTTTCCCTGTGTGTTCCGTCTTTTTTT	203
BX2-C063-AS-V2-2	TTTTTTTCCCTGTGTCTCCCGTCTTTTTTT	204
BX2-C063-AS-V3-2	TTTTTTTCCCCGTGTCTCCCGTCTTTTTTT	205
BX2-C063-AS-V4-2	TTTTTTTCCCCGTGTCTCCCCTCTTTTTTT	206
BX2-C067-SE-V1-1	TTTTTTTTCAGATCTCCAAGACTTTTTTTT	207
BX2-C067-SE-V2-1	TTTTTTTTCAGATCTTCAAGACTTTTTTTT	208
BX2-C067-SE-V3-1	TTTTTTTTCAGATCTACAAGGCTTTTTTTT	209
BX2-C067-SE-V4-1	TTTTTTTTCAGATCTGCAAGACTTTTTTTT	210
BX2-C067-SE-V5-1	TTTTTTTTAGATCTGCAAGGCTTTTTTTTT	211
BX2-C067-SE-V6-1	TTTTTTTTCGGAACATGAAGGTTTTTTTTT	212
BX2-C067-SE-V7-1	TTTTTTTTCAGAAGTACAAGCGCTTTTTTT	213
BX2-C067-SE-V8-1	TTTTTTTTCAGATCTAGAAGACTTTTTTTT	214

BX2-C067-SE-V6-2	TTTTTTTTACGGAACATGAAGTTTTTTTTT	215
BX2-C067-SE-V7-2	TTTTTTTTCAGAAGTACAAGCGTTTTTTTT	216
BX2-C067-SE-V8-2	TTTTTTTTAGATCTACAAGACCTTTTTTTT	217
BX2-C069-SE-V1-1	TTTTTTTTTAAGACCAACACATTTTTTTTT	218
BX2-C069-SE-V2-1	TTTTTTTTTAAGGCCCAGGCACTTTTTTTT	209
BX2-C069-SE-V3-1	TTTTTTTTTAAGGCCAAGGCACTTTTTTTT	220
BX2-C069-SE-V4-1	TTTTTTTTTAAGGCCTCCGCGCTTTTTTTT	221
BX2-C069-SE-V5-1	TTTTTTTTTAAGCGCCAGGCACTTTTTTTT	222
BX2-C069-SE-V1-2	TTTTTTTTTAGACCAACACACTTTTTTTTT	223
BX2-C069-SE-V2-2	TTTTTTTTAAGGCCCAGGCACATTTTTTTT	224
BX2-C069-SE-V3-2	TTTTTTTTAAGGCCAAGGCACATTTTTTTT	225
BX2-C069-SE-V4-2	TTTTTTTTGAAGGCCTCCGCGCTTTTTTTT	226
BX2-C069-SE-V5-2	TTTTTTTTCAAGCGCCAGGCATTTTTTTTT	227
BX2-C069-SE-V4-3	TTTTTTTCGAAGGCCTCCGCGCTTTTTTTT	228
BX2-C074-SE-V1-1	TTTTTTTTCAGACTTACCGAGCTTTTTTTT	229
BX2-C074-SE-V2-1	TTTTTTTTACAGACTGACCGATCTTTTTTT	230
BX2-C077-SE-V1-1	TTTTTTTTGCAGGCTCTCTCGTCTTTTTTT	231
BX2-C077-SE-V2-1	TTTTTTTTGCAGGTTCTCTCGTCTTTTTTT	232
BX2-C077-SE-V1-2	TTTTTTTTGCAGGTCCTCTCGTCTTTTTTT	233
BX2-C077-SE-V2-2	TTTTTTTTGCAGGCTCACTCGTCTTTTTTT	234
BX2-C077-SE-V3-2	TTTTTTTTGCAGGCCCACTCGTCTTTTTTT	235
BX2-C081-SE-V1-2	TTTTTTTTTGGAACCTGCGCGTTTTTTTTT	236
BX2-C081-SE-V2-2	TTTTTTCTCGGATCGCGCTCCGTCTTTTTT	237
BX2-C081-SE-V3-2	TTTTTTTTGCACCGCGCTCCGTCTTTTTTT	238
BX2-C081-SE-V4-2	TTTTTTCTCGGACCCTGCTCCGTCTTTTTT	239

TABLE 6

Probe	Sequence	SEQ ID
BX3-C094-SE-V1-1	TTTTTTTTCCTCACACCCTCCTTTTTTTTT	240
BX3-C094-SE-V2-1	TTTTTTTTTCTCACATCATCCATTTTTTTT	241
BX3-C094-AS-V1-1	TTTTTTTTCGGAGGGTGTGAGTTTTTTTTT	242
BX3-C094-AS-V2-1	TTTTTTTTTGGATGATGTGAGATTTTTTTT	243
BX3-97/99-SE-V1-1	TTTTTTTTAGAGGATGTACGGTCTTTTTTT	244
BX3-97/99-SE-V2-1	TTTTTTTTAGAGCATGTACGGTCTTTTTTT	245
BX3-97/99-AS-V1-1	TTTTTTTTTCCGTACATCCTCTTTTTTTTT	246
BX3-97/99-AS-V2-1	TTTTTTTTTCCGTACATGCTCTTTTTTTTT	247
BX3-C103-SE-V1-1	TTTTTTTTTACGTGGGGCCGTTTTTTTTTT	248
BX3-C103-SE-V3-1	TTTTTTTTTACCTGGGGCCGTTTTTTTTTT	249
BX3-C103-AS-V1-1	TTTTTTTTCTCGGCCCCACGTTTTTTTTTT	250
BX3-C103-AS-V3-1	TTTTTTTTCTCGGCCCCAGGTTTTTTTTTT	251
BX3-C114-SE-V1-1	TTTTTTTTGGGCATAACCAGTCTTTTTTTT	252
BX3-C114-SE-V2-1	TTTTTTTTTGGGCATGACCAGCTTTTTTTT	253
BX3-C114-AS-V1-1	TTTTTTTCCACTGGTTATGCCCTTTTTTTT	254
BX3-C114-AS-V2-1	TTTTTTTCTCTGGTCATGCCCTTTTTTTTT	255
BX3-C116-SE-V1-1	TTTTTTTCTACCAGTACGCCTATTTTTTTT	256
BX3-C116-SE-V2-1	TTTTTTTCTACCAGTCCGCCTATTTTTTTT	257
BX3-C116-AS-V1-1	TTTTTTTTTAGGCGTACTGGTATTTTTTTT	258
BX3-C116-AS-V2-1	TTTTTTTTTTAGGCGGACTGGTTTTTTTTT	259
BX3-C121-SE-V1-1	TTTTTTTTGGCAAGGATTACATTTTTTTTT	260
BX3-C121-SE-V2-1	TTTTTTTTGGCAAAGATTACATCTTTTTTT	261
BX3-C121-AS-V1-1	TTTTTTTATGTAATCCTTGCCTCTTTTTTT	262
BX3-C121-AS-V2-1	TTTTTTTGATGTAATCTTTGCCTCTTTTTT	263
BX3-C131-SE-V1-1	TTTTTTTTTGACCTGAGCTCCCTTTTTTTT	264
BX3-C131-SE-V2-1	TTTTTTTTTACCTGCGCTCCTTTTTTTTTT	265
BX3-C131-AS-V1-1	TTTTTTTTTGGAGCTCAGGTCTCTTTTTTT	266
BX3-C131-AS-V2-1	TTTTTTTTTAGGAGCGCAGGTTTTTTTTTT	267

BX3-C135-SE-V1-1	TTTTTTTTTACCGCGGCGGATTTTTTTTTT	268
BX3-C135-SE-V2-1	TTTTTTTTTACCGCCGCGGATTTTTTTTTT	269
BX3-C135-AS-V1-1	TTTTTTTTCTTCCGCCGCGGTTTTTTTTTT	270
BX3-C135-AS-V2-1	TTTTTTTTCTTCCGCGGCGGTTTTTTTTTT	271
BX3-C143-SE-V1-1	TTTTTTTTCTCAGATCACCCATTTTTTTTT	272
BX3-C143-SE-V2-1	TTTTTTTTCTCAGATCTCCCATTTTTTTTT	273
BX3-C143-AS-V1-1	TTTTTTTTTTGGGTGATCTGAGTTTTTTTT	274
BX3-C143-AS-V2-1	TTTTTTTTTTGGGAGATCTGAGTTTTTTTT	275
BX3-C145-SE-V1-1	TTTTTTTTCCCCAGCGCAAGTCTTTTTTTT	276
BX3-C145-SE-V2-1	TTTTTTTTCCCCAGCTCAAGTGTTTTTTTT	277
BX3-C145-AS-V1-1	TTTTTTTTTACTTGCGCTGGGCTTTTTTTT	278
BX3-C145-AS-V2-1	TTTTTTTTCACTTGAGCTGGGCTTTTTTTT	279
BX3-C152-SE-V1-1	TTTTTTTTTTCCCGTGTGGCGTTTTTTTTT	280
BX3-C152-SE-V2-1	TTTTTTTTTTCCCGTGAGGCGTTTTTTTTT	281
BX3-C152-AS-V1-1	TTTTTTTTCTCGCCACACGGGTTTTTTTTT	282
BX3-C152-AS-V2-1	TTTTTTTTCTCGCCTCACGGGTTTTTTTTT	283
BX3-C156-SE-V1-1	TTTTTTTTAGCAGCTGAGAGCTCTTTTTTT	284
BX3-C156-SE-V3-1	TTTTTTTTTAGCAGCGGAGAGTTTTTTTTT	285
BX3-C156-AS-V1-1	TTTTTTTTTCTCTCAGCTGCTCTTTTTTTT	286
BX3-C156-AS-V3-1	TTTTTTTTTCTCTCCGCTGCTTTTTTTTTT	287
BX3-C163-SE-V1-1	TTTTTTTTTGGCCTGTGCGTGTTTTTTTTT	288
BX3-C163-SE-V2-1	TTTTTTTTTGGCGAGTGCGTGTTTTTTTTT	289
BX3-C163-AS-V1-1	TTTTTTTCTCACGCACAGGCCTCTTTTTTT	290
BX3-C163-AS-V2-1	TTTTTTTCTCACGCACTCGCCTCTTTTTTT	291

EXAMPLE 4

Microarray manufacture

Microarrays are manufactured using methods described in Belosludtsev *et al.* (17). Briefly, substrates used are silica slides cleaned in an ultrasonic bath with detergent (2 min) followed by washing with distilled water and methanol and drying (30 min at 40°C).

Slides are silanized with 3-aminopropyltrimethoxysilane or 3-glycidoxy-propyltrimethoxysilane in vapor phase in equilibrium with a 50% silane/p-xylene solution in a vacuum oven at 25 in. Hg overnight at 70-80°C with cleaned slides assembled in a rack.

Commercially available oligonucleotides (Midland Certified Reagent Co. Midland, TX) are deposited as 1 nl solutions in distilled water at 5 μM for aminosilanized slides upon the silanized surface using a microarrayer. After "printing" oligonucleotides, the slides are dried for 15 min at 40°C or overnight at room temperature and capped with acetic anhydride in vapor phase by placing a Petri dish with a solution of 3 ml of acetic anhydride in 3 ml of dimethyl formamide (DMF) in a vacuum oven at 22 in Hg for 1 h at 50°C. The acetylated amino-derivatized slides are capped with succinic anhydride by dipping slides in a tank with 0.5 M succinic anhydride in DMF at room temperature for 1 h. Slides are cleaned by washing in acetone (3x), in distilled water (2x), and again in acetone (2x). As a quality control, 30 ml of fluorescent-labeled oligonucleotide in hybridization buffer is deposited onto the slide surface. If no background is observed after 15 min, slides would be considered ready for hybridization experiments.

A comparison of adsorptive vs covalent strategies for oligonucleotide attachment to a planar glass substrate was made. Covalent attachment was obtained by reaction of a 5'-amino-modified oligonucleotide with an epoxysilanized surface, a standard method in the art for covalent attachment to surfaces, which yields a terminal secondary amine linkage. Comparison of experimental and calculated density gives evidence that a densely packed monolayer of oligonucleotides was formed during the adsorption process and length dependency studies demonstrate that a densely packed probe film can be formed for probes as long as 36 bases. In further determinations of the stability and structure of the immobilized target, it was found that for short oligonucleotides up to 36 mers, a ribbon form of single stranded DNA is formed on the surface (23). The hybridization selectivity and specificity are similar to a fully upright single stranded DNA molecule covalently attached to the surface.

EXAMPLE 5

Validation of sample collection and extraction

The "mouthwash" method (21) can be used to collect samples. This technique of sample collection is ideally suited for the collection of buccal-derived DNA for large-scale population studies and for collecting samples from geographically dispersed large-scale population studies. The "mouthwash" method does not require any medical supervision. The technique has been shown to preserve the integrity of the sample for up to 1 week as compared to freshly prepared DNA (Figs. 2A-2B), even on exposure to a variety of temperature conditions.

Briefly, buccal cells are collected by vigorous mouth washing for about 45 seconds with a mouthwash liquid or any other biocompatible liquid followed by spitting into a 10 ml jar. The jar is sealed and mailed. On arrival, cells are pelleted and the cell pellet is dissolved at 25°C in 100 µl of a solution of Tris-EDTA and 1% SDS for 1 hour with intermittent vortexing. This suspension is applied directly to GenVault elements (13 µl each). Subsequent to drying, DNA is isolated by two saline washes, at which time it remains bound to the porous element. DNA is then released from the element by a single 5 minute wash at 25°C in GenVault release buffer. DNA thus released, about 1 µg/element, can be used for PCR without additional purification.

Buccal cells can also be collected using a Fitzco Dacron cheek swab. Briefly, a swab is collected either from the right or left cheek and the swab tip is placed in a spin basket within a microfuge tube, air dried overnight and then capped for storage. The tip when required is rehydrated by direct addition of 200 ml SRB and then heated for 4 hours at 55°C. Sample is then harvested from the spin basket and centrifuged for 1minute at 10,000G. At least 90% fluidic recovery is obtained. The resulting material can either be processed by Argylla PrepParticles or by batchwise clean-up with a Qiagen column, per manufacturers recommendations. Table 7 compares DNA yields from cheek swabs extracted with the FBI's stain extraction buffer and then using Argylla (A) and Qiagen (Q) for each of 5 volunteers.

Table 7

Volunteer	DNA conc (ng/μl) (x 45 μl eluate) =	DNA (ng)
BI (A)	15.4	693
BI (Q)	41.3	1859
RE (A)	59.2	2664
RE (Q)	50	2250
MH (A)	67.9	3056
MH (Q)	44.1	1985
DI (A)	14.9	671
DI (Q)	13.2	594
JU (A)	29.6	1332
JU (Q)	18.5	833

Figure 3A demonstrates that the DNA extracted from two of the volunteers by the cheek swap method provides an excellent PCR product for HLA-B. UCLA standards are compared to the PCR products derived from buccal DNA samples. As seen in Figure 3B, HLA-B specific PCR reactions from buccal swab collection are quantitatively similar to those obtained from the pure UCLA reference standards.

EXAMPLE 6

Image analysis and pattern recognition

Digitally captured microarray images are analyzed using the algorithms in the ImageAnalyzer software. Briefly, the edge detection methods are employed to substract the background (Fig. 4A) from the images, to produce a clean pattern at a pre-defined threshold above the background (Fig. 4B). The pattern recognition and the generation of a 2-D bar code for the assignment of the HLA type is based on the hybridized spot patterns and their spatial relationships to each of the HLA type. Each of the alleles of the HLA type has a set pattern of spots which would be used as a barcode.

EXAMPLE 7

K-ras model: genotyping and detection of SNPs on adsorptive microarrays

Amplicon generation and oligonucleotide probes

The microarrays used in HLA-typing have been used successfully in identifying and assigning genotypes, differing from the wild type by a single nucleotide polymorphism, of K-ras locus.The 152-bp K-ras amplicon was generated by the polymerase chain reaction. Wild-type amplicaon (K-ras 1) was obtained by amplification of a commercial genomic DNA source (Sigma). K-ras 2 and K-ras 7 mutants were obtained by amplification of human genomic DNA from cell lines A549 and SW 480, respectively. The PCR protocol was the following: one pre-PCR cycle at 94° C for 12 min, 60° C for 1 min and 72° C for 1 min; 35 PCR cycles at 95° C for 1 min, 57° C for 1 min, 72° C for 1 min; hold cycle at 72° C for 7 min, 4° C hold. PCR primers for k-ras amplicons were labeled with digoxigenin at their 5' ends and had the following sequences: 5'-DIG-ACTGAATATAAACTTGTGGTAGTTGGACCT-3' (SEQ ID NO: 292) and 5'-DIG-TCAAAGAATGGTCCTGCACC-3' (SEQ ID NO: 293). K-ras amplicons had different point mutations in codon 12. Specific oligonucleotides were designed to serve as microarray capture probes as shown in Table 8. The underlined nucleotide corresponds to the point mutation.

Table 8

SEQ ID	Sense	SEQ ID	Antisense
294	GACCTGGTGGCG	301	CGCCACCAGGTC
295	GACCTAGTGGCG	302	CGCCACTAGGTC
296	GACCTTGTGGCG	303	CGCCACTAGGTC
297	GACCTCGTGGCG	304	CGCCACGAGGTC
298	GACCTGATGGCG	305	CGCCATCAGGTC
299	GACCTGCTGGCG	306	CGCCACGAGGTC
300	GACCTGTTGGCG	307	CGCCACAAGGTC

Hybridization and pattern detection

A prehybridization solution, containing 150 mM sodium citrate, with respect to sodium ion concentration, 5X Denhardt's solution, pH 8.0 was applied to the array for at least 10 min. The solution was vacuumed off and hybridization solution (1 nM amplicon, 0.1 mM chaperone, 150 mM sodium citrate with respect to sodium, 5X Denhardt's solution, pH 8.0) was applied to the array. In these studies, only amplicons complementary to capture probes K-ras 1, K-ras 2, K-ras 7 were used. After 2 h of hybridization, the array was washed two times in 100 mM sodium citrate with respect to sodium, 10 min each, followed by a brief rins in 13 SSC. The digoxigenin-labeled amplicon was detected using anti-digoxigenin antibody linked to alkaline phosphatase (Boehringer Mannheim) at 1:1000 dilution in the blocking buffer from the ELF-97 mRNA In Situ Hybridization Kit (Molecular Probes), followed by washing in buffer A from the same kit and by application of ELF as described in the kit, which is a substrate for alkaline phosphatase. After cleavage by alkaline phosphatase, ELF molecules precipitate and become fluorescent under UV excitation. The fluorescence intensities were detected with an Alpha Imager 2000 apparatus and processed using Sigma Plot 3.0 software (Figs. 5A-5D).

EXAMPLE 8

SNP detection in HLA-B

HLA-B validation model

UCLA has assembled a library of 75 highly characterized DNA samples with known HLA type, which is used worldwide as reference standards. This reference set was obtained and the HLA-B hypervariable region was resequenced to obtain a higher resolution understanding of sequence variation in the reference set. As seen in Figures 6A-6B, the UCLA-derived data are quite accurate, yielding only 4-5 discrepancies with respect to one-pass re-sequencing.

HLA-B target preparation

A 500-bp fragment from exon 2 of HLA-B is obtained by the polymerase chain reaction (PCR) using the primers designed above from test case purified human genomic DNA samples. The following PCR protocol is used to generate the amplified 500 bp

fragment: one pre-PCR cycle 94°C for 12 min, 60°C for 1 min, and 72°C for 1 min, 35 PCR cycles 95°C for 1 min, 57°C for 1 min, 72°C for 1 min; hold cycle 72°C for 7 min, 4°C hold. PCR primers for HLA-B amplicons are modified to contain a T7 polymerase recognition sequence 5' ATGTAATACGACTCACTATAG 3' (SEQ ID NO: 317).

The double-stranded PCR products are isolated by micro-column purification, then in vitro transcribed in the presence of biotin labeled-ribonucleotides using the HighYield RNA Transcript Labeling Kit (Enzo Labs, Farmingdale, NY). Briefly, the in vitro transcription reaction to generate a single stranded cRNA was done in a 20 microliter reaction volume containing a mix of 75 mM NTPs with rUTP fraction containing 25% biotinylated-rUTP, 10x reaction buffer, T7 Polymerase enzyme (Ambion, Austin, TX). The reaction mix is incubated at 37°C for four hours. In vitro transcribed biotin labeled cRNA is purified using Qiagen RNeasy kit (Qiagen, Valencia, CA), quantified, and is fragmented at 94°C for 35 min in the presence of 1x fragmentation buffer (40 mM Tris-acetate, pH 8.0, 100 mM Kac, 30 mM MgAc). The quality of the RNA is checked before and after fragmentation using formaldehyde agarose gels.

Hybridization and detection

Prehybridization solution, containing 150 mM sodium citrate, with respect to sodium ion concentration, 5X Denhardt's solution, pH 8.0, is applied to the array for at least 10 min. The prehybridization solution then is vacuumed off and a hybridization solution containing the single stranded labeled cRNA targets in 150 mM sodium citrate, with respect to sodium, and 5X Denhardt's solution, pH 8.0 was applied to the array for hybridization to the HLA microarray. After 2 h of hybridization, the array is washed two times in 100 mM sodium citrate, with respect to sodium, 10 min each, followed by a brief rinse in 1X SSC. Streptavidin linked phycoerythrin is bound to the biotin on the cRNA target (22). The fluorescence intensities are detected with a CCD-based microarray imager (Array Worx, API, Issaquah, WA).

HLA-B chip

Figures 7A-7F show hybridization data for codon 9 of HLA-B. In this array, 9-12 base long probes, lacking the oligo-T flanking sequence, are on the upper right and the

corresponding 30 mer oligo-T flanked derivatives are on the left of the microarray. Hybridization was performed with a Cy-3 labeled, 281bp nested PCR product of known sequence variation obtained from the UCLA HLA reference DNA library (samples 72, 21, 27, 57). Also, shown in Figure 7F-G are microarray hybridization data to 281bp products obtained from volunteers (MH, BI). As seen, keeping the probe recognition sequence constant, addition of the oligo-T flank gives rise to a 10-fold increase in hybridization signal relative to the short probe homologues.

In Figure 7G, spot intensity was manually quantified within the arrays for codon 9. Data in Figure 7G is presented as six clusters. The first four clusters correspond to hybridization data from UCLA reference samples of known allelotype at codon 9. Codon 9 was chosen because it is triallelic and thus three hybridization probes are required to interrogate known allelic variation. Within each cluster, two sets of probe type were tested: a "long probe" and a "short probe" where the sequence specific sequence at the center had been decreased by one base, in order to determine if specificity could be enhanced. Of importance, it can be seen that for all four of the UCLA reference samples, measured specificity is nearly perfect at codon 9 (Figure 7G). Those "spots" which should hybridize, as predicted from the UCLA standards are clearly detected. Specificity with respect to single nucleotide resolution is seen to be in excess of 10-fold and in some instances as much as 50-fold among these probes, which allows for unambiguous, hands-free analysis.

Nearly identical specificity factors are seen for Codon 50, Figure 7H. In Figure 7E-7F the microarray data derived from buccal swab DNA (MH, BI). These samples are buccal DNA obtained from volunteers, so the HLA type is not known. Thus, the values listed beneath the graph are the "HLA calls" for these two samples, rather than external validation: MH being an apparent V2/V3 heterozygote and BI being a V2/V2 homozygote at Codon 9. The bar graphs reveal that, as was the case for the 4 UCLA standards in Figures 7A-7D, clearly defined microarray analysis of buccal DNA was obtained from a standardized 5ng HLA-B specific PCR reaction. Given that the DNA yield per buccal swab is 500-2000ng the data of Figure 7G-7H demonstrates that each buccal swab collects at least 100 PCR equivalents of DNA. The data also demonstrates that extremely clear HLA calls can be obtained by simple inspection of the data. Because the probes are not chemically modified,

and because the oligo-T flanking sequence is a physical "filler", probe synthesis and purification remain very low cost for these microarrays.

EXAMPLE 9

Automated microarray signal analysis

Before intensities can be extracted from a raw image (as in Fig. 7A), the choice of the appropriate gridding technique must be made (Fig. 8). The grid's structural information is used to layout estimated spot positions. Since microarray spots are symmetric around their centers, an optimization routine can be executed to search for the optimal spot center within the spot's neighboring region. Starting with a raw microarray image, the gridding process includes raw gridding estimation, localized spot center adjustments and structural spot center adjustments. After gridding is complete, it is required to determine how many pixels can be considered to be a part of the spot. For a radius R (measured in pixels), a square of size 2R + 1 can easily be constructed in which the spot will be contained. To determine if a pixel is inside the spot the distance D between the given pixel and the center pixel of the square is calculated and seen if it is smaller or equal to the radius R, as illustrated in Figures 7A-7F. By making use of efficient data structures to store information for each pixel, the method detailed above can handle microarray images with thousands of spots with modest memory and computational time consumption.

A typical spot in Figure 7A has radius 15 (pixels), yielding 709 pixels considered to be in the spot. Having defined the boundary of each spot explicitly, the intensities of pixels contained in a certain spot were converted into signal of the spot. The metric exploited takes the arithmetic mean of n most intensive pixels of a spot after the s most intensive pixels are removed from consideration. By doing so the errors involving spikes in intensity values, noise, dust and other extraneous factors are removed. The choice of s and n should reflect the characteristics of the image.

The probes were tested on a series of 12 UCLA samples with the same enhanced 30mer probe set. The outcome of each test is an image similar to the one illustrated in Figure 7A. To make a statistically sound statement that this technique is successful at performing SNP calling, signals of the same probe from different images (samples) were combined. This necessitated normalization of the signals across images. Given the nature of

the experiment, images in HLA typing are expected to mostly contain spots with intensities at the two extremes (If a probe is present in the sample, the corresponding spot will have extremely high intensity. On the other hand, if the probe is absent from the sample, the corresponding spots will have extremely low intensity). The number of spots at each extreme is probe- and sample- specific. A sophisticated normalization scheme to accommodate the characteristics of the images was selected for HLA typing. It takes the arithmetic mean of the markers spots (the 6 bright spots on the left and bottom of Figure 7A, typically they have intensities of 255 in a 8-bit per pixel format image) as representative of spots with extremely high intensities. The arithmetic mean intensity of the background pixels was chosen as representative of spots with extremely low intensities. The range of an image was then defined as the difference between the arithmetic mean of the marker spots and the background. The images were normalized such that every image had the same range after normalization.

The normalized data of each probe were then divided into two sets, the present set and the absent set. The present set of a probe contains signals from images (samples) where the probe is present. On the other hand, the absent set contains signals from images (samples) where the probe is not present. The Mann-Whitney U tests were performed between the present set and the absent set for each probe. The resulting two-tailed p-values were in the range of 2% to 5×10^{-7}%, which indicated that the difference in signals of a probe being present in the sample and absent in the sample were statistically significant at the 2% level. These analyses are presented in Table 9. The U statistic of the test is 44, which is the highest for set size 22 and 2, i.e., 2.16×10^{-2} is the smallest p-value for set size 22 and 2.

TABLE 9

Probe		Set Size	p-value	Significance
Codon 50 (A)	Present set	28	4.76×10^{-9}	highly
	Absent set	20		highly
Codon 50 (G)	Present set	24	1.81×10^{-9}	highly
	Absent set	24		highly
Codon 09 (C)	Present set	8	8.88×10^{-5}	highly
	Absent set	16		highly
Codon 09 (G)	Present set	2	2.16×10^{-2}	significant
	Absent set	22		significant
Codon 09 (T)	Present set	20	1.95×10^{-3}	highly
	Absent set	4		highly

After the microarray images are converted into SNP calls making use of the Image Analyzer and the data analysis method, the SNP calls are converted to allele calls in an automatic fashion. Accurate and robust allele calls are enabled by the following properties of the final integrated set of probes on the HLA chip: (1) each allele in consideration must be identifiable by at least a certain number of probes; and (2) subsets that identify any two alleles must be different by at least a certain number of probes.

The allele calls are performed in a decision-tree-based approach. For example, let $( ) k p p P ,... 1 =$ be the result of the SNP calls, where $i p , k i ,..., 1 =$, denotes the presence/absence of an individual probe. Let $\{ \} n a a A ,... 1 =$ be the set all of alleles in consideration, where $j a , n j ,..., 1 =$, denotes an individual allele. Let $i A, k i ,..., 1 =$, denote the set of alleles that can be identified by probe i, i.e., at least one allele in $i A$ is expected to be present if $i p$ = Present and vice versa. The allele-calling decision tree is illustrated in Figure 9. Each node in the tree denotes a subset of A members of which are likely to be present at the current stage of decision-making. The root of the tree is A. Each leaf is a final set of alleles that are determined to be present in the sample. Given the SNP calls of a sample, the set of alleles present in the sample is determined by traversing the decision tree from the

root node to one of the leaves based on the presence/absence of each probe. The allele calling process is now a simple tree traversal, which has time complexity linear to the number of probes on the "HLA chip". Thus, the SNP calls can be converted into allele calls on a standard desktop PC (with a 1GHz CPU and 1GB RAM) in a matter of seconds. It is possible that the final set of alleles is empty, which implies a hybridization pattern that is unique from all possible patterns given the set of alleles in consideration has been encountered. In this case, the sample is suspected to contain new allele(s) and should be subjected to additional experiments or even sequencing.

EXAMPLE 9

Knowledge-based literature screening tool to determine HLA alleles of scientific interest

Although more than 10,000 HLA alleles are known, to develop HLA chips, which offer best clinical advantage, it is important to focus on that subset of the total alleles associated with the greatest body of accumulated scientific interest. A knowledge-based literature screening tool to determine such alleles, as defined by citations within the PubMed database is developed in the instant invention. Based upon that first-order knowledge-based screen, a set of approximately 210 alleles have been identified of greatest potential interest. The measurement of significance that was adapted is the number of hits returned while performing a PubMed search with the allele name as the search phrase. Entrez PubMed provides a set of seven server-side programs, known as eUtils, to allow access to Entrez data outside of the regular web query interface. Esearch is an eUtil which searches and retrieves primary IDs and term translations, and optionally retains results for future use in the user's environment. It retrieves the IDs based on different options such as relative date, date ranges, retrieval mode, sort order etc. A script that posts an eUtil URL to NCBI for each search phrase has been written to search for HLA alleles. It searches for the search phrase in the title, the abstract and the full text of the PubMed entries. The script then retrieves the results, which are in XML format. The number of hits for each search phrase, which is contained in the "count" tag in the XML file, can be obtained by parsing the file. Such a search was performed on alleles from 21 HLA genes. Table 10 lists the results for the automated PubMed Search for the 21 HLA Allele Citations.

TABLE 10

PubMed Citations	No. of Alleles	PubMed Citations	No. of Alleles
0	1970	51-100	0
1	104	101-200	4
2-5	62	201-500	1
6-10	18	501-1000	2
11-20	7	1001-2000	0
21-30	1	>2000	9
31-50	1		

Table 9 shows that out of the 2179 alleles, 105 of them have two or more citations in PubMed. 210 have one or greater citations. For the 105 alleles that are found to be most significant, i.e., 2 or greater citations, their frequency information in the four major populations in the United States was collected. This list can be further edited relative to the current understanding of the role of HLA in vaccine response, and a set of approximately ten PCR reactions and 400 SNP-specific probes will be designed which uniquely identify the allele set derived from the original set of 210 allele candidates. These serves as the raw material from which to build the HLA Chip and associated protocols for sample processing. All 210 "1 or more" PubMed alleles is listed in the chart in Figure 10. The number of PubMed citations is listed to the right.

EXAMPLE 10

Candidate probe selection for HLA chips

IMGT/HLA database provides multiple alignment for each HLA locus. The aligned sequences are in a format such that the SNPs are clearly marked. The SNPs are what make one allele different from another. Combinations of SNPs can uniquely represent an allele. From the instant computations it is know that there are 125 SNPs in HLA-B exon 2 and 93 in exon 3, encoding 553 and 562 alleles, respectively. At each location where SNPs occur a short n-mers (13- to 15- nucleotide long) surrounding the SNPs were "filtered" out as

templates for candidate probe sequences. There are two types of templates, the simple templates and the composite templates. A simple template contains only one SNP. For instance, there is a SNP at position 36 in exon 3 such that there are no other SNPs occurring within 6 nucleotides on either side. A 13-mer template, consisting of the SNP in the middle and six nucleotides from both sides, was made for this SNP. The template is 5'-TGCGACXTGGGGC (SEQ ID NO: 318), where X denotes the SNP. In allele B*7301, it is an "A" at this position. In allele B*0712 and another 126 alleles, it is a "C". In the rest of HLA-B alleles, including the reference allele, it is a "G". Thus, three probe candidates were made from this template, with the A-, C- and G- polymorphism. In this case, allele B*0712 and all other alleles with a "C" at the SNP site are said to be identified by the probe candidates with the C-polymorphism. Meanwhile, allele B*7301 is said to be uniquely identified by the probe candidate with the A-polymorphism.

Multiple SNPs can occur adjacently or within very short distance such that they fall into the same template. Such templates are known as the composite templates. The 13-mer starting position 254 in exon 3 is an example of a composite template. It contains two SNPs, at position 260 and, separated by a single nucleotide in the middle. The composite template is 5'-GAAGGAXAYGCTG (SEQ ID NO: 319), where X and Y denote the two SNPs. Combinations of the SNPs were taken into considerations. Three probe candidates were made from this template. 31 alleles are identified by the probe candidate with the C- and C- polymorphism. 478 alleles are identified by the probe candidate with the G- and C-polymorphism. The remainder of the HLA-B alleles, including the reference allele, is identified by the probe candidate with the C- and A- polymorphism.

In the next step, the candidate probe sequences that may occur in the rest of the human genome, as well as in ~1000 micro organisms that are expected to appear in the clinical samples with 1-, 2- or 3- mismatches canbe excluded from further consideration. This is enabled by a recently developed computational ability known as the "background-blind" technology. The new technology enables performing exact analysis of presence/absence of all subsequences (n-mers) of size up to 22 nucleotides in sequences with order of magnitude of human genome (3Gb) in a reasonable amount time. Furthermore, it allows explicit consideration of all subsequences deriving from each sequence of interest with 1, 2, and 3 mismatches (in contrast with traditionally used heuristics based alignments like BLAST bases

probe/primers design applications).

When a master list such as shown in Figure 10 is provided, the above computations allow creation of a "complete set" and a "minimal set" of SNP specific probes. The "Complete Set" is the highly redundant compilation of all probes which identify all SNPs in all alleles of a master list like Figure 10. The Minimal Probe Set can identify the same set of alleles that the complete set identifies, but with a minimal number of probes. This minimization will be done by systematically deleting probes (SNP sites) from the complete set. In each step, the algorithm chooses a probe for deletion. If the remaining set of probes excluding the chosen probe can identify the same set of alleles among the master list, then this particular probe can be deleted without losing any coverage. Otherwise, the algorithm chooses another probe and repeats the above process. It halts when deleting any probe in the set will cause the set to identify fewer alleles. At this point, the set of probes is minimal. To this minimal set, probes that uniquely identify single alleles will be added as controls. Given two samples of allele sequences, these sets of probes can be used to measure the similarity between the two samples, i.e., which allele sequences are present in both sets. This can be extremely useful for transplantation where the compatibility between two individuals is the focus of interests. In other cases, differences among individuals can be identified. For example, two patients may react differently to the same drug or treatment because of the differences in their HLA genes. This is the main motivation behind the concept of the optimal set. An optimal set of probes of a set of alleles can uniquely identify every allele in the set with minimal number of probes. Decision-tree based algorithms can be developed to generate the optimal set.

EXAMPLE 11

Optimal number of PCRs to amplify loci of interest

Table 9 shows the estimates of the number of primary and in some instances secondary PCR reactions that will be required to amplify the ten loci of interest. Both primary and secondary (nested) PCR will be required for A, B and C. The three primary PCRs can be multiplexed as one reaction, followed by a second multiplex reaction for the nested steps. Thus only 2 PCR reactions would be required for the entire set of Class I genes.

For the Class II genes (Table 11) only 1 exon is required to analyze major SNP

variation among the Class II genes. If required nested PCR may be applied to cleanly isolate DRB1. DPA1, DPB1, DQA1 and DQB1 may not require nesting to discriminate against pseudogenes. If the primary PCR for DRB1 can be multiplexed then Class II amplification can be accomplished with one primary and one secondary PCR reaction.

TABLE 11

	Primary PCR	2nd PCR ex 2	2nd PCR ex 3
Class I			
HLA A	1	1	1
HLA B	1	1	1
HLA C	1	1	1
Class II			
HLA DRB1	1	1	no
HLA DRB3	1	1	no
HLA DRB4	1	1	no
HLA DRB5	1	1	no
HLA DPA1	0	1	no
HLA DPB1	0	1	no
HLA DQA1	0	1	no
HLA DQB1	0	1	no

The following references are cited herein:

1. Charron, D. (1997) *Genetic Diversity of HLA: Functional and Medical Implications.* EDK, Paris, France.

2. Marsh, *et al.* (2000) *HLA FactsBook.* Academic Press, London, UK.

3. Trowsdale and Campbell (1992) *Eur J Immunogenet.* 19, 45-55.

4. Little, A. M. and Parhams P. (1999) *Rev Immunogenet.* 1999, 105-123.

5. Trachtenberg, *et al.* (2003) *Nat Med.*, 9, 928-935.

6. Kruskall, *et al.* (1992) *J Exp Med.* 175, 495-502.

7. McCloskey, *et al.* (1993) *Handbook of HLA typing techniques.* Hui, K. M., Bidwell, J. L., Eds. Boca Raton, FL. CRC Press, Inc, 175-247.

8. Wordswoth, P. (1991) *Immunol Lett.*, 29, 37-39.

9. Olerup and Setterquist (1993) *Handbook of HLA typing techniques.* Hui, K. M., Bidwell, J. L., Eds. Boca Raton, FL. CRC Press, Inc, 149-174.

10. Teutsch, *et al.* (1996) *Eur J Immunogenet.*, 23, 107-120.

11. Hurley, C. K. (1997) *Tissue Antigens*, 49, 323-328.

12. Dinauer, *et al.* (2000) *Tissue Antigens*, 55, 364-368.

13. Schena, *et al.* (1995) *Science* 270, 467-70

14. Saiki, *et al.* (1989) Proc Natl Acad Sci USA 86 6230-6234.

15. Guo, *et al.* (2001) *Genome Res.*, 12, 447-457.

16. Wang, *et al.* (2003) The Eighth Annual Structural Biology Symposium of Sealy Center for Structural Biology, p 157, Galveston, Texas.

17. Belosludtsev, *et al.* (2001) *Anal. Biochem.* 292, 250-256.

18. Robinson, *et al.* (2003) *Nuc. Acids Res.* 31, 311-314.

19. Fofanov, *et al.* (2002) The 2002 Bioinformatics Symposium, Keck/GCC Bioinformatics Consortium, p 14.

20. Fofanov, *et al.* (2002) The Seventh Structural Biology Symposium of Sealy Center for Structural Biology, p 51, Galveston, Texas.

21. Hayney, *et al.* (1995) *Mayo Clin Proc.* 70, 951-954.

22. Mitra, *et al.* (2003) *Analytica Chemica Acta*, 469, 141-148.

23. Lemeshko, *et al.* (2001) *Nuc. Acids Res.* 29, 3051-3058.

Any patents or publications mentioned in this specification are indicative of the levels of those skilled in the art to which the invention pertains. Further, these patents and publications are incorporated by reference herein to the same extent as if each individual publication was specifically and individually incorporated by reference.

One skilled in the art will appreciate readily that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those objects, ends and advantages inherent herein. Changes therein and other uses which are encompassed within the spirit of the invention as defined by the scope of the claims will occur to those skilled in the art.

WHAT IS CLAIMED IS:

1. Primers for amplifying an HLA gene.

2. The primers of claim 1, wherein the HLA gene is HLA A, B or DRB1 or an exon therein.

3. The primers of claim 2, wherein the HLA-A primers have sequences shown in SEQ ID NOS: 14-15, the HLA-A exon 2 primers have the sequences shown in SEQ ID NOS: 20-21 and the HLA-A exon 3 primers have the sequences shown in SEQ ID NOS: 22-26.

4. The primers of claim 2, wherein the HLA-B primers have sequences shown in SEQ ID NOS: 16-19, the HLA-B exon 2 primers have the sequences shown in SEQ ID NOS: 27-28 and the HLA-B exon 3 primers have the sequences shown in SEQ ID NOS: 29-31.

5. The primers of claim 2, wherein the HLA-DRB1 primers have sequences shown in SEQ ID NOS: 32-37 and the HLA-B exon 2 primers have the sequences shown in SEQ ID NOS: 38-47.

6. Hybridization probes for detecting single nucleotide polymorphisms (SNPs) in an HLA gene, comprising: about a 9-15mer oligonucleotide complementary to a region containing the SNP; and 5' and 3' flanking sequences.

7. The hybridization probes of claim 6, wherein the flanking sequences are oligo-thymidines or an oligo-thymidine-like polyanionic polymer.

8. The hybridization probes of claim 6, wherein the SNPs are located in HLA-A exon 2 or exon 3, HLA-B exon 2 or exon 3 or HLA-DRB1 exon 2.

9. The hybridization probes of claim 8, wherein the HLA-A exon 2 probes have the sequences shown in SEQ ID NOS: 48-49 and the HLA-A exon 3 probes have the sequences shown in SEQ ID NOS: 100-153.

10. The hybridization probes of claim 8, wherein the HLA-B exon 2 probes have the sequences shown in SEQ ID NOS: 154-237 and the HLA-B exon 3 probes have the sequences shown in SEQ ID NOS: 238-239.

11. A microarray device for allelotyping an HLA gene, comprising:

a substrate having a cationic surface; and

a monolayer comprising one or more of the hybridization probes of claim 6 adsorbed thereto.

12. The microarray device of claim 11, wherein the cationic surface comprises an aminosilane, a quanidinium, tin oxide, aluminum oxide or zirconium oxide or other equivalently charged moiety.

13. The microarray device of claim 11, wherein the substrate is glass, plastic or metal.

14. The microarray device of claim 11, further comprising:

an oligo-thymidine co-absorbed with the hybridization probes.

15. The microarray device of claim 14, wherein the oligo-thymidine has about 20 to about 40 thymidine.

16. The microarray device of claim 14, further comprising:

a fluorescent dye linked to the oligo-thymidine

17. The microarray device of claim 11, further comprising:

a capping agent.

18. A kit for population-scale HLA genotyping, comprising:

gene-specific primers for amplifying an HLA gene; and

the microarray device of claim 11.

19. The kit of claim 18, wherein the primers have the sequences shown in SEQ ID NOS: 14-47.

20. The kit of claim 18, further comprising:

buffers and polymerases for a PCR reaction or a fluorescent dye or a combination thereof.

21. A system for real-time high throughput population-scale HLA allelotyping in a field environment, comprising:

the microarray device of claim 11;

means for collecting and purifying DNA samples from individuals comprising a population;

means for generating by PCR cRNA target amplicons of one or more HLA genes of interest from the collected DNA; and

means for assigning an HLA allelotype to each individual HLA gene of interest; wherein individual means and devices comprising said system are portable and operable in real time within the field environment.

22. The system of claim 21, wherein the HLA gene is HLA-A, HLA-B or HLA-DRB1.

23. The system of claim 21, wherein the means for collecting DNA samples comprises:

a container suitable to receive a buccal wash sample, a buccal swab sample or a blood sample collected from the individuals.

24. The system of claim 21, wherein the means for generating target amplicons comprises HLA gene-specific primers for amplifying the HLA gene of interest.

25. The system of claim 24, wherein the gene-specific primers have sequences shown in SEQ ID NOS: 14-47.

26. The system of claim 21, wherein said means for assigning an HLA-allelotype to each individual comprises:

an imaging device adapted to detect hybridization patterns formed on the microarray device after hybridization of the target to the hybridization probes adsorbed thereto; and

pattern recognition software comprising a set of algorithms adapted to recognize the imaged hybridization patterns as HLA allelotypes.

27. The system of claim 26, wherein the hybridization probes have sequences shown in SEQ ID NOS: 48-289.

28. The system of claim 21, wherein real time high throughput genotyping is about 200 to about 300 HLA allelotypes per hour per system operated.

29. A method for real time population-scale HLA allelotyping in a field environment, comprising:

collecting DNA from one or members of the population;

purifying the DNA for analysis;

generating a target amplicon from an HLA gene of interest comprising the DNA using gene specific primers;

contacting the hybridization probes comprising the microarray of claim 11 with the target; and

imaging the hybridization pattern formed after the contact wherein each HLA allelotype has a pattern associated therewith.

30. The method of claim 29, further comprising storing the collected DNA.

31. The method of claim 29, wherein the DNA is collected from blood, with a buccal wash or with a buccal swab.

32. The method of claim 29, wherein the gene-specific primers have sequences shown in SEQ ID NOS: 14-47.

33. The method of claim 29, wherein the hybridization probes have sequences shown in SEQ ID NOS: 48-289.

34. The method of claim 29, further comprising:

assessing a risk of infection by a biological agent or weapon for each individual based on the assigned allelotype.

35. The method of claim 29, further comprising:

assessing a response to a particular vaccine against the biological agent or weapon by each individual.

36. The method of claim 29, wherein the allelotype assigned to each individual of the population comprises a means of identification thereof.

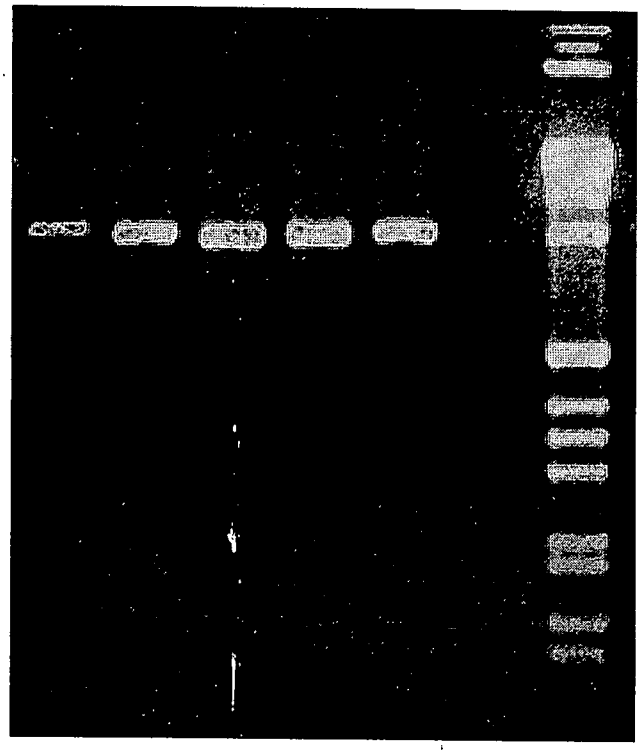

Fig. 1A

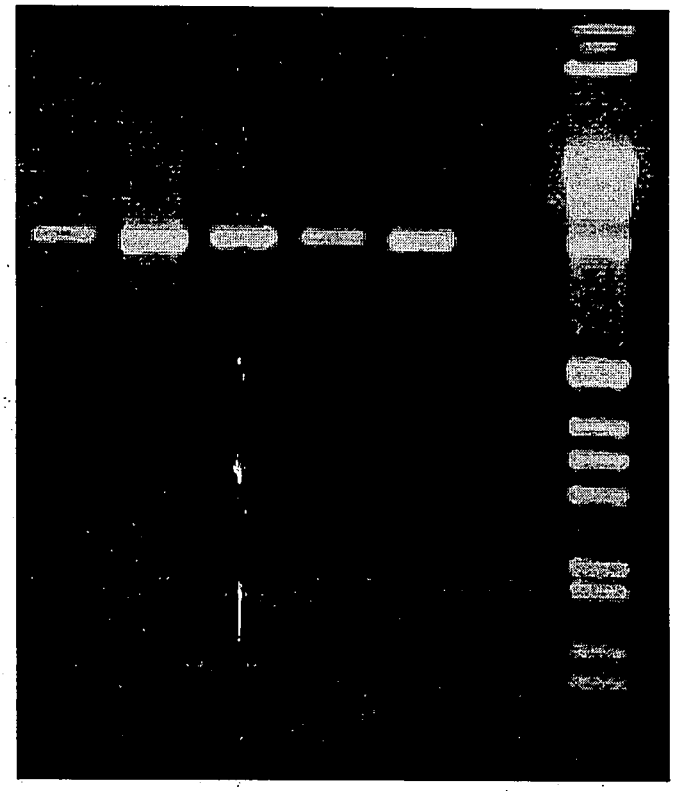

Fig. 1B

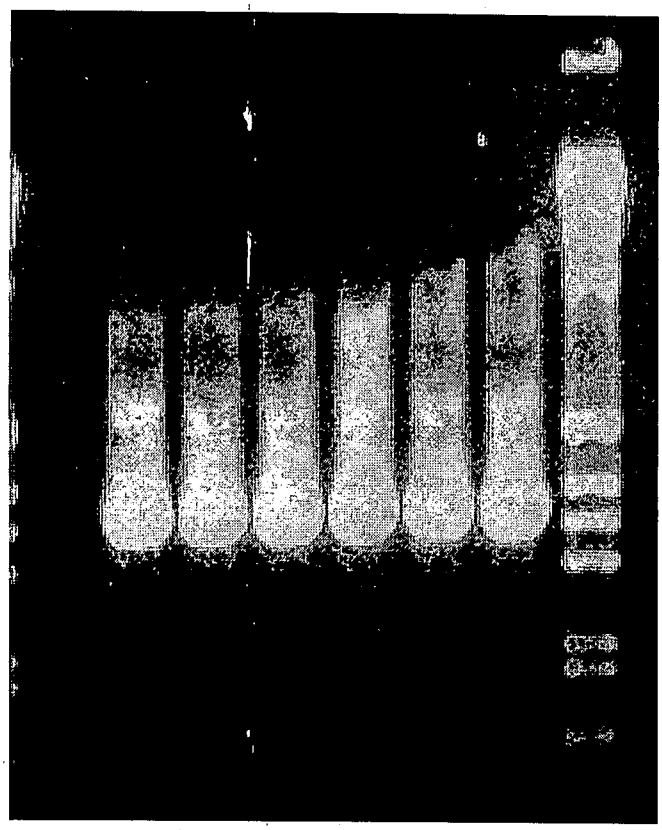

Fig. 1C

Fig. 1D

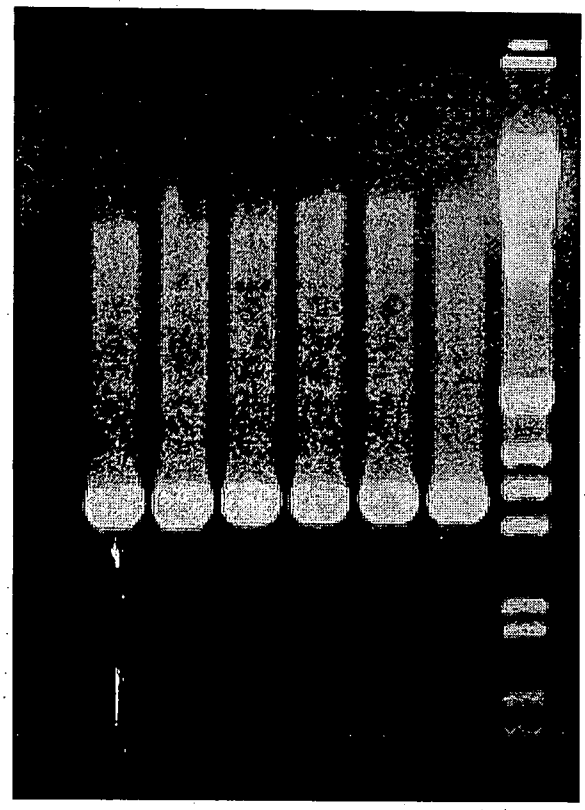

Fig. 1E

Fig. 1F

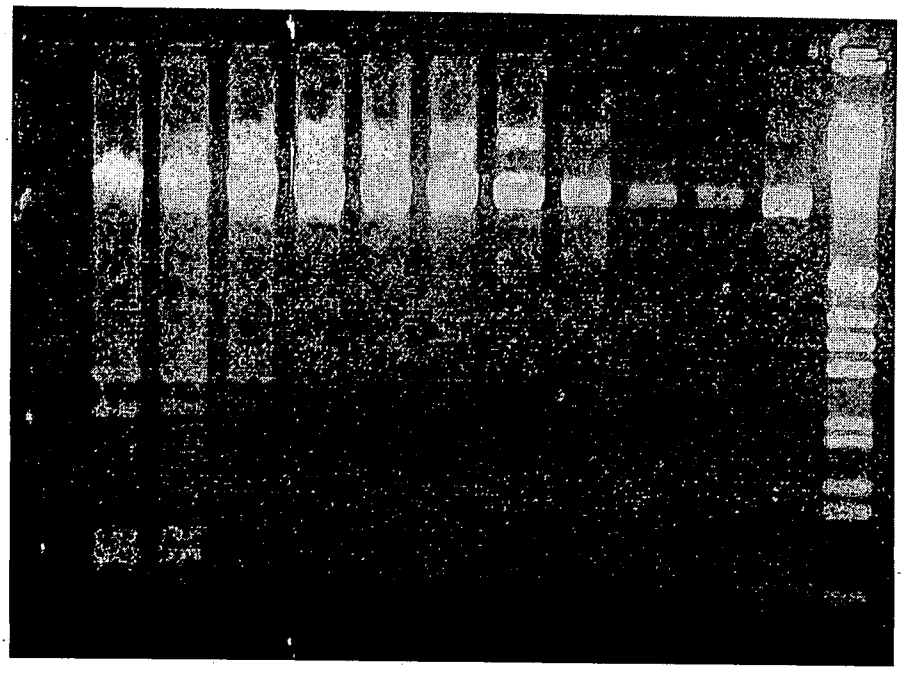

Fig. 1G

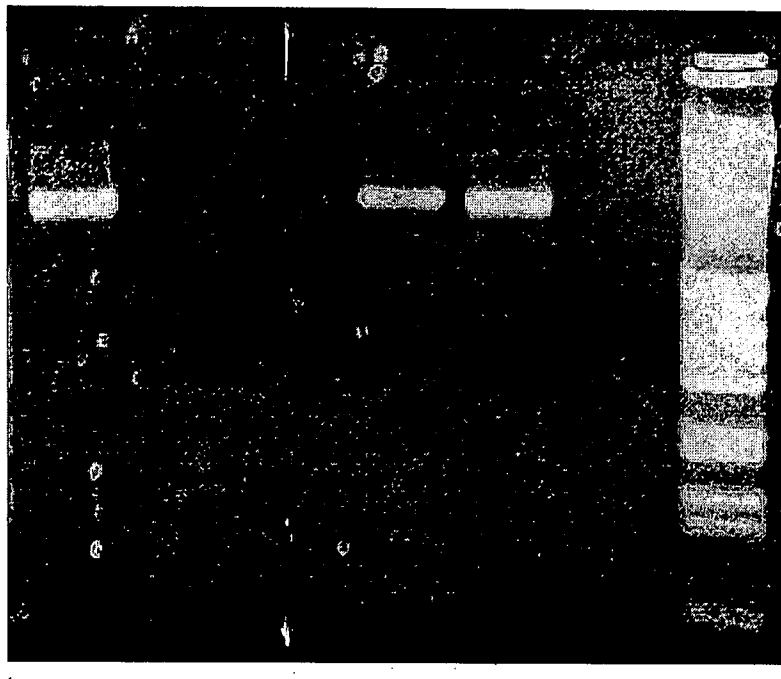

Fig. 1H

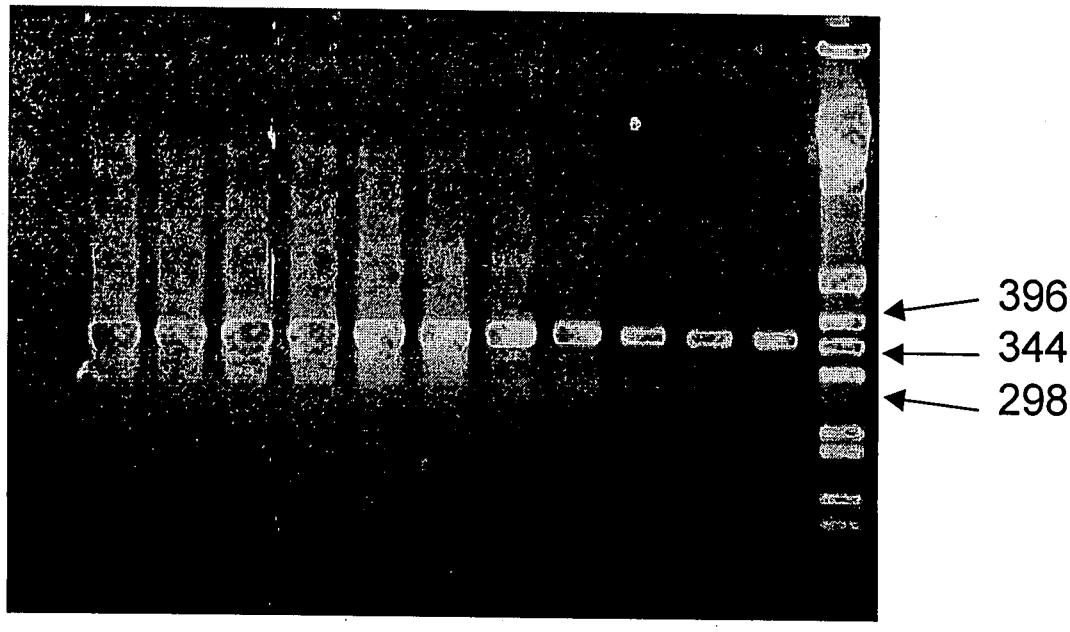

Fig. 1I

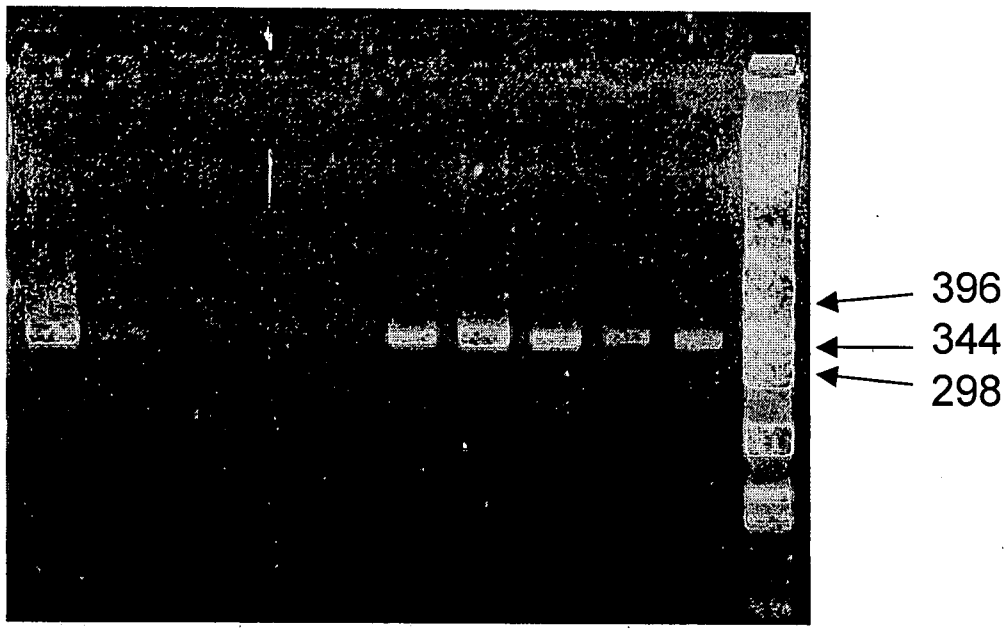

Fig. 1J

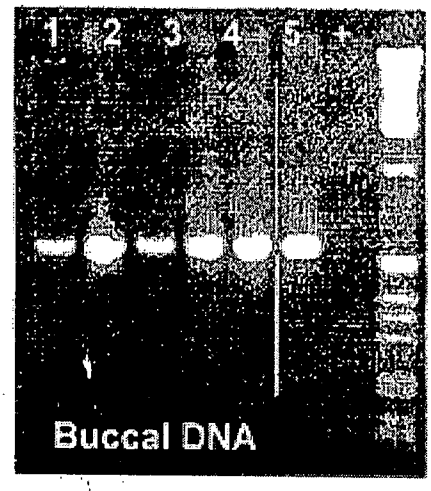

Fig. 2A

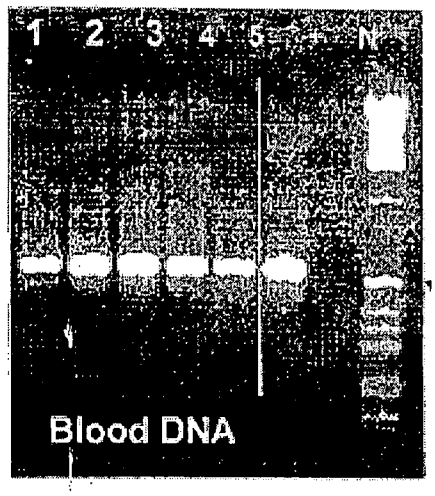

Fig. 2B

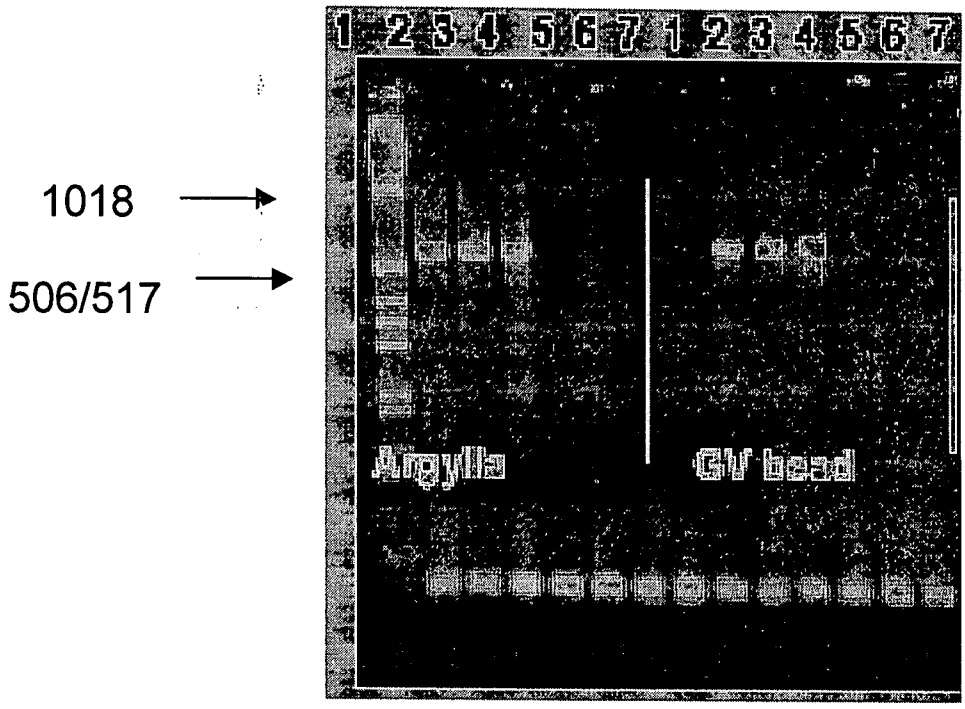
1 2 3 4 5 6 7 1 2 3 4 5 6 7
1018
506/517
Argyll
GV bead

Fig.3A

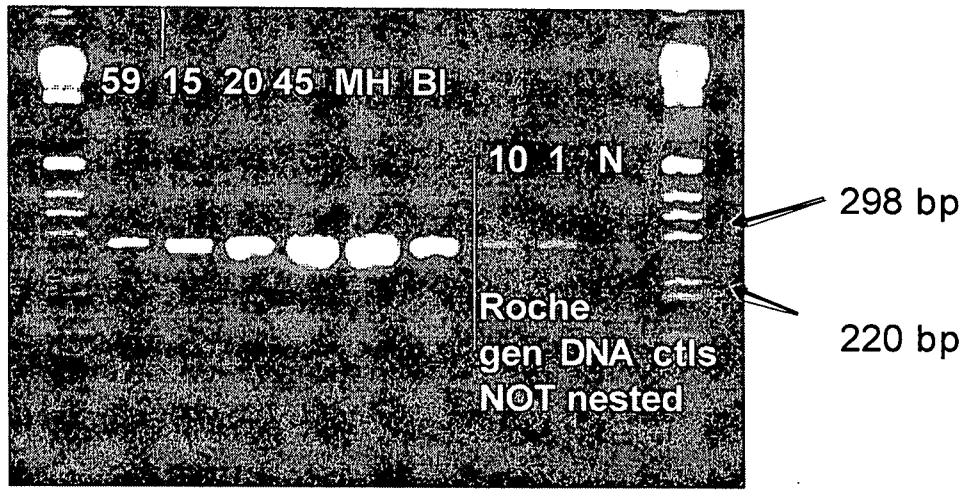
59 15 20 45 MH BI
10 1 N
298 bp
220 bp
Roche
gen DNA ctls
NOT nested

Fig.3B

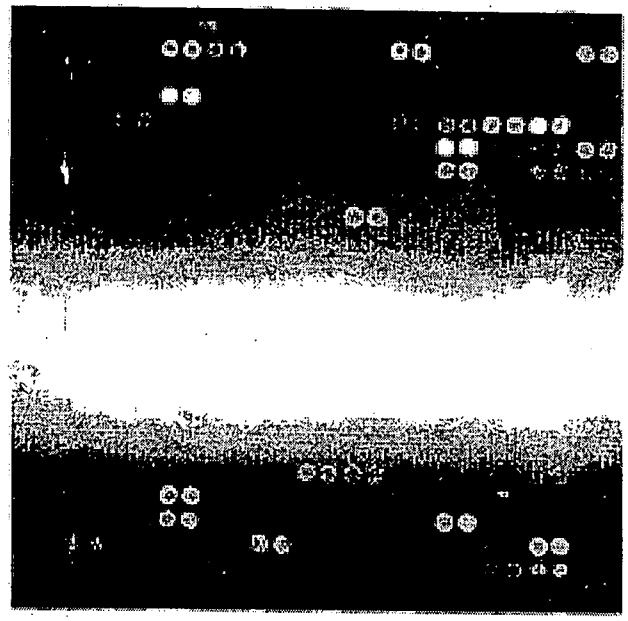

Fig.4A

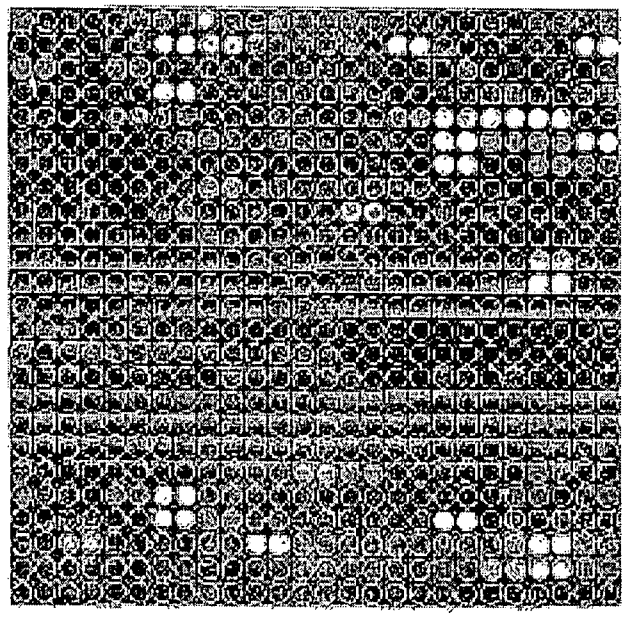

Fig.4B

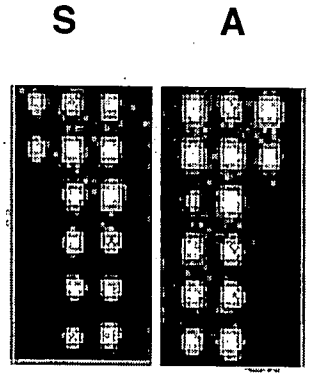

Fig.5A

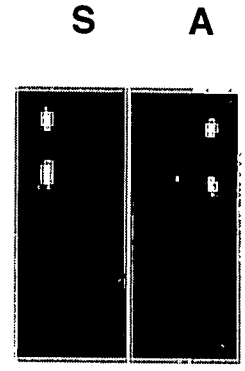

Fig. 5B

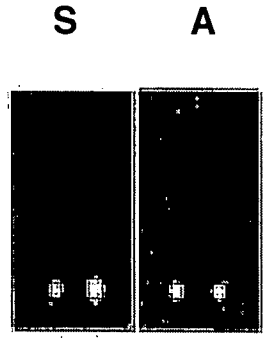

Fig.5C

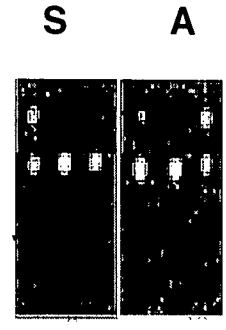

Fig.5D

Probe designs	Codon 09	Codon 20	Codon 24	Codon 31	Codon 50	Codon 57	Codon 59
Concensus Sequence Variant 1	TAC	CCC	TCA	ACC	CCG	CCG	TAT
Variant 2	GAC	CTC	GCA	ACG	CCT	CCT	TTT
Variant 3	CAC						
UCLA Cl-072 *0702/*3508	TAC/TAC	CCC/CCC	TCA/GCA	ACS/ACC	CCG/CCA	CCK/CCG	TAT/TAT
Sequencing	TAC	CCC	TCA/GCA	ACC	CCG/CCA	CCG	TAT
Array	TAC	CCC	TCA/GCA	ACC	CCG/CCA	CCG	TAT
UCLA Cl-023 *1401/*5501	TAC/TAC	CCC/CCC	TCA/GCA	ACG/ACG	CCG/CCG	CCG/CCG	TAT/TAT
Sequencing	TAC	CCC	TCA/GCA	ACG	CCG	CCG	TAT
Array	TAC	CCC	TCA/GCA	ACG	CCG	CCG	TAT
UCLA Cl-021 *27XX/*7301	CAC/CAC	CCC/CCC	ACC/ACC	ACG/ACC	CCG/CCG	CCG/CCG	TAT/TAT
Sequencing	CAC	CCC	ACC/ACC	ACC	CCG	CCG	TAT
Array	CAC	CCC	NP/NP	ACC/2SNP	CCG	CCG	TAT
UCLA Cl-055 *0702/*8101	TAC/TAC	CCC/CCC	TCA/TCA	ACC/ACC	CCG/CCG	CCG/CCG	TAT/TAT
Sequencing	TAC	CCC	TCA	ACC	CCG	CCG	TAT
Array	TAC	CCC	TCA	ACC	CCG	CCG	TAT
UCLA Cl-060- *5502/*5601	TAC/TAC	CCC/CCC	GCA/GCA	ACG/ACG	CCG/CCG	CCG/CCG	TAT/TAT
Sequencing	TAC	CCC	GCA	ACG	CCG	CCG	TAT
Array	TAC	CCC	GCA	ACG	CCG	CCG	TAT
UCLA Cl-007 *1521/*15XX	TAC/TAC	CCC/CCC	GCA/DCM	ACG/ACS	CCA/CCR	CCG/CCG	TAT/TAT
Sequencing	TAC	CCC	GCA	ACC	CCA	CCG	TAT
Array	TAC	CCC	GCA	ACC	CCA	CCG	TAT
UCLA Cl-026 *4403/*5801	TAC/TAC	CCC/CCC	ACC/GCA	ACG/ACC	CCA/CCA	CCG/CCG	TAT/TAT
Sequencing	TAC	CCC	GCA	ACC	CCA	CCG	TAT
Array	TAC	CCC	NP/GCA	ACC	CCA	CCG	TAT
UCLA Cl-027 *1515/*4001	TAC/CAC	CCC/CCC	GCA/ACC	ACC/ACG	CCA/CCR	CCG/CCG	TAT/TAT
Sequencing	TAC/CAC	CCC	RCM	ACC/ACG	CCA	CCG	TAT
Array	TAC/CAC	CCC	GCA/NP	ACC/ACG	CCA	CCG	TAT
UCLA Cl-057 *0702/*0801	TAC/GAC	CCC/CCC	TCA/TCA	ACS/ACG	CCG/CCG	CCK/CCG	TAT/TAT
Sequencing	TAC/GAC	CCC	TCA	ACC/ACG	CCG	CCG	TAT
Array	TAC/GAC	CCC	TCA	ACC/ACG	CCG	CCG	TAT
UCLA Cl-051 *4006/*4601	CAC/TAC	CCC/CCC	ACC/GCA	ACG/ACC	CCA/CCA	CCG/CCG	TAT/TAT
Sequencing	CAC/TAC	CCC	RCM	ACG/ACC	CCA	CCG	TAT
Array	CAC/TAC	CCC	NP/GCA	2SNP/ACC	CCA	CCG	TAT
UCLA Cl-039 *4011/*4101	CAC/CAC	CCC/CCC	ACC/ACC	ACG/ACG	CCA/CCA	CCG/CCG	TAT/TAT
Sequencing	CAC	CCC	ACC	ACG	CCA	CCG	TAT
Array	CAC	CCC	NP	2SNP	CCG/CCA	CCG	TAT
UCLA Cl-044 *5101/*1501	TAC/TAC	CCC/CCC	GCA/GCA	ACC/ACS	CCR/CCA	CCG/CCG	TAT/TAT
Sequencing	TAC	CCC	GCA	ACC	CCA	CCG	TAT
Array	TAC	CCC	GCA	ACC	CCA	CCG	TAT

Fig. 6A

Probe designs	Codon 67	Codon 69	Codon 74	Codon 84
Concensus Sequence Variant 1	TAC	GCCCAGGCA (SEQ ID NO:308)	GAC	TAC
Variant 2	TGC	ACCAACACA (SEQ ID NO:309)	TAC	CAC
Variant 3	TTC			
UCLA CI-072 *0702/*3508	TAC/TTC	GCCCAGGCA/ACCAACACA	GAC/TAC	TAC/TAC
Sequencing	TAC/TTC	RCCMASRCA (SEQ ID NO: 310)	GAC/TAC	TAC
Array	TAC/TTC	GCCCAGGCA/ACCAACACA	GAC/TAC	TAC
UCLA CI-023 *1401/*5501	TGC/TAC	GCCCAGGCA/ACCAACACA	GAC/GAC	TAC/TAC
Sequencing	TGC/TAC	RCCMASRCA	GAC	TAC
Array	TGC/TAC	GCCCAGGCA/ACCAACACA	GAC	TAC
UCLA CI-021 *27XX/*7301	TNC/TGC	RCCMASRCA/GCCAAGGCA (SEQ ID NO: 311)	KAC/GAC	TAC/TAC
Sequencing	TGC	*AMBIGUOUS*	GAC	TAC
Array	TGC	**NP/NP**	GAC	TAC
UCLA CI-055 *0702/*8101	TAC/TAC	GCCCAGGCA/GCCCAGGCA	GAC/GAC	TAC/TAC
Sequencing	TAC	GCCCAGGCA	GAC	TAC
Array	TAC	GCCCAGGCA	GAC	TAC
UCLA CI-060- *5502/*5601	TAC/TAC	GCCCAGGCA/GCCCAGGCA	GAC/GAC	TAC/TAC
Sequencing	TAC	GCCCAGGCA	GAC	TAC
Array	TAC	GCCCAGGCA	GAC	TAC
UCLA CI-007 *1521/*15XX	TAGC/TNC	ACCAACACA/RYCHMSRCR (SEQ ID NO:312)	TAC/KAC	TAC/TAC
Sequencing	TSC	ACCAACACA	TAC	TAC
Array	**2SNP**	ACCAACACA	TAC	TAC
UCLA CI-026 *4403/*5801	TCC/ATG	ACCAACACA/GCCTCCGCG (SEQ ID NO:313)	TAC/TAC	TAC/TAC
Sequencing	*ATG*	*GCCTCCGCG*	TAC	TAC
Array	NP/NP	NO CALL/NP	TAC	TAC
UCLA CI-027 *1515/*4001	TCC/TCC	ACCAACACA/ACCAACACA	TAC/TAC	TAC/TAC
Sequencing	TCC	ACCAACACA	TAC	TAC
Array	**NP/NP**	ACCAACACA	TAC	TAC
UCLA CI-057 *0702/*0801	TAC/TTC	GCCCAGGCA/ACCAACACA	GAC/GAC	TAC/TAC
Sequencing	TAC/TTC	GCCCAGGCA/ACCAACACA	GAC	TAC
Array	TAC/TTC	GCCCAGGCA/ACCAACACA	GAC	TAC
UCLA CI-051 *4006/*4601	TCC/TAC	ACCAACACA/SSCCAGGCA (SEQ ID NO: 314)	TAC/GAC	TAC/TAC
Sequencing	TCC/TAC	MSCMASRCA (SEQ ID NO: 315)	TAC/GAC	TAC
Array	**NP/2SNP**	**ACCAACACA/NP**	**2SNP/GAC**	TAC
UCLA CI-039 *4011/*4101	TCC/TCC	ACCAACACA/ACCAACACA	TAC/TAC	TAC/TAC
Sequencing	TCC	ACCAACACA	TAC	TAC
Array	**NP/NP**	ACCAACACA	TAC	TAC
UCLA CI-044 *5101/*1501	TTC/TCC	ACYAACACA (SEQ ID NO: 316)/ACCAACACA	TAC/TAC	TAC/TAC
Sequencing	TYC	ACCAACACA	TAC	TAC
Array	**TTC/NP**	ACCAACACA	TAC	TAC

Fig. 6B

72

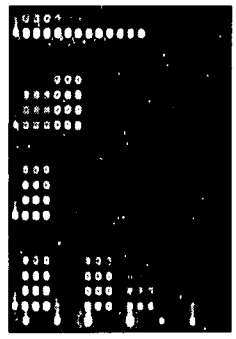

Fig.7A

21

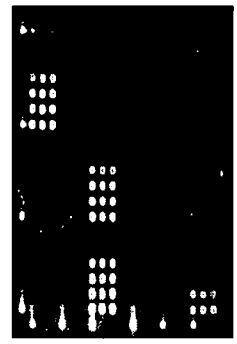

Fig.7B

27

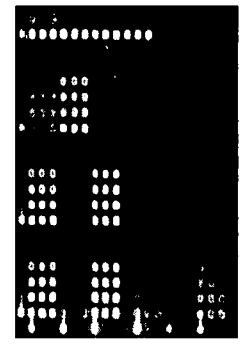

Fig.7C

57

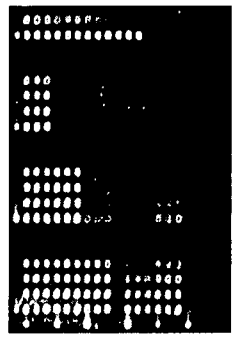

Fig.7D

MH

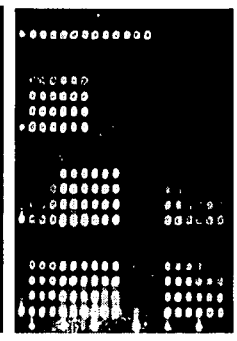

Fig.7E

BI

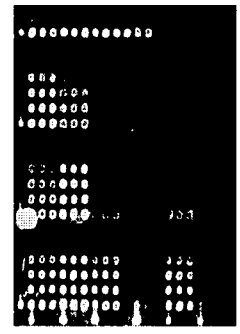

Fig.7F

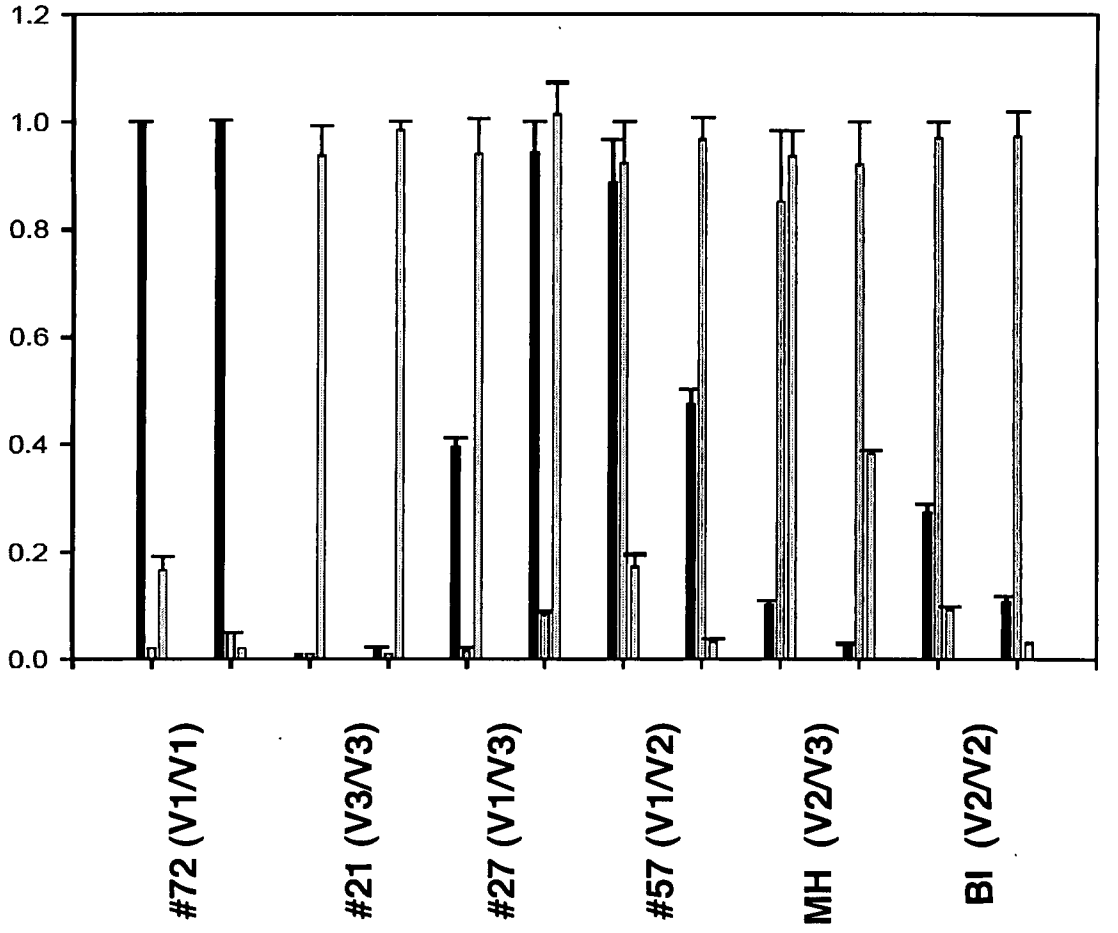
1.2
1.0
0.8
0.6
0.4
0.2
0.0
#72 (V1/V1)
#21 (V3/V3)
#27 (V1/V3)
#57 (V1/V2)
MH (V2/V3)
BI (V2/V2)

Fig. 7G

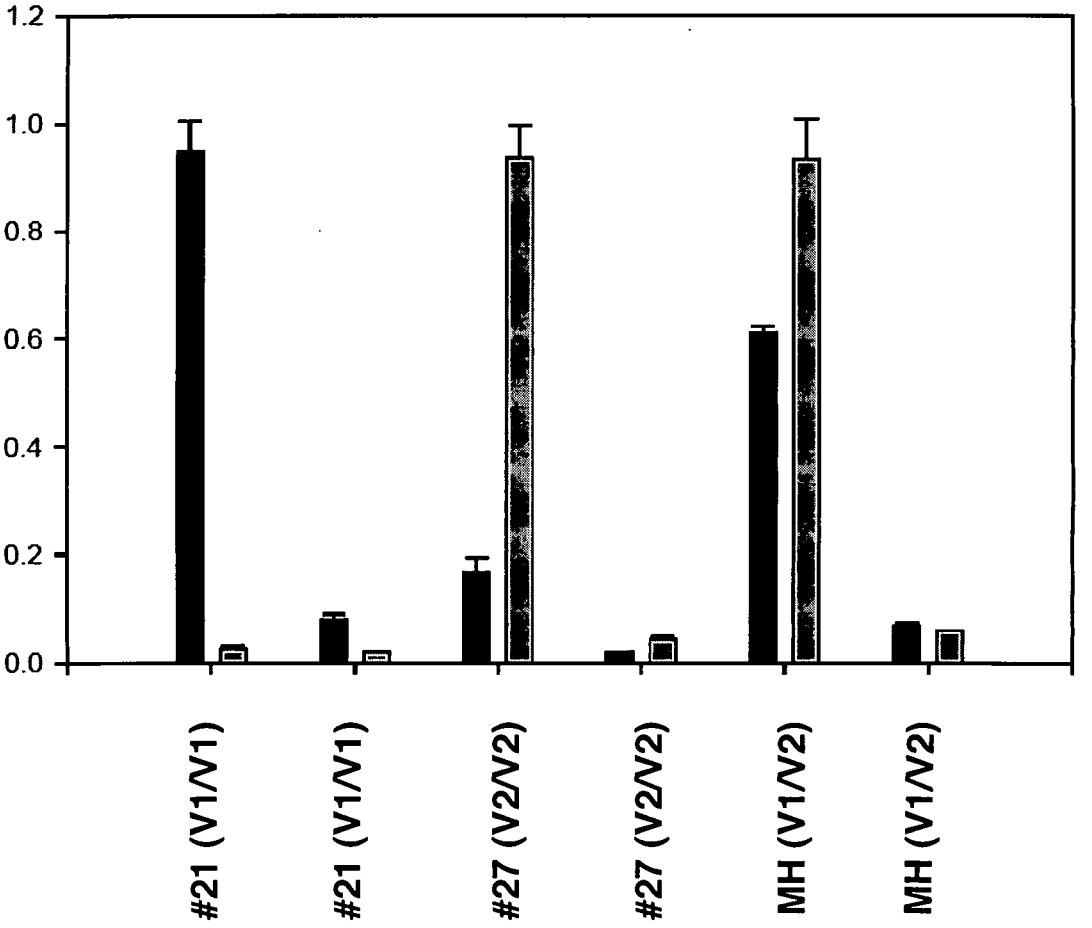
1.2
1.0
0.8
0.6
0.4
0.2
0.0
#21 (V1/V1)
#21 (V1/V1)
#27 (V2/V2)
#27 (V2/V2)
MH (V1/V2)
MH (V1/V2)

Fig. 7H

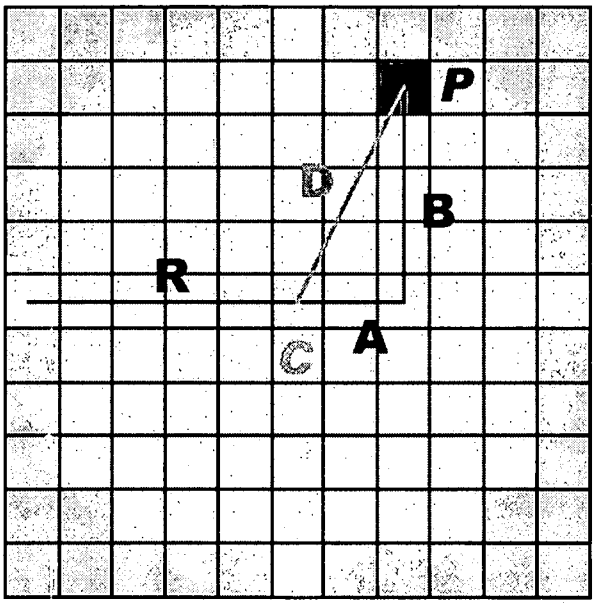
P
D
B
R
C
A

Fig. 8

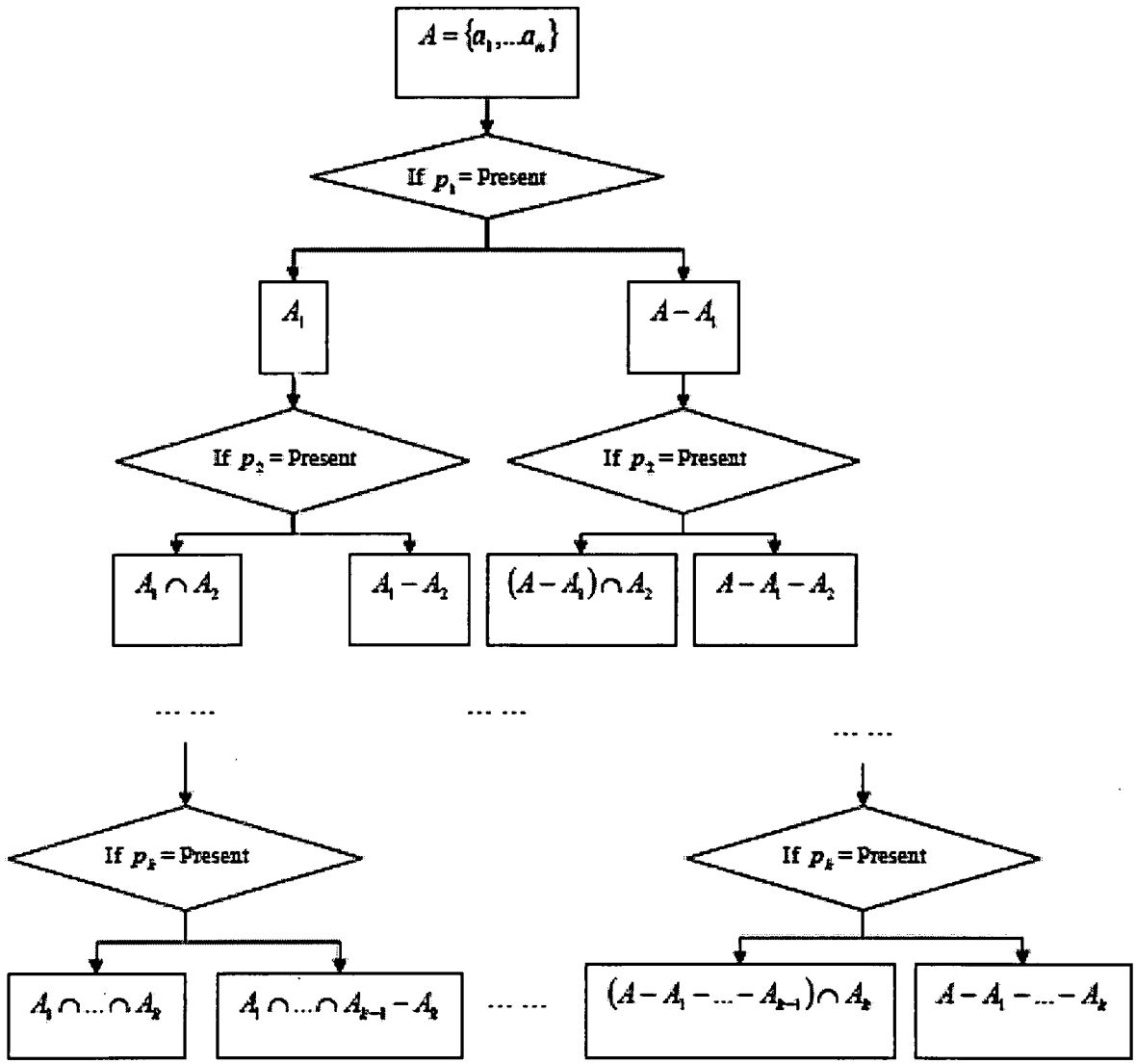
$A = \{a_1, \ldots a_n\}$
If p_1 = Present
A_1
$A - A_1$
If p_2 = Present
If p_2 = Present
$A_1 \cap A_2$
$A_1 - A_2$
$(A - A_1) \cap A_2$
$A - A_1 - A_2$
If p_k = Present
If p_k = Present
$A_1 \cap \ldots \cap A_k$
$A_1 \cap \ldots \cap A_{k-1} - A_k$
$(A - A_1 - \ldots - A_{k-1}) \cap A_k$
$A - A_1 - \ldots - A_k$

Fig. 9

HLA-A	hits	HLA-B	hits	HLA-B cont	hits	HLA-C	hits
A*0102	2	B*0703	3	B*4101	2	Cw*0105	1
A*0103	2	B*0707	1	B*4102	1	Cw*0303	1
A*0104N	1	B*0708	2	B*4104	2	Cw*0304	1
A*0202	4	B*0709	2	B*4201	1	Cw*0305	1
A*0204	6	B*0720	1	B*4202	3	Cw*0403	2
A*0205	7	B*0726	1	B*420502	1	Cw*0408	1
A*0207	9	B*0732	1	B*4405	4	Cw*040502	2
A*0209	2	B*0802	1	B*4413	1	Cw*0502	33
A*0211	4	B*0804	2	B*4415	2	Cw*050202	1
A*0214	1	B*0809	1	B*4417	1	Cw*0507	1
A*0226	1	B*0811	1	B*4422	1	Cw*120201	1
A*0232N	1	B*1301	3	B*4423N	1	Cw*1209	1
A*0240	1	B*1304	1	B*4425	1	Cw*1210	1
A*0240	1	B*1311	1	B*4426	1	Cw*1507	1
A*0249	1	B*1403	3	B*4427	1	Cw*1602	2
A*0252	1	B*1502	3	B*4431	1	Cw*1701	1
A*0302	2	B*1503	4	B*4437	1		
A*0310	1	B*1513	2	B*4501	2		
A*1102	2	B*1521	1	B*4601	0		
A*1107	1	B*1545	1	B*4703	3		
A*1108	1	B*1546	1	B*4704	1		
A*2301	7	B*1567	1	B*4001	6		
A*2306	1	B*1560	1	B*4902	1		
A*2402010202	1	B*1017N	1	B*5002	4		
A*2602	2	B*2701	5	B*5103	2		
A*2608	2	B*2702	15	B*5105	2		
A*2614	1	B*2703	6	B*5107	1		
A*2615	1	B*2706	8	B*5108	5		
A*2902	2	B*2709	15	B*5110	1		
A*3004	1	B*2712	2	B*5111N	1		
A*3301	4	B*2723	1	B*5401	6		
A*3306	1	B*3503	11	B*5502	2		
A*6601	2	B*3520	3	B*5504	1		
A*6802	9	B*3531	1	B*5509	9		
A*6808	1	B*3701	1	B*5514	1		
A*6812	1	B*3705	1	B*5601	4		
A*6813	1	B*3801	1	B*5602	1		
A*6901	2	B*3805	1	B*5603	1		
A*8001	4	B*3905	1	B*560502	1		
		B*3909	1	B*5606	9		
		B*3910	2	B*5614	4		
		B*3911	1	B*5801	0		
		B*3924	1	B*5802	1		
		B*3928	1	B*5809	1		
		B*4012	1	B*7301	4		
		B*4013	1	B*8101	4		
		B*4024	1	B*8102	1		
		B*4046	1	B*8201	1		

Fig. 10A

HLA-DRB	hits	HLA-DPB1	hits	HLA-DQ	hits	Others	hits
DRB1*0101	16	DPB1*0402	10	DQA1*0103	11	DMA*0101	2
DRB1*0102	3033	DPB1*0501	14	DQA1*0105	695	DMA*0104	2
DRB1*0103	3033	DPB1*0201	4	DQA1*0201	670	DRA*0101	3
DRB1*0402	3041	DPB1*0401	362	DQA1*0102	6	F*0101101	1
DRB1*0404	28	DPB1*0501	1	DQA1*0103	4	F*0101102	1
DRB1*0403	3	DPB1*9001	1	DQA1*0401B02	1	G*0105N	9
DRB1*0410	5			DQA1*0402	1		
DRB1*0411	1			DQA1*0404	1		
DRB1*0412	3033			DQA1*0505	1		
DRB1*0416	1			DQB1*0202	3		
DRB1*0704	1			DQB1*0104	1		
DRB1*0806	1			DQB1*0401	10		
DRB1*0807	2			DQB1*0402	5		
DRB1*0808	1			DQB1*0602	109		
DRB1*0901N1	3033			DQB1*0603	12		
DRB1*1117	3033			DQB1*0609	1		
DRB1*1135	3033						
DRB1*1143	1						
DRB1*1303	1						
DRB1*1312	3033						
DRB1*1406	3						
DRB1*1408	1						
DRB1*1602	8						
DRB1*1604	3						
DRB1*1606N2	1						
DRB1*1617	3033						
DRB1*1703	4						
DRB3*0105	120						
DRB3*0107	120						
DRB3*0202	120						

Fig. 10B