nature biotechnology

Multiplexed genotyping with sequence-tagged molecular inversion probes

Paul Hardenbol[1,3], Johan Banér[2], Maneesh Jain[1,3], Mats Nilsson[2], Eugeni A Namsaraev[1,3], George A Karlin-Neumann[1,3], Hossein Fakhrai-Rad[1,3], Mostafa Ronaghi[1], Thomas D Willis[1,3], Ulf Landegren[2] & Ronald W Davis[1]

We report on the development of molecular inversion probe (MIP) genotyping, an efficient technology for large-scale single nucleotide polymorphism (SNP) analysis. This technique uses MIPs to produce inverted sequences, which undergo a unimolecular rearrangement and are then amplified by PCR using common primers and analyzed using universal sequence tag DNA microarrays, resulting in highly specific genotyping. With this technology, multiplex analysis of more than 1,000 probes in a single tube can be done using standard laboratory equipment. Genotypes are generated with a high call rate (95%) and high accuracy (>99%) as determined by independent sequencing.

The availability of large collections of SNPs along with recent large-scale linkage disequilibrium mapping efforts[1] have brought the promise of personalized whole-genome association studies to the field of human genetics. To achieve this goal, methodologies that permit screening of hundreds of thousands of SNPs will be needed to implement such large-scale association studies on a routine basis. These methods not only will have to be inexpensive per SNP screened, but will need to consume very little genomic DNA—that is, no more than is typically obtained from a patient's blood sample. In addition, such technologies should ideally require minimal investment in infrastructure so that the technology can be made broadly available.

The challenge of genotyping the approximately 150 molecules of a given SNP locus present in 1 ng of genomic DNA is commonly met by PCR amplification of the locus before genotyping is done[2]. However, an increase in the number of target sequences for simultaneous amplification by PCR quickly leads to unmanageable levels of cross-reaction among primer pairs[3,4], whereas parallel hybridization on microarrays[5,6] lacks the specificity and sensitivity required to genotype large genomes directly.

There are only a limited number of genotyping technologies with sufficient specificity to identify an SNP from genomic DNA without prior PCR amplification. Flap endonucleases have been used to generate a sequence-specific endonuclease cascade in an isothermal fashion that can be assessed with FRET probes[7,8]. However, this technology is not readily multiplexed for high-throughput applications. Padlock probes are linear oligonucleotides, whose two ends can be joined by ligation when they hybridize to immediately adjacent target sequences[9]. As shown before[10–12], padlock probes provide sufficient specificity analyze SNPs directly, without previous amplification of the target sequences.

Unlike amplification strategies such as PCR and the Invader assay that require two specific primers, cross-reactive padlock probes can easily be distinguished from the desired circular products by methods such as exonucleolysis[9]. This offers the opportunity to add a complex pool of padlock probes to individual DNA samples to investigate large sets of genes in parallel, without a concomitant increase in the risk of cross-reactivity between different probes.

Here we present a strategy that combines DNA detection specificity and sensitivity with the potential to analyze large numbers of target sequences in parallel. Sets of padlock probes with universal tag sequences were reacted with target DNA, molecularly inverted, amplified together and identified in a multiplex analysis yielding more than 1,000 genotypes simultaneously. Using molecular inversion probes, the information content of the SNPs was reformatted into tag sequences that could be detected using a universal oligonucleotide detection array[13]. We report the application of this technique at unprecedented levels of multiplexing, resulting in a lowering of the scale, cost and sample requirements of high-throughput genotyping. The approach retained high accuracy through multiple hybridization and enzymatic processing events, and provided inherent quality control checking.

RESULTS

Selection for circularized probes using exonucleases

Most genotyping methods require PCR amplification of the region spanning the sequence variation. However, when sets of n PCR primer pairs are combined in one reaction to evaluate n target sequences, any of the $2n^2 + n$ possible pairwise primer combinations may give rise to nonspecific amplification products[3]. With padlock probes the corresponding cross-reactive ligation products create linear dimeric molecules, easily distinguished from circularized

[1]Stanford Genome Technology Center, Stanford University, 855 California Avenue, Palo Alto, California 94304, USA. [2]The Beijer Laboratory, Department of Genetics and Pathology, Rudbeck Laboratory, Se-751 85 Uppsala, Sweden. [3]Present address: ParAllele BioScience 384 Oyster Point Blvd Suite 8, S. San Francisco, California 94080, USA. Correspondence should be addressed to M.R. (mostafa@stanford.edu).

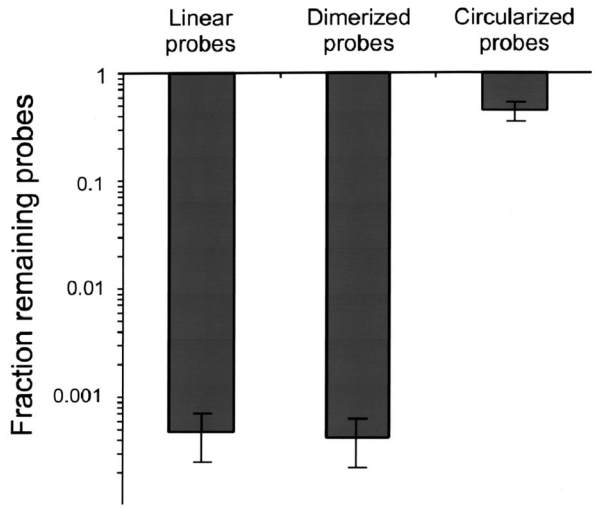

Figure 1 Selection for circularized padlock probes. Effect of exonuclease on linear monomer, dimer or on circularized padlock probes were measured by real-time PCR. Dimerized probes were produced using a ligation template that allowed two different padlock probes to be joined. The results were converted to numbers of molecules by reference to a standard dilution series. The fractions of remaining probe were calculated by dividing each reaction by the respective starting number. Error bars denote s.d. of the ratios from eight reactions.

probes by exonucleolytic degradation[9,14]. The exonuclease treatment protocol reduces the number of such linear monomeric and dimeric molecules by almost three orders of magnitude with negligible effects on circularized probes as measured by real-time PCR (**Fig. 1**). The removal of unreacted probes further reduces ligation-independent amplification events that may otherwise occur through accidental priming or templating of polymerization by the large number of linear probes (data not shown).

Molecular inversion probe (MIP) genotyping

Initially we combined pairs of padlock probes specific for alternate alleles in SNP loci. This permitted parallel genotyping of several loci in a single reaction before amplification and identification of the reaction products on tag arrays (**Fig. 2b**). Before increasing the multiplexing level, we redesigned the padlock probes to be locus-specific to avoid the need for balancing allele-specific probes at every locus (**Fig. 2a**). With this strategy only one probe was required per locus. To achieve this, the polymorphic nucleotide at the 3′ end of the probe was left out, creating a gap between the probe ends. This gap was then filled in four separate allele-specific polymerization (A, C, G and T in four different tubes) and ligation reactions[15]. Next, the probes were released from the genomic DNA by removing the uracil residues between primer sequences to avoid topological inhibition of the polymerization reaction[16]. The oligonucleotide probe undergoes a unimolecular rearrangement before amplification (**Fig. 2b**). Each probe contains a unique 20-base tag sequence that is complementary to a sequence on an Affymetrix GenFlex Tag Array. The tags are selected to be similar in melting temperature (T_m) and base composition, and maximally orthogonal in sequence complementarity. These tags amplify and hybridize under a single set of conditions with minimal cross-hybridization to each other and to other features on the microarray.

After amplification, the products are hybridized on four DNA microarrays and the components are decoded by measuring the fluorescence signals at the corresponding complementary tag site on the DNA array (**Fig. 3a**). An image of 938 amplified probes hybridized to a DNA array is shown (**Fig. 3b**). Four intensity values for each probe are generated. The two values for the expected allelic bases are compared to determine whether the sample is homozygous or heterozygous for the given SNP, and the two non-allele bases are compared to the allele bases to determine the signal-to-noise ratio (SNR) for the probe (**Table 1**). The two non-allele bases serve as internal controls that are used to reduce incorrect genotype calls owing to missing, degraded or noisy probes.

Assay performance

To investigate the performance of the method, probes were generated for 1,121 SNPs from the SNP consortium (TSC) database (http://snp.cshl.org) for a 16-megabase region on chromosome 6 centered on the linkage peak for IgA nephropathy[17] (**Table 2**). Markers were selected from the database based on map position. Of the 1,121 probes, 183 (16%) were inactive during a single synthesis step, possibly owing to such problems as errors in the database, probe design, or failures of oligonucleotide synthesis, probe synthesis or the assay itself. In a pilot study, 25 different individuals were genotyped with the 938 active probes for a total of 23,450 assays. We successfully called 21,336 full genotypes (two chromosomes) and 1,746 half genotypes (single chromosome) (95%) with a median SNR of 16.7 for allele-specific signal to non-allele signal. Half genotypes are reported when the identity of only one of two chromosomes is certain. A cluster plot of data of four of the probes used to genotype 25 individuals is shown (**Fig. 4**). No substantial differences were seen in the call rates of probes designed for all allele combinations (**Table 3**).

Accuracy was determined through independent sequencing. 1,517 loci were genotyped in a 1,517-probe multiplex analysis with ten individuals. Forward and reverse Sanger sequencing was performed on a subset (129) of PCR amplicons of 1,517 loci amplified from the same 10 individuals. Conservative reads were made manually with the identity of the forward and reverse loci blinded at the time of sequence interpretation. Accuracy of Sanger sequencing was measured by comparing reads for which the sequence of both strands existed. 359 of 367 sequence pairs were identical, for an

Table 1 Data generated from the first 10 probes from individual NA17203

Probe ID	Allele	Base call[a]	Signal A[b]	Signal G	Signal C	Signal T	SNR[c]
2,515	A/G	G/G	139	1,472	216	202	6.8
2,516	A/G	A/A	437	21	30	31	14.1
2,517	A/G	A/G	1,538	1,494	95	94	16.2
2,518	A/G	A/G	343	474	39	30	12
2,519	A/G	A/A	3,574	39	51	65	55.2
2,520	A/G	G/G	147	1,702	175	172	9.8
2,521	A/G	G/G	59	1290	45	38	28.5
2,522	A/G	A/G	478	382	110	87	4.4
2,523	A/G	G/G	36	1,234	49	62	19.9
2,524	A/G	G/G	62	1,492	59	115	13

[a]A base-call is made if the SNR is at least 3, and the ratio of the higher allele signal to the lower allele signal is >6:1 for homozygous calls and <2:1 for heterozygous calls. [b]Signals are normalized fluorescence intensity values produced by the microarray scanner. [c]The SNR was calculated for each marker by dividing the maximum allele signal (A and G) by the maximum signal in the two background channels (C and T).

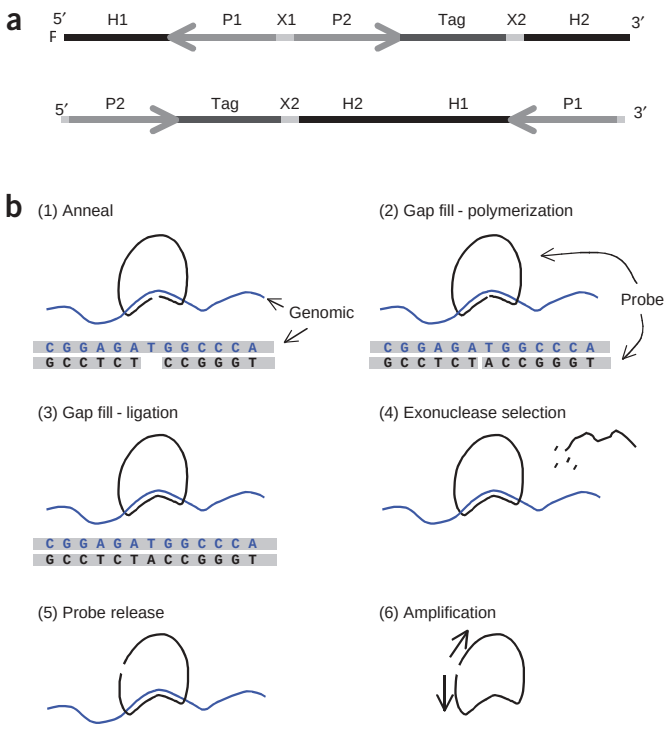

Figure 2 Molecular inversion probes. (**a**) Unreacted probe (top) and inverted probe (bottom). A single probe is used to detect both alleles of each SNP and consists of seven segments: two regions of homology to target genomic DNA, H1 and H2 (unique to each probe) at the termini of the probe, two PCR primer regions common to all probes, one bar code specific for each locus and two common cleavage sites, X1 and X2. Successfully reacted probes are amplified using primers P1 and P2. A universal detection tag sequence, one of 16,000, is for array detection of amplified probe. Cleavage sites X1 and X2 are used to release the circularized probe from genomic DNA and for post-amplification processing, respectively. (**b**) Enzymatic probe inversion. (1) A mixture of Genomic DNA, 1,000 or more probes, and thermostable ligase and polymerase is heat-denatured and brought to annealing temperature. Two sequences targeting each terminus of the probe hybridize to complementary sites in the genome, creating a circular conformation with a single-nucleotide gap between the termini of the probe. (2) Unlabeled dATP, dCTP, dGTP or dTTP, respectively, is added to each of the four reactions. In reactions where the added nucleotide is complementary to the single-base gap, DNA polymerase adds the nucleotide and (3) DNA ligase closes the gap to form a covalently closed circular molecule that encircles the genomic strand to which it is hybridized. (4) Exonucleases are added to digest linear probes in reactions where the added nucleotide was not complementary to the gap and excess linear probe in reactions where circular molecules were formed. The reactions are then heated to inactivate the exonucleases. (5) To release probes from genomic DNA, uracil-*N*-glycosylase is added to depurinate the uracil residues in the probes. The mixture is then heated to cleave the molecule at the abasic site and release it from genomic DNA. (6) PCR reagents are added, including a primer pair common to all probes. The reactions are then subjected to thermal cycling, with the result that only probes circularized in the allele-specific gap-fill reaction are amplified.

accuracy of 97.8%. The accuracy of the 359 agreeing sequence pairs was therefore 99.95%. In the genotyping data set, 312 full genotypes and 23 half genotypes were identical with the Sanger sequence pairs (643/647 chromosome comparisons), for an accuracy of 99.4%. Similar accuracy was achieved when the data were compared with pyrosequencing[18,19] data on different sets of SNPs. The four discordant genotypes were found in a single locus (probe 2101) in four individuals. Because the sequencing data were nonpolymorphic and subsequent sequencing of these loci with newly generated amplicons was concordant with our genotyping results, it is likely that the original amplification for Sanger sequencing was contaminated with other template DNA or PCR product from another individual.

To measure the reproducibility of MIP genotyping, an individual was assayed three times with 888 markers (5,328 chromosomes). 5,006 of 5,011 chromosome comparisons were concordant (99.9%) (**Table 2**).

We investigated the effect of increasing the multiplexing level tenfold. The performance of 75 probes either in a 75-probe multiplexed reaction or embedded in a 938-probe multiplex reaction was compared on the same individual's DNA (**Table 3**). The average call rate in seven repetitions of the same individual for the 75-probe multiplex was 92.6%. Call rate for the same 75 probes in the 938-probe multiplex was 93.4% (average of 25 individuals). The assay conditions were identical in every respect except the number of probes added.

Because DNA array costs represent a substantial fraction of the overall cost of this method, we compared four-chip–one-color detection to two-chip–two-color detection in otherwise identical experiments. The dyes were carboxyfluorescein directly coupled to the labeling oligonucleotide, and phycoerythrin that was coupled to the labeling oligonucleotide via biotin-streptavidin in post–chip hybridization staining[20]. Call rate and SNR in the two-chip–two-color experiment (96.1% and 30, respectively) were very similar to those in the four-chip–one-color experiment (95.8% and 31).

Table 2 Performance metrics of the genotyping assay by molecular inversion probes

Measured parameter	Result
Design rate[a]	84%
Call Rate[b]	95%
Concordance with Sanger sequencing[c]	>99.4%
Repeatability[d]	99.9%
Highest multiplex level	1,517
Average SNR[e]	17
Genomic DNA used / SNP[f]	2 ng

[a]183 of 1,121 probes failed to generate data. [b]An average of 891 of 938 probes called per individual for 25 experiments. [c]Two of 396 chromosomes were discordant with pyrosequencing. [d]5,006 of 5,011 chromosome comparisons were concordant. [e]Average of the ratio of maximum allele signal to maximum non-allele signal of called probes. [f]2 μg genomic DNA used to genotype 1,121 markers per individual.

DISCUSSION

The MIP genotyping method described here has several advantages over alternative techniques. No singleplex PCR amplification is required before mutation detection, thereby reducing labor and expense. PCR is applied only after mutation detection, at which time all molecular inversion probes are converted to standard-length oligonucleotides of similar sequence composition and common primers. This results in a high degree of multiplexing capacity. We have not observed any change in performance in multiplexing from a single probe up to 1,500 probes and speculate that a further increase to 10,000 probes might be possible because sufficient signal is generated in the assay to support that many probes. The data

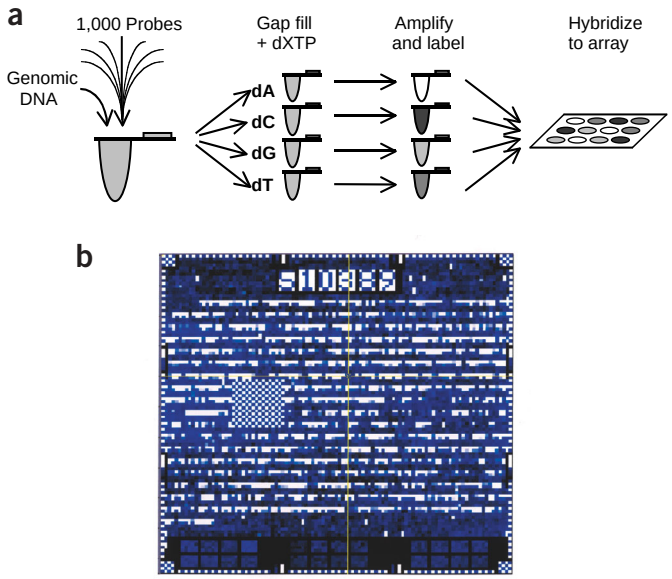

Figure 3 Process flow and array image. (**a**) Genotyping process flow. 1,000 or more probes are mixed with genomic DNA and gap-fill enzymes (see **Fig. 2**). The reaction is split into four tubes and one of four unmodified nucleotides is added. Reactions are subsequently amplified and a label is added. Reactions are combined and hybridized to the microarray. Relative intensities of two expected allele bases and two background bases indicate genotype and probe performance. (**b**) Data from 938 amplified probes hybridized to a GenFlex universal DNA array. The relative base incorporation is measured by the fluorescence signals at the corresponding complementary tag site on the DNA array.

presented here were generated using four microarrays per sample. Currently we use two microarrays with two-color detection per sample as previously described[20], and we obtain equivalent call rates and SNRs. In theory, genotyping 16,000 markers with this method would require 44 reactions and 2 oligonucleotide arrays (1,500-probe multiplex with 16,000 element Affymetrix Tag 3 array using two-color detection). Thus only a very modest infrastructure is needed to use this approach: a small number of thermocyclers, microarray washing instruments and microarray scanners. This compares very favorably with the robotic infrastructure and detection instrumentation required to set up thousands of PCR reactions and analyze the results.

The intramolecular nature of the MIP genotyping allows higher multiplexing than any other current approach because only the self-self interacting molecules are amplified, while cross-interactions are greatly suppressed. This should allow the current DNA usage of 2 ng per SNP reaction (2 μg/1,000 probes) to be further reduced to 0.2 ng per SNP reaction (2 μg/10,000 probes) as the degree of multiplexing is increased to 10,000 probes.

Several levels of intrinsic specificity are built into this assay. First, the dual recognition sequences at the 3′ and 5′ ends of probes are physically constrained to interact locally. A molecular inversion probe hybridizes much more quickly than two independent probes because the second recognition sequence hybridizes instantaneously after the first. As a result, probe-genomic complexes form at probe concentrations that do not favor nonspecific cross-interactions between probes. Specificity is then increased by the action of the gap-fill enzymes. DNA polymerase selectively extends the correct nucleotide, and DNA ligase ligates only perfectly hybridized DNA. An error requires both misextension and misligation to occur. Probes that have undergone the correct interaction and circle formation are further selected by exonuclease treatment before amplification. Finally, the tag sequences are selected to achieve high hybridization specificity and thereby eliminate cross-talk at the detection step. The synergism of the individually optimized steps comprising the MIP genotyping results in the high degree of multiplexing described here.

An unusual aspect of the approach is the built-in quality control of SNR through monitoring of the background allele channels. Biallelic markers such as SNPs have only two possible base alleles. Because this assay monitors all four base possibilities, the SNR is measured with each call and suspect calls can be efficiently discarded.

Molecular inversion requires a single probe per marker, reducing the requirement for probe synthesis. Moreover, any damage or loss of performance of that probe will affect both alleles equally and will therefore not lead to spurious genotypes such as can occur with allele-specific oligonucleotides.

For molecular inversion technology, the rate at which a functional probe is generated from an SNP chosen at random from a database in a single synthesis attempt is 84%. The rate at which all functional probes produce high-quality data over many individuals is 95%. As mapping and cSNP (SNPs that are found in exons) discovery efforts proceed, it will be increasingly important to assay a particular SNP rather than any SNP within a region. This will place increasing emphasis on the ability of a given technology to assay any SNP.

Cost is a fundamental driver for the development of alternative SNP genotyping technologies. There are three main costs associated with SNP genotyping: probe cost, assay cost and detection cost. Although molecular inversion probes are longer than PCR primers, the total number of unique bases that must be synthesized for each probe is comparable to that for a PCR-based approach and much lower than for methods that require allele-specific oligonucleotides, such as the oligonucleotide ligation assay[21]. The locus-specific probes do not require any fluorescent or modified bases and are

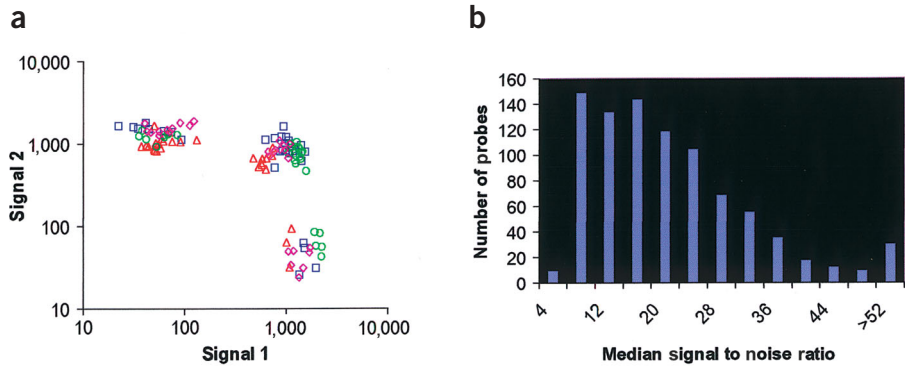

Figure 4 Assay performance. (**a**) Fluorescence signal from four markers tested on 25 individuals in 32 experiments in a 1,121-probe multiplex assay. Markers 1, 2 and 3 are A/G alleles and marker 4 is a C/G allele. A and C signal is plotted as signal 1 and G signal as signal 2. (**b**) Median ratio of maximum allele signal to maximum background (non-allele) signal for 938 probes.

Table 3 Data on allele type, median call rate and SNR obtained by large-scale genotyping

Allele	No. tested	Average call rate	Median SNR
A/C	77	94%	19
A/G	349	95%	18
A/T	52	94%	22
C/G	90	94%	19
C/T	302	95%	22
G/T	68	94%	21

therefore inexpensive to synthesize. Also, owing to the kinetic advantage of intramolecular interactions, only 12 amol of each probe are used in a single assay. Typical synthesis scales of 1 nmol thus represent millions of assays worth of material. These probes will thus persist as a valuable resource for subsequent genotyping. Assay costs are amortized by the high degree of multiplexing involved, resulting in a very inexpensive assay in the current format (<$0.01 per assay). Detection costs are dominated by the cost of the microarray. To minimize this cost, arrays should be fully occupied and larger arrays used to amortize the cost of the arrays over more genotypes. Such arrays are now available from multiple suppliers either on or off the shelf (GenFlex and Tag3 arrays; Affymetrix) or on a service basis (Agilent, Amersham, NimbleGen), and the cost per array feature ranges from less than $0.01 to $0.10. Taken together, we believe these advantages amount to an enabling advance in genotyping technology that will allow the power of whole-genome analysis to be realized.

METHODS

Oligonucleotide synthesis. Example of a probe constructed for NCBI SNPcluster ID: rs 1389629 (TSC0559094) 5′-TGGATCCCATTATCCTCCATTACGGCTCAACGTTCCTATTCGGTTUUUTTGCAAATGTTATCGAGGTCCGGCACGCACAGGTTATGAATCTCTTTAAACTCCCACAGTGAGGAGC-3′. Molecular inversion probes were ordered from ParAllele BioScience and Eurogentec. Other oligonucleotides were synthesized by Operon, Inc.

Samples. All samples were obtained from the Coriell Institute human variation collection of African American (HD100AA) and Caucasian (HD100CAU) genomic DNA.

Exonucleolysis assay. Circularization of 3 fmol MIP probe A (5′-TGATGGACGTCTGGAAAGCAACCGAAGCTTGTGCGCGTGCAGGGTCACCAGCAGGCATGAGCCCGGTCAACTTCAAGCTCCTAAGCCGGCAGGCAATGCACAGCACCG-3′) was done using as template 6 fmol of the oligonucleotide target 5′-TGCTTTCCAGACGTCCATCACGGTGCTGTGCATTGCCTGC-3′, in 20 mM Tris-HCl pH 8.5, 100 mM KCl, 9 mM $MgCl_2$, 1 mM dithiothreitol, 0.1% Triton X-100 and 4 units Ampligase (Epicentre). Reactions were thermally cycled ten times between 95 °C for 30 s and 58 °C for 1 min. Exonucleolysis was performed by addition of 8 units each of exonuclease I and exonuclease III (New England Biolabs) and incubation at 37 °C for 3 h, followed by denaturation at 95 °C for 10 min. Linear or circularized forms of A were amplified using primers F1 (5′-TGATGGACGTCTGGAAAGCAA-3′) and R1 (5′-CGGTGCTGTGCATTGCCTGCC-3′), or F2 (5′-CACGCGCACAAGCTTCGG-3′) and R2 (5′-CAGGGCACCAGCAGGCA-3′), respectively. Dimerization of the two padlock probes A (sequence above) and B (5′-TGTTCACTGCTGGCCTCCGCAAGCGTGTAGTGTCCGTCGAATAT-TCGTTCTGCAGCATCGCACACAGAAGGTCGATTGCTAGGTGACTGCCACCCAAGGGG-3′) was done using the template oligonucleotide 5′-GCGGAGGCCAGCAGTGAACACGGTGCTGCATTGCCTGC-3′, and the ligation product was amplified with primers F2 and R2. Real-time amplification was done with an ABI Prism 7700 thermocycler (Applied Biosystems) in a total volume of 30 µl containing 2.5 µl buffer A (Applied Biosystems), 2 mM $MgCl_2$, 0.2 mM of each dNTP, 0.3 µM of each primer, 0.8 µM TaqMan probe 5′-FAM-CCCGGTCAACTTCAAGCTCCTAAGCC-TAMRA-3′ and 0.6 units AmpliTaq Gold (Applied Biosystems). The temperature profile was 95 °C for 10 min followed by 50 cycles of denaturation at 95 °C for 30 s and annealing and extension at 58 °C for 45 s.

Molecular inversion probe assay. Denaturing and annealing: four identical reactions containing 400 ng of genomic DNA, 12 amol each of 1,000 probes, 0.0625 units Ampligase (Epicentre) and 0.5 units Stoffel fragment DNA polymerase (Applied Biosystems) in 9 µl of 20 mM Tris-HCl (pH 8.3), 25 mM KCl, 10 mM $MgCl_2$, 0.5 mM NAD and 0.01 % Triton X-100 were incubated for 4 min at 20 °C, 5 min at 95 °C and 15 min at 60 °C. Gap-fill reaction: 1 µl of each of four nucleotides was added to the four reactions and incubated for 10 min at 60 °C and then 1 min at 37 °C. Exonuclease selection: 10 units exonuclease I and 200 units exonuclease III (United States Biochemical) in a 2-µl volume were added and the mixture incubated for 14 min at 37 °C, 2 min at 95 °C and 1 min at 37 °C. Uracil depurination and cleavage: 2 units of uracil-*N*-glycosylase (New England Biolabs) were added in 25 µl of 1.6 mM $MgCl_2$, 10 mM Tris-HCl (pH 8.3), 50 mM KCl and incubated for 9 min at 37 °C and 20 min at 95 °C. Amplification: 2 units of AmpliTaq Gold (Applied Biosystems), 16 pmol primer 1 (5′-CCGAATAGGAACGTTGAGCCGT-3′), and 16 pmol primer 2 (5′-GCAAATGTTATCGAGGTCCGGC-3′) in 25 µl of 1.6 mM $MgCl_2$, 10 mM Tris-HCl (pH 8.3), 50 mM KCl and 112 µm dNTP were preactivated for 10 min at 95 °C and then added to the genotyping reactions. The reactions were amplified in 28 cycles of 95 °C for 20 s, 65 °C for 45 s and 72 °C for 10 s. Sample processing: 20 units of exonuclease I and 10 units *Dra*I (New England Biolabs) were incubated with 60 µl of each amplification product at 37 °C for 1 h and then 80 °C for 30 min.

Microarray hybridization. Approximately 1.25 pmol of each amplified and processed product were hybridized overnight at 39 °C to a GenFlex Tag Array (Affymetrix) DNA array with 55 µl 2× MES, 2.2 µl 50× Denhardt buffer, 110 pmol primer 1 complement (5′-ACGGCTCAACGTTCCTATTCGG-3′), 110 pmol primer 2-FAM (5′-FAM-GCAAATGTTATCGAGGTCCGGC-3′), 1.1 fmol (each) GenFlex control oligonucleotide (Affymetrix). Microarrays were washed and scanned as recommended by the manufacturer (Affymetrix). Data analysis was performed on the raw signal data for each tag feature generated by the Affymetrix image analysis software.

Data analysis. Four signals were generated for each genotype, one for each base reaction. The raw signal was background subtracted and then normalized using the GenFlex hybridization controls that were hybridized in equimolar amounts to each microarray. Base-calls were performed as follows: a small systematic noise was added to all four channels (A,C,G,T). Measured signals were adjusted as follows: $S_{adjusted} = \sqrt{(S_{measured}^2 + (7 \times \sigma)^2)}$, where σ = 3 times the standard deviation of features that were not used in a particular experiment (array noise). The SNR was calculated for each marker by dividing the maximum signal in the two expected allele channels by the maximum signal in the two background channels. The two criteria for calling genotypes were a signal to background ratio of at least 3:1, and the ratio of the higher allele signal to the lower allele signal greater than 6:1 for homozygous calls and less than 2:1 for heterozygous calls.

ACKNOWLEDGEMENTS

We thank K. Tran, T. Brundage, J. Patterson, E. Prakash and C. Bruckner for technical assistance and T. Jones and P. Oefner for helpful discussions. This work was supported by NIH HG00205. The work in Uppsala was supported by the Beijer and Wallenberg Foundations, the Research Councils of Sweden for natural science and for medicine, the Swedish Cancer Fund and Polysaccharide Research AB (Uppsala).

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Biotechnology* website for details).

Received 18 October 2002; accepted 21 February 2003
Published online 5 May 2003; doi:10.1038/nbt821

1. Patil, N. *et al.* Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294**, 1719–1723 (2001).
2. Syvanen, A.C. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat. Rev. Genet.* **2**, 930–942 (2001).
3. Nilsson, M. *et al.* Making ends meet in genetic analysis using padlock probes. *Hum. Mutat.* **19**, 410–415 (2002).
4. Elnifro, E.M., Ashshi, A.M., Cooper, R.J. & Klapper, P.E. Multiplex PCR: optimization and application in diagnostic virology. *Clin. Microbiol. Rev.* **13**, 559–570 (2000).
5. Lockhart, D.J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**, 1675–1680 (1996).
6. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
7. Lyamichev, V. *et al.* Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat. Biotechnol.* **17**, 292–296 (1999).
8. Eis, P.S. *et al.* An invasive cleavage assay for direct quantitation of specific RNAs. *Nat. Biotechnol.* **19**, 673–676 (2001).
9. Nilsson, M. *et al.* Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* **265**, 2085–2088 (1994).
10. Lizardi, P.M. *et al.* Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* **19**, 225–232 (1998).
11. Faruqi, A.F. *et al.* High-throughput genotyping of single nucleotide polymorphisms with rolling circle amplification. *BMC Genomics* **2**, 4 (2001).
12. Antson, D.O., Isaksson, A., Landegren, U. & Nilsson, M. PCR-generated padlock probes detect single nucleotide variation in genomic DNA. *Nucleic Acids Res.* **28**, E58 (2000).
13. Shoemaker, D.D., Lashkari, D.A., Morris, D., Mittmann, M. & Davis, R.W. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nat. Genet.* **14**, 450–456 (1996).
14. Zhang, D.Y., Brandwein, M., Hsuih, T.C. & Li, H. Amplification of target-specific, ligation-dependent circular probe. *Gene* **211**, 277–285 (1998).
15. Abravaya, K., Carrino, J.J., Muldoon, S. & Lee, H.H. Detection of point mutations with a modified ligase chain reaction (Gap-LCR). *Nucleic Acids Res.* **23**, 675–682 (1995).
16. Banér, J., Nilsson, M., Mendel-Hartvig, M. & Landegren, U. Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Res.* **26**, 5073–5078 (1998).
17. Gharavi, A.G. *et al.* IgA nephropathy, the most common cause of glomerulonephritis, is linked to 6q22-23. *Nat. Genet.* **26**, 354–357 (2000).
18. Fakhrai-Rad, H., Pourmand, N. & Ronaghi, M. Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. *Hum. Mutat.* **19**, 479–485 (2002).
19. Ronaghi, M. Pyrosequencing sheds light on DNA sequencing *Genome Res.* **11**, 3–11 (2001).
20. Fan, J.B. *et al.* Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res.* **10**, 853–860 (2000).
21. Landegren, U., Kaiser, R., Sanders, J. & Hood, L. A ligase-mediated gene detection technique. *Science* **241**, 1077–1080 (1988).