



Haplotype Counting by Next-Generation Sequencing for Ultrasensitive Human DNA Detection

Marija Debeljak,* Donald N. Freed,[†] Jane A. Welch,[‡] Lisa Haley,* Katie Beierl,* Brian S. Iglehart,[§] Aparna Pallavajjala,* Christopher D. Gocke,* Mary S. Leffell,[§] Ming-Tseh Lin,* Jonathan Pevsner,[†] Sarah J. Wheelan,[¶] and James R. Eshleman*[¶]

From the Departments of Pathology,* Neuroscience,[†] Genetics,[‡] Medicine,[§] and Oncology,[¶] Johns Hopkins University, Johns Hopkins Medical Institutions, Baltimore, Maryland

CME Accreditation Statement: This activity ("JMD 2014 CME Program in Molecular Diagnostics") has been planned and implemented in accordance with the Essential Areas and policies of the Accreditation Council for Continuing Medical Education (ACCME) through the joint sponsorship of the American Society for Clinical Pathology (ASCP) and the American Society for Investigative Pathology (ASIP). ASCP is accredited by the ACCME to provide continuing medical education for physicians.

The ASCP designates this journal-based CME activity ("JMD 2014 CME Program in Molecular Diagnostics") for a maximum of 48 AMA PRA Category 1 Credit(s)[™]. Physicians should only claim credit commensurate with the extent of their participation in the activity.

CME Disclosures: The authors of this article and the planning committee members and staff have no relevant financial relationships with commercial interests to disclose.

Accepted for publication
April 22, 2014.

Address correspondence to
James R. Eshleman, M.D.,
Ph.D., The Sol Goldman
Pancreatic Cancer Research
Center, Departments of
Pathology and Oncology, Johns
Hopkins University School of
Medicine, CRB II, Room 344,
1550 Orleans St, Baltimore,
MD 21231. E-mail: jeshlem@jhmi.edu.

Human identity testing is critical to the fields of forensics, paternity, and hematopoietic stem cell transplantation. Most bone marrow (BM) engraftment testing currently uses microsatellites or short tandem repeats that are resolved by capillary electrophoresis. Single-nucleotide polymorphisms (SNPs) are theoretically a better choice among polymorphic DNA; however, ultrasensitive detection of SNPs using next-generation sequencing is currently not possible because of its inherently high error rate. We circumvent this problem by analyzing blocks of closely spaced SNPs, or haplotypes. As proof-of-principle, we chose the *HLA-A* locus because it is highly polymorphic and is already genotyped to select proper donors for BM transplant recipients. We aligned common *HLA-A* alleles and identified a region containing 18 closely spaced SNPs, flanked by nonpolymorphic DNA for primer placement. Analysis of cell line mixtures shows that the assay is accurate and precise, and has a lower limit of detection of approximately 0.01%. The BM from a series of hematopoietic stem cell transplantation patients who tested as all donor by short tandem repeat analysis demonstrated 0% to 1.5% patient DNA. Comprehensive analysis of the human genome using the 1000 Genomes database identified many additional loci that could be used for this purpose. This assay may prove useful to identify hematopoietic stem cell transplantation patients destined to relapse, microchimerism associated with solid organ transplantation, forensic applications, and possibly patient identification. (*J Mol Diagn* 2014, 16: 495–503; <http://dx.doi.org/10.1016/j.jmoldx.2014.04.003>)

Myeloablative conditioning and allogeneic stem cell transplantation have historically been limited to the treatment of lethal hematological malignancies in children or young adults. More recently, with the advent of highly immunosuppressive, nonmyeloablative regimens, the clinical use of allogeneic stem cell transplantation has expanded to include older, less fit patients with hematological malignancies and patients with nonmalignant disorders, such as sickle-cell disease.^{1–4} Nonmyeloablative conditioning regimens offer the additional safeguard of recovery of autologous hematopoiesis in the event

of graft rejection and may be a safer option in patients at risk for immune-mediated rejection of the donor graft.

Chimerism testing at set intervals is an effective method for detecting graft rejection or recurrence of the original hematopoietic neoplasm after allogeneic hematopoietic stem cell transplantation (HSCT) [with either bone marrow (BM) or peripheral blood stem cells]. Decades ago, BM

Supported in part by The Sol Goldman Foundation.
Disclosures: None declared.

engraftment (BME) monitoring was performed using Southern blot analysis and minisatellite or variable number of tandem repeats loci.⁵ Today, short tandem repeat (STR) or microsatellite loci are most commonly used for this purpose.^{6–8} STRs are composed of 10 to 60 tandemly repeated units, in which each unit is 1 to 6 bases in length. They are widely distributed throughout the human genome and highly variable between individuals; therefore, they allow for excellent differentiation between individuals, including patient and donor, even if they are closely related. Most laboratories use multiplex PCR-based kits, originally developed for forensics analysis using Combined DNA Index System loci.^{7,9,10} STR analysis most commonly involves PCR amplification using fluorescently labeled primers, followed by amplicon separation by capillary electrophoresis.

Other polymorphic DNAs that could be used to monitor BME include single-nucleotide polymorphisms (SNPs).^{11–13} SNPs are theoretically superior to STR-based analyses because analysis of STR loci by capillary electrophoresis is relatively insensitive [limit of detection (LD), 1% to 5%] and microsatellite alleles of varying length amplify with different efficiencies, thus making them inherently biased. STR amplification can also be difficult in the setting of highly degraded DNA. However, SNPs are less attractive as targets because of their inherently lower informativity (eg, only two possible bases for a bi-allelic SNP versus ≥ 10 alleles for some microsatellites), requiring many more SNPs to be tested to identify those that distinguish donor from recipient. For example, we previously estimated that one would need to screen >20 to 30 individual SNPs to confidently identify one SNP where the donor is homozygous for one allele and an unrelated recipient is homozygous for the other allele.¹¹ Fewer would need to be included if heterozygotes were included, but more would have to be analyzed for related individuals.

Recently emerging next-generation sequencing (NGS) technologies, along with their decreasing costs, are now feasible for clinical testing. However, all NGS technologies currently have high error rates, in the range of 0.04% to 1% at each base,¹⁴ which precludes their use for ultrasensitive detection of one SNP. One solution to this problem is sequencing blocks of closely spaced SNPs (ie, haplotypes). Haplotypes are regions of the genome, where polymorphic areas are sufficiently close that they are inherited together, including either genes (eg, *HLA-A* and *HLA-B*) within a locus or multiple SNPs within a region of DNA.

Herein, we first used the *HLA-A* locus as proof-of-principle to demonstrate that this approach permits high sensitivity, precision, and accuracy. We then studied BM samples from a cohort of patients who engrafted after HSCT and tested as all donors by STRs, and found that low-level patient DNA is commonly present. To identify additional loci that could be used for this purpose, we comprehensively analyzed the human genome and identified other regions with highly informative haplotypes. We discuss additional situations where routine haplotyping patient samples could improve patient safety.

Materials and Methods

Sample Collection and Preparation

Cell lines HCT116 (*A*01:01:01* and *A*02:01:01*) and DLD-1 (*A*02:01:01* and *A*24:02:01:01*) were chosen because of their available *HLA-A* haplotypes, and we confirmed them in the Immunogenetics Laboratory of the Johns Hopkins University (Baltimore, MD), as well as their DNA fingerprint profile using the AmpFISTR Profiler Plus PCR Amplification Kit (Life Technologies, Carlsbad, CA). Many cells were expanded to permit making cell dilutions, because we have found that cell dilutions are generally more accurate than DNA dilutions (data not shown). Cell-to-cell dilutions were made, where an appropriate number of HCT116 cells was added to 10 million DLD-1 cells for each dilution. DNA was extracted from each cell pellet using a DNeasy Blood and Tissue kit (Qiagen, Valencia, CA). Extracted DNA was quantified by Quantifiler (Applied Biosystems, Carlsbad, CA) and stored at -20°C .

BME Samples

Samples from 18 patients who underwent allogeneic HSCT were obtained from the Molecular Diagnostic Laboratory at the Johns Hopkins Hospital on an Institutional Review Board–approved protocol. Samples were selected on the basis of the disease type, the ability to distinguish patient from donor alleles, their classification as donor by STR analysis, and the fact that at least 600 ng (100,000 genomes) of DNA was available to test (Table 1). Each BM sample was prepared for sequencing, as described below, and sequenced on the Ion Torrent Personal Genome Machine (PGM) to ensure approximately 100,000 reads were obtained.

HLA-A PCR Amplification

Forward and reverse primers were ordered with Ion Torrent–specific adaptors (A- and P1-adaptors) added at their 5' ends. Briefly, the first round PCR included 600 ng (100,000 genomes) of DNA, 200 nmol/L forward and reverse primers, in Platinum PCR SuperMix High Fidelity (Invitrogen, Carlsbad, CA) in a total 100- μL reaction volume. The forward primer was as follows: 5'-CCATCTCATCCCTGCGTGTCTCCGAC*tca*gAGGACCTGCGCTCTTGGAC-3' [A-adaptor (underlined letters), library key (lowercase letters), and *HLA-A* primer (italicized letters)]. The reverse primer was as follows: 5'-CCTCTCTATGGGCAGTCGGTGATCGTTCTCCAGGTATCTGCGGA-3' [P1-adaptor (underlined letters) and *HLA-A* reverse (italicized letters)]. The expected size of the full-length PCR product was 298 bp, including adaptors. After amplification, adaptor containing PCR product was visualized by gel electrophoresis (Novex TBE Gels; Life Technologies) and quantified using high-sensitivity double-stranded DNA reagents and a Qubit 2.0 fluorometer version 3.10 (Invitrogen). With the desire to paint a single molecule on each bead for

Table 1 BMT Patient Characteristics and Patient DNA Detected

Patient no.	Disease	Patient <i>HLA-A</i>	Donor <i>HLA-A</i>	Difference*	Patients (%)
BME14	AML	<i>A*02:05:01/A*03:01:01:01</i>	<i>A*02:05:01/A*01:01:01:01</i>	11,6	0.000
BME20	MDS	<i>A*29:02:01:01/A*26:01:01</i>	<i>A*01:01:01:01/A*26:01:01</i>	12,6	0.001
BME18	AML	<i>A*03:01:01:01/A*30:01:01</i>	<i>A*11:01:01/A*30:01:01</i>	7,5	0.019
BME10	AML	<i>A*32:01:01/A*34:02:01</i>	<i>A*01:01:01:01/A*34:02:01</i>	12,3	0.030
BME19	AML	<i>A*24:02:01:01/A*68:01:01:01</i>	<i>A*32:01:01/A*68:01:01:01</i>	6,7	0.052
BME25	ALL	<i>A*01:01:01:01/A*03:01:01:01</i>	<i>A*33:01:01/A*03:01:01:01</i>	12,11	0.060
BME30	AML	<i>A*03:01:01:01/A*24:02:01:01</i>	<i>A*03:01:01:01/A*01:01:01:01</i>	7,6	0.136
BME27	AML	<i>A*02:01:01:01/A*24:02:01:01</i>	<i>A*02:01:01:01/A*02:01:01:01</i>	7	0.140
BME16	AML	<i>A*02:XX/A*26:XX</i>	<i>A*02:xx/A*03:xx</i>	6,9	0.189
BME11	AML	<i>A*30:02:01/A*74:01</i>	<i>A*68:02:01:01/A*74:01</i>	8,2	0.190
BME4	AML-M6	<i>A*02:01:01:01/A*29:02:01:01</i>	<i>A*02:01:01:01/A*68:01:01:01</i>	6,5	0.312
BME26	ALL	<i>A*11:01:01/A*30:01:01</i>	<i>A*25:01:01/A*30:01:01</i>	6,9	0.328
BME1	MM	<i>A*02:01:01:01/A*23:01:01</i>	<i>A*30:xx/A*02:xx</i>	4,7	0.416
BME22	CLL	<i>A*02:01:01:01/A*66:01</i>	<i>A*01:01:01:01/A*66:01</i>	11,9	0.640
BME21	MCL	<i>A*68:02:01:01/A*29:02:01:01</i>	<i>A*30:01:01/A*29:02:01:01</i>	8,6	0.750
BME29	AML	<i>A*11:01:01/A*26:01:01</i>	<i>A*11:01:01/A*11:01:01</i>	6	0.863
BME23	DLBCL	<i>A*02:01:01:01/A*26:01:01</i>	<i>A*02:01:01:01/A*02:01:01:01</i>	9	1.110
BME24	HD	<i>A*01:01:01:01/A*29:02:01:01</i>	<i>A*30:02:01/A*29:02:01:01</i>	13,12	1.470

*Number of SNP differences between unique alleles (bold) and between the unique patient and shared alleles.

ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; CLL, chronic lymphocytic leukemia; DLBCL, diffuse large B-cell lymphoma; HD, Hodgkin disease; MCL, mantle cell lymphoma; MDS, myelodysplastic syndrome; MM, multiple myeloma.

emulsion PCR, samples were diluted to a 0.03 nmol/L working concentration. During emulsion PCR, each single molecule produces a clone of progeny molecules (estimated to be approximately 500,000) for sequencing.

Ion Torrent PGM Library Preparation and Sequencing

Emulsion PCR using 20 to 25 µL working stock, NGS, and mapping to hg19 were all done per the manufacturer’s protocol (Life Technologies). Assessment of the percentage amplicon-containing beads was performed per the manufacturer’s protocol (Life Technologies) and measured with a Qubit 2.0 fluorometer (Invitrogen). Amplicon-coated beads were analyzed on 314 and 316 chips using the Ion PGM Sequencing 200 kit on the Life Technologies’ Ion Torrent PGM, semiconductor sequencer, which detects dNTP incorporation using the hydrogen ion that is released (along with pyrophosphate) when a dNTP is incorporated into an elongating DNA strand.^{15,16}

Microsatellite Analysis

PCR amplification of nine microsatellites (AmpFISTR Profiler kit; Applied Biosystems) or 15 microsatellites (Identifiler; Applied Biosystems) was performed according to the manufacturer’s instructions. Amplicons were resolved on a capillary electrophoresis ABI3130xl Genetic Analyzer (Applied Biosystems).

Bioinformatics/Analysis

Initial processing of the data was done using the Ion Torrent platform-specific pipeline software Torrent Suite version 3.2.1 to generate sequence reads, trim adapter sequences,

and remove poor quality reads. Resulting sequence files were aligned to an hg19 reference sequence. Further analysis of Fastq files was done using Geneious Pro 5.5.7, where perfectly matched reads were counted. Some samples were log₁₀ transformed and graphed (GraphPad Prism version 5.04; GraphPad Software Inc., San Diego, CA).

Analysis of molecular specificity was done with homozygous *HLA-A*01:01:01:01* and *HLA-A*02:01:01:01* samples (hereafter simply *A*01* and *A*02*, respectively). Each sample was analyzed for the other allele, followed by the introduction of the other allele’s base at 11 SNP positions. For example, homozygous *A*01* was analyzed for perfectly matched *A*01* reads, followed by perfect *A*02* reads. On this, one *A*02*-specific base was introduced into the *A*01* sequence at each of the 11 SNP positions and analyzed. After introducing one *A*02*-specific base at all of the possible positions, two *A*02*-specific bases were introduced, repeating the process for 11 possible combinations. The process was continued all of the way to substituting 10 *A*02*-specific bases. The number of reads found, for each combination, was averaged and the percentage of reads found was graphed.

The BM transplant samples were analyzed for the amount of patient (patient unique allele) in the sample. After eliminating reads representing a haplotype shared by both individuals, we calculated percentage patient DNA (patient/patient + donor). For cases where three alleles were shared, the equation [(2 × unique patient)/(unique patient + shared patient – donor)] was used.

Bioinformatics Analysis of 1000 Genomes Database

By use of 1000 Genomes release version 3_20110521 from build hg18 (<http://www.1000genomes.org>, last accessed

A

Exon 3 position	154 (C/T)	163 (A/G)	178 (C/T)	184 (A/C/ G/T)	195 (C/T)	196 (A/G/T)	202 (C/T)	216 (A/G)	217 (C/G)	227 (C/G)	228 (G/T)	
cDNA position	497	506	521	527	538	539	545	559	560	570	571	
Sample	A1	A1	T	G	T	C	C	G	T	C	G	G
Probe	A2	C	A	C	T	T	T	C	A	C	G	T
		-	A	-	-	-	-	-	-	-	-	-
		-	-	-	-	T	-	-	-	-	-	-
		-	-	-	-	-	-	C	-	-	-	-
		-	-	-	-	-	-	-	C	-	-	-
Total 10A1 + 1A2												9
No. of Reads												201,605

B

Sample	Probe	C	A	C	T	T	T	C	A	C	G	T	No. of Reads
A2	A2	C	A	C	T	T	T	C	A	C	G	T	187,630
A1	A1	T	G	T	C	C	G	T	C	G	C	G	0
		-	-	T	-	-	-	-	-	-	-	-	2
		-	-	-	-	-	-	-	-	-	-	-	1
Total 10A2 + 1A1													3

C

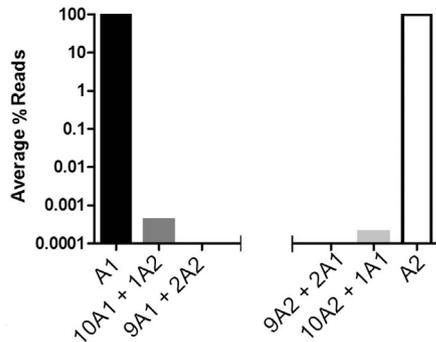


Figure 3 Molecular specificity. Samples homozygous for the *A*01* allele (A) and the *A*02* allele (B) were PCR amplified and NGS sequenced. Each was analyzed for the presence of the other allele and for hybrid molecules containing one or more SNPs of the other allele. C: Waterfall plot of the percentage reads for each combination of the 11 SNP differences for the two samples. For example, the 10A1 + 1A2 bar reflects the mean number of erroneous reads (9/11 = 0.82), in the pure *A*01* sample where any one of the 11 SNP positions matches the *A*02* allele (0.82/201,605, from A). Likewise, the 9A1 + 2A2 bar reflects the lack of any reads in the pure *A*01* sample, where two adjacent SNPs match the *A*02* allele at any of the positions.

sample with cytosine at all four SNPs, this would indicate the presence of true patient-specific DNA (Figure 1C).

Analysis of *HLA-A* Alleles

We performed alignments of common *HLA-A* alleles in the European-origin population using the Major Histocompatibility Complex database, a publically accessible platform for DNA and clinical data related to the human major histocompatibility complex (<http://www.ncbi.nlm.nih.gov/gv/mhc/main.cgi?cmd=init>, last accessed March 28, 2014). Regions with a high density of SNPs and flanked by non-polymorphic DNA were identified. One region, *HLA-A* exon 3, contained 18 possible SNPs and at least 15 major alleles in the European-origin population. We tested a series of primers surrounding this region and selected the best pair on the basis of amplification efficiency and specificity (as described in *Materials and Methods*) (Figure 1D). The number of SNP differences in this region between the most common alleles of the European-origin population is given (Figure 2); some

combinations of *HLA-A* alleles are easily differentiated, whereas others are more difficult. For example, 11 SNPs differentiate allele *A*01* from *A*02*, so that even with a relatively high base substitution error rate (eg, 1% per base), a sample homozygous for *A*01* should not contain any false-positive *A*02* allele reads. In contrast, *A*02* and *HLA-A*68:01:01:01* (hereafter *A*68*) have only one SNP difference, so a pure sample homozygous for *A*02* will likely contain reads matching *A*68* because of the relatively high error rate intrinsic to current NGS technologies.

Specificity of the Assay for SNPs

To test the possible cross talk between molecules that vary by 11 SNPs, we sequenced two samples, one homozygous for *A*01* and another homozygous for *A*02*, and analyzed each for the other allele (Figure 3). The *A*01* and *A*02* samples contained approximately 200,000 perfect matching reads. Neither pure sample contained any perfect reads of the other haplotype. We then examined the *A*01* sample files for reads containing a single *A*02* SNP at each of the 11 positions and found an average of 0.3% to 0.8% reads that contained a single error from the perfect haplotype (Figure 3A). When two SNPs of the opposite haplotype were searched, no reads were obtained. Similar results were obtained with the pure *A*02* sample (Figure 3B). A double-waterfall plot shows that when enough discriminating SNPs between two individuals' alleles exist, the assay is highly specific (Figure 3C and Supplemental Tables S1 and S2).

***HLA-A* Dose-Response Curve, Accuracy, Precision, and LD**

To assess the accuracy and LD, we generated a dilution series from two cell lines with known *HLA-A* genotypes. These samples were chosen because the two alleles of interest (*A*01* and *A*02*) vary from one another by 11 SNPs, and both vary from the commonly shared allele (*A*24*) by seven SNPs (Figure 2). Dilutions were made with cell mixes varying from 1 in 1 million (0.0001%) to 1 in 100 (1%) using a total of 10 million cells for each dilution. DNA was isolated and PCR was performed using 600 ng of DNA. We chose this relatively large amount of DNA on the basis of the desire to achieve an LD of at least 1:10,000 (0.01%) and to exceed that target LD by using 10× excess DNA. This relatively high DNA input reflects approximately 100,000 genomes (on the basis of approximately 6 pg/haploid genome) and was chosen to prevent bottlenecking and resultant allele dropout. For example, if DNA representing only 100 genomes were analyzed to 100,000 × depth of coverage, the LD is input DNA limited, not depth of coverage limited, and is at best 1%. A sample with a minor allele frequency of 0.1% would likely not be detected. Each sample was sequenced at least twice, and results were graphed as the percentage *HLA-A*01*-bearing cells (Figure 4A). Least-squares analysis generated a straight line demonstrating excellent accuracy ($R^2 = 0.93$). To determine

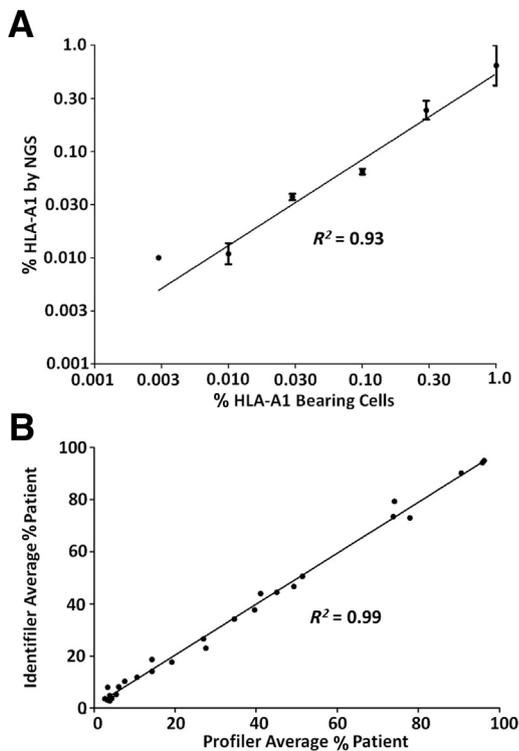


Figure 4 Linearity NGS haplotyping and microsatellite analysis. Linearity of the SNP haplotyping approach. *A*01/A*02* cells were serially diluted into *A*02/A*24* cells, and DNA was isolated and tested by *HLA-A* NGS. **A:** NGS *HLA-A*01* counts were expressed as a percentage, as a function of the percentage of *HLA-A*01* bearing cells. Error bars: ± 1 SEM. The error bar for 0.1% bearing cells is too small to be seen. The error bar for 0.003% is absent because only half the replicates were amplified. The NGS haplotyping data begin at 1% and are serially diluted down from there. **B:** Post-transplant patient samples with various percentage of patient DNA were analyzed using the Profiler and Identifier kits. Note the lack of the ability to detect DNA lower than the LD of 3%.

precision, we performed additional replicates at 0.1% and 0.01% dilutions (four total) (Supplemental Table S3). The assay was highly precise at 0.1% cell mix (mean, 0.065%; coefficient of variation, 0.12%), but less so at the 0.01% cell mix, as expected (mean, 0.012%; coefficient of variation, 0.40%). Because we detected the minor *HLA-A* allele in four of four replicates at 0.01%, but only one of two replicates at 0.003%, we concluded that 0.01% is the lower LD for this assay (Supplemental Table S3). These results compare favorably with those of microsatellites, which are highly linear and generally accurate, but with a much poorer LD of only approximately 3% (Figure 4B).

SNP Haplotype Assay Detects Patient DNA in BM Samples That Tested All Donors by STR Analysis

We selected 18 patients whose donor-patient *HLA* genotypes varied by at least four SNPs (Table 1). The allele unique to donor varied from that unique to patient by 4 to 13 SNPs. To further prevent false-positive results, the unique patient allele varied from the shared allele by 2 to 12 SNPs. We also required that 600 ng be available for testing, so that

the number of genomes (approximately 100,000) exceeded by an order of magnitude the desired LD. Several samples were excluded that could not meet these criteria.

All samples tested positive for some level of patient DNA, except for one sample. The positives ranged from 0.001% to 1.47% patient DNA (mean, 0.373) (Figure 5), confirming that the haplotype-based assay was more sensitive. We also noted that of the four samples with the highest levels of patient DNA, three of them were from patients with lymphoma.

The Human Genome Contains Clusters of SNPs That Can Be Amplified to Give Information on Haplotypes

We used variant calls from the 1000 Genomes project (<http://www.1000genomes.org>) to identify regions containing at least nine SNPs in all three populations (Europeans, Asians, and Africans; as described in *Materials and Methods*) within 300 bp. We further required these regions to be flanked by at least 20 bp of nonpolymorphic DNA. The nonpolymorphic regions provide potential primer sites, whereas the variable regions are suitable for haplotype counting. We identified 4349 such loci across the genome (Supplemental Figure S1), requiring that nine or more variants have minor allele frequencies >9% in all three populations (Supplemental Table S4). One concern is that a cluster of SNPs may not be that informative because it simply represents a high-frequency haplotype. For example, a haplotype that exists within a population at a 2% level and containing 20 SNPs, in conjunction with a second haplotype with the other base at each of the 20 positions present in 98% of the population, is a relatively uninformative marker. In contrast, a locus with 20 SNPs with 10 different haplotypes, each of which is present at 10% in the population, is highly informative.

By using the 4349 regions identified as SNP dense, we analyzed the 1000 (1092) Genomes project (<http://www.1000genomes.org>) phased-variant call files to calculate overall probability that two unrelated individuals would be different at each locus (as described in *Materials and Methods*). Briefly, at each locus, a file was generated that

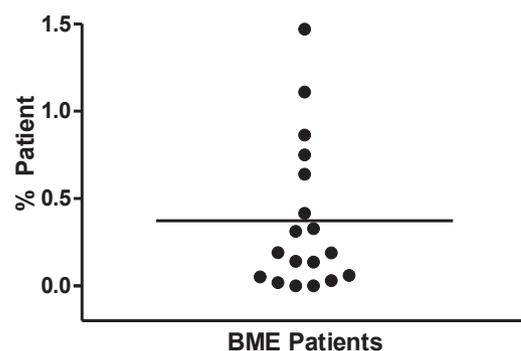


Figure 5 Percentage of patient DNA in BM samples. Unique patient DNA was detected in BM samples, which demonstrated all donors by conventional short tandem repeat assays, after hematopoietic stem cell transplantation.

Table 2 Alternate Loci for Haplotype Counting

Chromosome*	Gene [†]	Start [‡]	Stop	Length (bp)	No. of SNPs	No. of haplotypes	CEU	JPT	YRI	Mean
1	—	238855189	238855428	240	9	7	0.63	0.46	0.63	0.57
1	—	238416454	238416673	220	9	11	0.39	0.46	0.43	0.43
2	—	106142506	106142791	286	10	8	0.63	0.61	0.66	0.63
2	<i>VIT</i>	36927840	36928071	232	9	6	0.58	0.59	0.61	0.59
3	<i>ST6GAL1</i>	186654745	186655022	278	9	12	0.72	0.67	0.73	0.70
3	—	97907092	97907374	283	10	13	0.58	0.42	0.64	0.55
4	<i>SORCS2</i>	7447136	7447434	299	9	16	0.60	0.46	0.65	0.57
4	—	66995978	66996254	277	9	8	0.63	0.45	0.60	0.56
5	—	125498526	125498753	228	10	19	0.67	0.58	0.67	0.64
5	—	178259629	178259892	264	13	23	0.69	0.56	0.64	0.63
6	<i>HLA-DRB1</i>	32551493	32551765	273	9	306	0.98	0.95	0.96	0.96
6	<i>HLA-B</i>	31319390	31319668	279	10	27	0.93	0.91	0.88	0.91
7	—	64895079	64895355	277	10	12	0.59	0.61	0.66	0.62
7	—	9111002	9111297	296	9	11	0.60	0.64	0.59	0.61
8	—	6160141	6160442	302	10	13	0.56	0.67	0.63	0.62
8	<i>CSMD1</i>	3478267	3478568	302	10	13	0.63	0.51	0.66	0.60
9	—	132197723	132197947	225	11	11	0.66	0.45	0.53	0.55
9	—	95691312	95691534	223	10	14	0.39	0.35	0.46	0.40
10	—	133376146	133376435	290	9	14	0.74	0.74	0.78	0.75
10	—	123095093	123095377	285	10	18	0.74	0.78	0.69	0.74
11	<i>CNTN5</i>	99491238	99491538	301	9	7	0.62	0.50	0.60	0.57
11	—	5078880	5079122	243	10	15	0.65	0.40	0.61	0.55
12	—	82460264	82460531	268	9	7	0.65	0.66	0.66	0.66
12	—	17884472	17884719	248	9	10	0.67	0.61	0.66	0.65
13	<i>FARP1</i>	99084002	99084259	258	9	17	0.63	0.70	0.71	0.68
13	—	33553507	33553784	278	9	29	0.68	0.45	0.65	0.59
14	—	107094021	107094317	297	10	25	0.53	0.51	0.66	0.57
14	<i>TRA</i>	22736235	22736521	287	9	11	0.51	0.57	0.61	0.56
15	—	34750054	34750330	277	9	8	0.51	0.57	0.61	0.56
15	—	25047393	25047681	289	10	25	0.51	0.44	0.55	0.50
16	—	56576389	56576687	299	11	46	0.76	0.82	0.80	0.79
16	—	84540653	84540932	280	9	16	0.69	0.66	0.74	0.70
17	<i>TBCD</i>	80804085	80804372	288	9	33	0.70	0.70	0.74	0.71
17	<i>RBFOX3</i>	77148074	77148367	294	10	15	0.38	0.60	0.52	0.50
18	—	76597079	76597332	254	9	14	0.40	0.40	0.40	0.40
18	<i>CLUL1</i>	631210	631400	191	9	12	0.35	0.27	0.38	0.33
19	<i>CCDC61</i>	46500080	46500360	281	9	14	0.41	0.38	0.43	0.41
19	—	57525179	57525424	246	13	10	0.37	0.39	0.41	0.39
20	<i>PROKR2</i>	5289890	5290161	272	9	7	0.64	0.51	0.59	0.58
20	<i>SIRPA</i>	1895467	1895674	208	11	24	0.49	0.58	0.56	0.54
21	—	20548907	20549196	290	9	6	0.44	0.35	0.42	0.40
21	<i>UMODL1</i>	43528819	43529079	261	9	10	0.33	0.38	0.39	0.36

*Chromosome 22 is absent because of the lack of good polymorphic loci. X and Y were excluded.

[†]Gene, if present.

[‡]Hg19 coordinates.

—, No data.

contained all of the SNPs for each of the 2184 (2 × 1092) alleles. We then determined the number of distinct alleles and the number of times each of them was represented. We then calculated all possible combinations of alleles within donors and recipients, and finally determined the combined probability that two unrelated individuals would be informative (Table 2 and Supplemental Table S5). These putative alternate loci will need to be experimentally validated in large cohorts of patients, and experiments are in progress for this validation.

Discussion

Herein, we demonstrated that we can detect mixtures of human DNA down to an LD of 0.01% (1 in 10,000) using haplotype counting by NGS, 100-fold more sensitive than current STR-based methods. False positives from NGS are avoided by using haplotypes because if they vary from one another by enough SNPs, they should demonstrate no cross talk, even with a mutation frequency of 0.1% to 1% per base. Although we have chosen *HLA-A*^{19,20} as a proof-of-

concept in this assay, we report other haplotype loci from the 1000 Genomes project²¹ (<http://www.1000genomes.org>) that could be used for this purpose. Selection of suitable loci may be influenced by the patient's ethnic background, number of discriminating SNPs, and ease of primer placement. The set of haplotypes used for this purpose ideally would be suitable for transplant analysis of all patient ethnicities. Patient DNA is consistently detected in BM samples that test all donor by the conventional STR assay.

Detection of minimal residual disease in cases such as acute promyelocytic leukemia and subsequent early intervention has been associated with significantly higher survival.^{22,23} The haplotype counting-based assay may prove valuable to detect relapse in HSCT patients earlier than the existing microsatellite-based assays. One of the other non-chromosome 6 (containing human *HLA*) loci (Table 2) might be better for this purpose, because one mechanism that leukemic cells can escape donor anti-leukemic T cells is through the loss of the mismatched *HLA* allele, estimated to occur in 29.4% to 66.7% of such patients.^{24,25} Another limitation of the *HLA-A* haplotype approach is that some transplant donors are *HLA* identical, and loci other than *HLA* would be required to monitor such patients.

HSCT has traditionally been used to treat malignant and nonmalignant hematological disorders. In addition, with the transplantation of solid organs, some amount of lymphoid tissue may be transferred by cell migration from the donor organs, thereby generating a chimerism in the patient.²⁶ In contrast, intentional induction of a microchimerism, by injecting donor BM, is a strategy used to generate donor-specific tolerance in extremity transplantation.^{27–29} The development and persistence of a donor-recipient microchimerism may be associated with the acceptance of transplanted organs.²⁶ Using NGS to detect low levels of donor cells may provide an opportunity to document such a microchimerism (or lack thereof) and possibly manage immunosuppressive regimens to optimize engraftment.

The NGS of highly polymorphic regions may have applications outside of transplantation medicine. A microchimerism resulting from bidirectional exchange of cells between mother and fetus has been detected in women long after pregnancy using Y chromosome fluorescence *in situ* hybridization and PCR for the *SRY* gene.^{30–32} Our assay could be applied for such studies with the additional benefit of also being able to detect exchanged cells between a mother and a daughter. NGS haplotyping might also aid in the detection of rare tissue regenerative cells in the heart transplant setting. In sex-mismatched heart transplant and using Y chromosome fluorescence *in situ* hybridization, studies have shown a cardiac chimerism caused by migration of recipient cells to the grafted heart.^{33,34}

The field of forensics has relied on STR loci for identification of suspects and human remains. In theory, testing haplotypes could allow one to identify the presence of suspect in a large mixture of DNAs, such as in polysuspect cases.^{35,36} In addition to forensic applications, tools for quality control and sample tracking in a large inventory of human DNA samples would be valuable. Such tools have previously been reported using panels of SNPs,³⁷ and the NGS-haplotyping

assay could be applied in similar situations. Similar markers could be developed to distinguish among species.

To ensure patient identity and exclude the possibility of a mislabeled specimen, haplotyping to uniquely define the patient could easily be included in any NGS-based genetic test. They could also be used for any absolutely critical test, such as ABO typing of blood products,³⁸ and matched to the intended patient's genotype encoded in an implanted microchip immediately before transfusion. Although rare, wrong-patient adverse events occur in hospital settings, especially when patients share the same name or patients have similar appearances.³⁹ To circumvent these preventable errors, a biological identifier, such as those described herein, could be implemented to unequivocally distinguish patients.

NGS technologies, because of their declining costs and improved read lengths and base calling, have been making their way into the clinical setting. Haplotype counting using NGS described herein may become a valuable tool to analyze DNA mixes and perform identity testing because of its high sensitivity, precision, and accuracy.

Note Added in Proof

A group recently published that circulating DNA of donor origin can be used to detect rejection of the transplanted heart.⁴⁰

Acknowledgments

We thank Drs. Ephraim J. Fuchs, Richard J. Jones, Dwight Oliver, Allison Klein, Javier Bolanos-Meade, Bert Vogelstein, Alexis Norris, Soonweng Cho, Elliot Z. Chen, Laura Wood, Eric Stevens, Alexis Carter, Stacy Mosier, Rosie Jiang, and Jennifer Meyers for helpful discussions.

Supplemental Data

Supplemental material for this article can be found at <http://dx.doi.org/10.1016/j.jmoldx.2014.04.003>.

References

1. Slavin S, Nagler A, Naparstek E, Kapelushnik Y, Aker M, Cividalli G, Varadi G, Kirschbaum M, Ackerstein A, Samuel S, Amar A, Brautbar C, Ben-Tal O, Eldor A, Or R: Nonmyeloablative stem cell transplantation and cell therapy as an alternative to conventional bone marrow transplantation with lethal cytoreduction for the treatment of malignant and nonmalignant hematologic diseases. *Blood* 1998, 91:756–763
2. Khouri IF, Saliba RM, Giralt SA, Lee MS, Okoroji GJ, Hagemester FB, Korbling M, Younes A, Ippoliti C, Gajewski JL, McLaughlin P, Anderlini P, Donato ML, Cabanillas FF, Champlin RE: Nonablative allogeneic hematopoietic transplantation as adoptive immunotherapy for indolent lymphoma: low incidence of toxicity, acute graft-versus-host disease, and treatment-related mortality. *Blood* 2001, 98:3595–3599
3. Brodsky RA, Luznik L, Bolanos-Meade J, Leffell MS, Jones RJ, Fuchs EJ: Reduced intensity HLA-haploidentical BMT with post transplantation cyclophosphamide in nonmalignant hematologic diseases. *Bone Marrow Transplant* 2008, 42:523–527
4. Bolanos-Meade J, Fuchs EJ, Luznik L, Lanzkron SM, Gamper CJ, Jones RJ, Brodsky RA: HLA-haploidentical bone marrow transplantation

- with posttransplant cyclophosphamide expands the donor pool for patients with sickle cell disease. *Blood* 2012, 120:4285–4291
5. Jeffreys AJ, Wilson V, Thein SL: Individual-specific “fingerprints” of human DNA. *Nature* 1985, 316:76–79
 6. Scharf SJ, Smith AG, Hansen JA, McFarland C, Erlich HA: Quantitative determination of bone marrow transplant engraftment using fluorescent polymerase chain reaction primers for human identity markers. *Blood* 1995, 85:1954–1963
 7. Thiede C, Florek M, Bornhauser M, Ritter M, Mohr B, Brendel C, Ehninger G, Neubauer A: Rapid quantification of mixed chimerism using multiplex amplification of short tandem repeat markers and fluorescence detection. *Bone Marrow Transplant* 1999, 23:1055–1060
 8. Schichman SA, Suess P, Vertino AM, Gray PS: Comparison of short tandem repeat and variable number tandem repeat genetic markers for quantitative determination of allogeneic bone marrow transplant engraftment. *Bone Marrow Transplant* 2002, 29:243–248
 9. Mills KA, Even D, Murray JC: Tetranucleotide repeat polymorphism at the human alpha fibrinogen locus (FGA). *Hum Mol Genet* 1992, 1:779
 10. Huang NE, Schumm J, Budowle B: Chinese population data on three tetrameric short tandem repeat loci—HUMTHO1, TPOX, and CSF1PO—derived using multiplex PCR and manual typing. *Forensic Sci Int* 1995, 71:131–136
 11. Oliver DH, Thompson RE, Griffin CA, Eshleman JR: Use of single nucleotide polymorphisms (SNP) and real-time polymerase chain reaction for bone marrow engraftment analysis. *J Mol Diagn* 2000, 2:202–208
 12. Hochberg EP, Miklos DB, Neuberger D, Eichner DA, McLaughlin SF, Mattes-Ritz A, Alyea EP, Antin JH, Soiffer RJ, Ritz J: A novel rapid single nucleotide polymorphism (SNP)-based method for assessment of hematopoietic chimerism after allogeneic stem cell transplantation. *Blood* 2003, 101:363–369
 13. Gineikiene E, Stoskus M, Griskevicius L: Single nucleotide polymorphism-based system improves the applicability of quantitative PCR for chimerism monitoring. *J Mol Diagn* 2009, 11:66–74
 14. Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW: Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Comput Biol* 2013, 9:e1003031
 15. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al: An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011, 475:348–352
 16. Merriman B; Ion Torrent R&D Team, Rothberg JM: Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis* 2012, 33:3397–3417
 17. Howie BN, Donnelly P, Marchini J: A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009, 5:e1000529
 18. Delaneau O, Zagury JF, Marchini J: Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 2013, 10:5–6
 19. Gabriel C, Furst D, Fae I, Wenda S, Zollikofer C, Mytilineos J, Fischer GF: HLA typing by next-generation sequencing: getting closer to reality. *Tissue Antigens* 2014, 83:65–75
 20. De Santis D, Dinauer D, Duke J, Erlich HA, Holcomb CL, Lind C, Mackiewicz K, Monos D, Moudgil A, Norman P, Parham P, Sasson A, Allcock RJ: 16(th) IHIW: review of HLA typing by NGS. *Int J Immunogenet* 2013, 40:72–76
 21. Clarke L, Zheng-Bradley X, Smith R, Kulesha E, Xiao C, Toneva I, Vaughan B, Preuss D, Leinonen R, Shumway M, Sherry S, Flicek P; 1000 Genomes Project Consortium: The 1000 Genomes Project: data management and community access. *Nat Methods* 2012, 9:459–462
 22. Esteve J, Escoda L, Martin G, Rubio V, Díaz-Mediavilla J, González M, Rivas C, Alvarez C, Gonzalez San Miguel JD, Brunet S, Tomás JF, Tormo M, Sayas MJ, Sanchez Godoy P, Colomer D, Bolufer P, Sanz MA; Spanish Cooperative Group PETHEMA: Outcome of patients with acute promyelocytic leukemia failing to front-line treatment with all-trans retinoic acid and anthracycline-based chemotherapy (PETHEMA protocols LPA96 and LPA99): benefit of an early intervention. *Leukemia* 2007, 21:446–452
 23. Chendamurai E, Balasubramanian P, George B, Viswabandya A, Abraham A, Ahmed R, Alex AA, Ganesan S, Lakshmi KM, Sitaram U, Nair SC, Chandy M, Janet NB, Srivastava VM, Srivastava A, Mathews V: Role of minimal residual disease monitoring in acute promyelocytic leukemia treated with arsenic trioxide in frontline therapy. *Blood* 2012, 119:3413–3419
 24. Vago L, Perna SK, Zanussi M, Mazzi B, Barlassina C, Stanghellini MT, Perrelli NF, Cosentino C, Torri F, Angius A, Forno B, Casucci M, Bernardi M, Peccatori J, Corti C, Bondanza A, Ferrari M, Rossini S, Roncarolo MG, Bordignon C, Bonini C, Ciceri F, Fleischhauer K: Loss of mismatched HLA in leukemia after stem-cell transplantation. *N Engl J Med* 2009, 361:478–488
 25. Villalobos IB, Takahashi Y, Akatsuka Y, Muramatsu H, Nishio N, Hama A, Yagasaki H, Saji H, Kato M, Ogawa S, Kojima S: Relapse of leukemia with loss of mismatched HLA resulting from uniparental disomy after haploidentical hematopoietic stem cell transplantation. *Blood* 2010, 115:3158–3161
 26. Starzl TE, Demetris AJ, Trucco M, Murase N, Ricordi C, Ildstad S, Ramos H, Todo S, Tzakis A, Fung JJ, Nalesnik M, Zeevi A, Rudert WA, Kocova M: Cell migration and chimerism after whole-organ transplantation: the basis of graft acceptance. *Hepatology* 1993, 17:1127–1152
 27. Foster RD, Fan L, Neipp M, Kaufman C, McCalmont T, Ascher N, Ildstad S, Anthony JP, Niepp M: Donor-specific tolerance induction in composite tissue allografts. *Am J Surg* 1998, 176:418–421
 28. Foster RD, Ascher NL, McCalmont TH, Neipp M, Anthony JP, Mathes SJ: Mixed allogeneic chimerism as a reliable model for composite tissue allograft tolerance induction across major and minor histocompatibility barriers. *Transplantation* 2001, 72:791–797
 29. Schneeberger S, Gorantla VS, Brandacher G, Zeevi A, Demetris AJ, Lunz JG, Metes DM, Donnenberg AD, Shores JT, Dimartini AF, Kiss JE, Imbriglia JE, Azari K, Goitz RJ, Manders EK, Nguyen VT, Cooney DS, Wachtman GS, Keith JD, Fletcher DR, Macedo C, Planinsic R, Losee JE, Shapiro R, Starzl TE, Lee WP: Upper-extremity transplantation using a cell-based protocol to minimize immunosuppression. *Ann Surg* 2013, 257:345–351
 30. Bianchi DW, Zickwolf GK, Weil GJ, Sylvester S, DeMaria MA: Male fetal progenitor cells persist in maternal blood for as long as 27 years postpartum. *Proc Natl Acad Sci U S A* 1996, 93:705–708
 31. Cirello V, Perrino M, Colombo C, Muzza M, Filopanti M, Vicentini L, Beck-Peccoz P, Fugazzola L: Fetal cell microchimerism in papillary thyroid cancer: studies in peripheral blood and tissues. *Int J Cancer* 2010, 126:2874–2878
 32. Chan WF, Gurnot C, Montine TJ, Sonnen JA, Guthrie KA, Nelson JL: Male microchimerism in the human female brain. *PLoS One* 2012, 7:e45592
 33. Quaini F, Urbanek K, Beltrami AP, Finato N, Beltrami CA, Nadal-Ginard B, Kajstura J, Leri A, Anversa P: Chimerism of the transplanted heart. *N Engl J Med* 2002, 346:5–15
 34. Hocht-Zeisberg E, Kahnert H, Guan K, Wulf G, Hemmerlein B, Schlott T, Tenderich G, Korfer R, Raute-Kreinsen U, Hasenfuss G: Cellular repopulation of myocardial infarction in patients with sex-mismatched heart transplantation. *Eur Heart J* 2004, 25:749–758
 35. Nurit B, Anat G, Michal S, Lilach F, Maya F: Evaluating the prevalence of DNA mixtures found in fingernail samples from victims and suspects in homicide cases. *Forensic Sci Int Genet* 2011, 5:532–537
 36. Schmitt C, Benecke M: Five cases of forensic short tandem repeat DNA typing. *Electrophoresis* 1997, 18:690–694
 37. Pakstis AJ, Speed WC, Fang R, Hyland FC, Furtado MR, Kidd JR, Kidd KK: SNPs for a universal individual identification panel. *Hum Genet* 2010, 127:315–324
 38. Dzik WH, Corwin H, Goodnough LT, Higgins M, Kaplan H, Murphy M, Ness P, Shulman IA, Yomtovian R: Patient safety and blood transfusion: new solutions. *Transfus Med Rev* 2003, 17:169–180
 39. Seiden SC, Barach P: Wrong-side/wrong-site, wrong-procedure, and wrong-patient adverse events: are they preventable? *Arch Surg* 2006, 141:931–939
 40. De Vlaminck I, Valantine HA, Snyder TM, Strehl C, Cohen G, Luikart H, Neff NF, Okamoto J, Bernstein D, Weisshaar D, Quake SR, Khush KK: Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection. *Sci Transl Med* 2014, 6:241ra77