Gold nanoparticles for high-throughput genotyping of long-range haplotypes

Peng Chen1, Dun Pan2, Chunhai Fan1,2,*, Jianhua Chen1, Ke Huang1, Dongfang Wang2, Honglu Zhang2, You Li1, Guoyin Feng1, Peiji Liang3, Lin He1,4,5 and Yongyong Shi1,6*

Completion of the Human Genome Project1 and the HapMap Project2 has led to increasing demands for mapping complex traits in humans to understand the etiology of diseases3. Identifying variations in the DNA sequence, which affect how we develop disease and respond to pathogens and drugs, is important for this purpose, but it is difficult to identify these variations in large sample sets4–5. Here we show that through a combination of capillary sequencing and polymerase chain reaction assisted by gold nanoparticles, it is possible to identify several DNA variations that are associated with age-related macular degeneration6–8 and psoriasis9 on significant regions of human genomic DNA. Our method is accurate and promising for large-scale and high-throughput genetic analysis of susceptibility towards disease and drug resistance10–12.

Alleles, which are one of two or more versions of a gene located at specific positions (or loci) on a chromosome, determine distinct traits that are passed on from parents to their offspring. A single-nucleotide polymorphism (SNP), which is a single-base-pair mutation, occurs at a specific locus that usually consists of two alleles. Because SNPs are one of the most common types of genetic variation, measurement of SNPs between members of a species (or SNP genotyping) has been extensively used in studies of many human diseases. More recently, there has been increasing interest in determining sets of statistically associated SNPs on a single chromatid (known as haplotypes) to study causal variants identified in genome-wide association studies. Allele-specific polymerase chain reaction (AS-PCR), which selectively amplifies one of the alleles from a template13, is a common technique for SNP genotyping and haplotyping.14 However, conventional AS-PCR is not sufficient for genetic studies because it cannot effectively discriminate DNA mismatches in all situations. Because gold nanoparticles (AuNPs) have a unique combination of physical and chemical properties15,16, and have been successfully used to facilitate PCR17–19 and gene delivery20 processes by exploiting rich interactions between AuNPs and DNA molecules17,21,22, we studied the effects of AuNPs on the ability of AS-PCR to distinguish base mismatches for improved genotyping and haplotyping (see the schematic diagram in Fig. 1).

In a typical AS-PCR, there are 16 possible combinations of base pairing between the 3′-end primer and the template: 4 fully complementary and 12 one-base mismatched pairs. We first examined all these primer–template pairs by using lambda DNA with AuNPs-enhanced AS-PCR. PCR reactions were monitored in real time and the Ct value—cycle number at which the amount of PCR product reaches a threshold—was measured to quantify the efficiency of amplification. We found that the addition of an optimized concentration of 1 nM AuNPs (5 nm in diameter) significantly increased Ct values for all the mismatched pairs (Fig. 2a; see also Supplementary Fig. S1 for variations of Ct and Supplementary Table S1 for raw data). In contrast, the fully complementary primer–template pairs were minimally affected by AuNPs (Fig. 2b and Supplementary Table S1). Therefore, AuNPs can selectively inhibit the amplification of mismatched primer–template pairs and enhance the specificity of AS-PCR. Although the mechanism remains unclear17–19,23, this enhancement effect was attributed to the preferential binding of AuNPs to single-stranded DNA (for example, primer) over double-stranded DNA (for example, primer–template pair), a property resembling that of the single-stranded binding protein17. The difference between the Ct values of fully complementary and mismatched pairs (ΔCt values) increased for all examined AS-PCR, with the increased values dependent on the type of mismatches (see Supplementary Table S2 for ΔCt values). Parallel studies in human genome DNA confirmed that the AuNPs-enhanced specificity of AS-PCR is a generic effect (see Fig. 2c for four primer–template pairs; see also Supplementary Fig. S2 for the other 12 pairs, Supplementary Table S3 for raw data and Supplementary Table S4 for ΔCt values).

Four randomly selected SNP loci (rs7758706, rs3763197, rs4709716 and rs1059004), in a set of 24 human genomic DNA samples, were then used to further explain the enhancement effect of AuNPs. By taking rs7758706 with genotype A/A (allele A or G) as an example, although the mismatched primer for allele G led to delayed amplification (increased Ct) in AS-PCR (Fig. 2d), the presence of AuNPs selectively inhibited amplification of the mismatched primer with essentially no fluorescence increase (Fig. 2e). Parallel electrophoretic studies confirmed that the fully complementary primer produced a prominent amplification product band whereas the mismatched one did not (Fig. 2f). The 24 samples were sequenced to validate the accuracy, which showed 100% consistency with AuNPs-enhanced AS-PCR (data not shown). Analysis for the other three SNP loci displayed similarly increased SNP genotyping ability compared with conventional AS-PCR (Supplementary Fig. S3).

The ability of AuNPs to selectively inhibit the amplification of base mismatches is beneficial for the amplification of long haploid targets. A haploid contains the haplotype of a set of SNPs, which is very important for genetic studies. We amplified an ~7-kb target from rs11200630 to rs3750848 in a homozygous sample by
Figure 1 | Schematic showing the AuNPs-enhanced allele-specific sequencing (AuNAS) strategy. A heterozygous sample with two SNPs is amplified with AS-PCR in the presence and absence of AuNPs. a, The complementary (blue line) and mismatched (red line) template strands are annealed with the allele-specific primers (short purple line). For simplicity, the other strand of the template and the corresponding primer are omitted. Small polygons with four different colours represent four types of bases (A, green; T, red; C, blue; G, black). The presence of non-specific amplification in curves for mismatched primer–template (A/G) (b) and in the presence of polymerase, the templates direct primer extension following the typical Watson–Crick base pairing. Complementary primer–template pairs are extended normally (top panel of b and c). Because conventional AS-PCR (without AuNPs) has poor specificity for distinguishing single-base mismatches, primer extension for the mismatched primer–template pair passes the SNP1 site (bottom panel in b). In AuNPs-enhanced AS-PCR, primer extension is effectively terminated at the SNP1 site (bottom panel in c). Dotted black arrows represent the direction of primer extension. Dotted brown ovals represent areas intended for subsequent capillary sequencing shown in d and e. Sequencing gel maps for highlighted areas (SNP2 sites are underlined) of the heterozygous sample using conventional AS-PCR (d) and AuNPs-enhanced AS-PCR (e). Each peak is differently coloured to represent four types of bases (A, green; T, red; C, blue; G, black). The presence of non-specific amplification in d results in two peaks at the SNP2 site, which makes it difficult to identify this base (marked with a red ‘?’). In contrast, only one peak appears at this site in e, leading to unambiguous haplotyping of this heterozygous sample.

Figure 2 | Real-time amplification curves and Ct values of AS-PCR and AuNPs-enhanced AS-PCR using 5 nm AuNPs. a, b, Real-time PCR amplification curves for mismatched primer-template (A/G) (a) and complementary primer-template (A/T) (b) on lambda DNA in the presence of 0 (black), 0.5 (red) and 1 nM (blue) AuNPs. Increased Ct values in the presence of AuNPs suggest that PCR amplification of the mismatched primer-template was hampered. c, Effect of AuNPs on the Ct values for four primer-template pairs (A/A, A/G, A/C and A/T) on human genomic DNA. Increased Ct values confirm that amplification for mismatched pairs (A/A, A/G and A/C) was inhibited in the presence of AuNPs. d, e, Real-time amplification curves of AS-PCR (d) and AuNPs-enhanced AS-PCR (e) for SNP rs7758706 with genotype A/A (allele A or G). The black line represents complementary primer for allele A and the red line represents mismatched primer for allele G. The curves suggest that the presence of AuNPs increased the selectivity of amplification. f, Gel electrophoresis of PCR products from AS-PCR (lanes 1, 2) and AuNPs-enhanced AS-PCR (lanes 3, 4) using complementary primers containing T at the 3′-end (lanes 1, 3) and mismatched primers containing C at the 3′-end (lanes 2, 4) further confirms the ability of AuNPs to selectively inhibit amplification of base mismatches. Lane M is a DNA marker (DL2000).
using AuNPs-enhanced long-range AS-PCR (Fig. 3a). As shown in the gel electrophoresis, amplification yields decreased with increasing AuNP concentration (0.67, 1.33 and 2 nM) for both the mismatched and fully complementary primer-template pairs, with the mismatched pair decreasing much more rapidly than the complementary pair (Fig. 3b). Highest contrast was observed at an AuNP concentration of 1.33 nM, where complete elimination of mismatched amplification is seen with only a small reduction of amplification for the fully complementary pair.

We then amplified the 7-kb target in a heterozygous sample and evaluated the chromosome separation ability of AuNPs-enhanced AS-PCR with capillary sequencing using SNP rs2223089 as the model. Heterozygous samples contain different alleles at a particular gene locus on homologous chromosomes, whereas homozygous samples have the same alleles. In principle, there should be two sequencing peaks at the SNP locus in the capillary-sequencing samples because of the mismatched homologous chromosome were arising from the mismatched homologous chromosome were largely suppressed in AuNPs-enhanced AS-PCR, leading to well-separated allele sequencing peaks for each SNP at the optimal annealing temperature of 57 °C during amplification (Fig. 3c,f). This enhancement is due to the ability of AuNPs to inhibit non-specific amplification in PCR reactions.

On the basis of this new finding, we propose an AuNPs-enhanced allele-specific sequencing strategy. Because AuNPs-enhanced AS-PCR can amplify long fragments of diploid chromosomes, it is possible to place multiple heterozygous SNP loci in each long fragment, with at least one locus shared by the adjacent fragment (that is, ‘joint loci’). All overlapping fragments are amplified by jumping from one joint locus to the other and then sequenced for reconstruction of the whole haplotype. Because state-of-the-art PCR and next-generation sequencing technologies support high-throughput sample preparation and assays (Supplementary Scheme S1), it is convenient to run large-sized samples for disease analysis without a limit to theoretical distance using this nanotechnology-based haplotyping strategy. As a proof-of-concept, we haplotyped three long target regions in large sample sets: (1) a 34-kb region containing 15 common SNPs (Fig. 4a) in chromosome 10 for 147 samples from the Chinese Han population (74 wet age-related macular degeneration (AMD) cases and 73 normal controls); (2) an ~18-kb region containing seven common SNPs in chromosome 3 for the 147 Chinese Han population samples; and (3) an ~27-kb region containing nine common SNPs in chromosome 6 for 96 psoriasis samples and 96 controls (see data analysis in Supplementary Tables S8, S9 and S10).

Wet AMD is the aggressive form of this prevalent disorder that is most responsible for irreversible blindness among older individuals (see photographs of eye fundus caused by this disease in Supplementary Fig. S4). The 34-kb region is relevant to wet (neovascular) AMD, and connects the genes ARMS2 (age-related maculopathy susceptibility 2) and HTRA1 (high-temperature requirement factor A1) on chromosome 10q26. All genotypes of these samples were split into a subset of 12 fragments (frag1 to frag12), each containing one or two joint loci (Fig. 4b). The whole haplotype containing all 15 SNPs was reconstructed with the AuNPs-enhanced allele-specific sequencing strategy, which was
then subjected to association analysis by examining alleles of all 15 SNPs with the haplotype analysis software Haploview. We discovered 12 significant SNPs ($P < 0.05$) in our sample set, and rs11200638 (located in the promoter region and proved to affect the expression of the HTRA1 gene) and rs10490924 (non-synonymous SNP located in the ARMS2 gene) had the highest significance (Table 1), which coincides well with previous genome-wide association studies. We note that allele T is the risk allele of rs10490924 and allele A is the risk allele of rs11200638.

Studies on this region have attracted much attention recently. Dewan et al. first reported that the HTRA1 promoter polymorphism rs11200638 was a causal mutation of wet AMD based on genome-wide association studies. However, two groups later identified that AMD was instead strongly associated with SNP rs10490924 based on genome-wide association studies. We were able to identify the most probable causal variant by testing the significance of SNPs in haplotypes excluding those we were able to identify. Because we obtained the genuine molecular haplotypes, we were able to identify the most probable causal variant by testing the significance of SNPs in haplotypes excluding those we were able to identify.

One of the possible explanations is that both ARMS2 and HTRA1 contribute to the risk of AMD, because functional variants of them are independently associated with the disease. Nevertheless, further studies using animal model analysis are needed to eventually identify the exclusive causal variant. Our finding demonstrates the challenges of identifying causal alleles for complex genetic traits and the significance of molecular haplotyping in such analysis.

Software-based statistical methods (for example, PHASE, which is a software for haplotype reconstruction) are commonly used in haplotyping, particularly when molecular haplotyping is not available. However, previous studies have suggested that inference from these statistical methods might result in haplotype misclassification, particularly when linkage disequilibrium is not strong. Indeed, the PHASE software analysis of 10q26 led to relatively low accuracy (84.4%, Supplementary Table S8). Therefore, although PHASE remains a convenient and cost-effective tool for genetic analysis, molecular haplotyping is important for obtaining
unbiased haplotyping information and providing direct evidence to study the causal variant of diseases.

In conclusion, our AuNPs-enhanced allele-specific sequencing strategy relies simply on PCR and capillary sequencing, which provides a straightforward approach for molecular haplotyping in routine laboratories. Given that only minute amounts of AuNPs are required in these assays and that AuNPs are inexpensive and readily available, current laboratory instrument and assay protocols can be easily adapted for unambiguous molecular haplotyping in the long range. This method is scalable and compares favourably with traditional labour-intensive cloning-sequencing methods in both speed and throughput. Comparison studies have clearly identified the advantage of AuNPs-enhanced allele-specific sequencing for the discovery of causal variants. More recently, emerging technologies based on microfluidic devices or next-generation sequencing have held great promise for whole-genome haplotyping29,30. Despite the elegance of these advanced haplotyping technologies, their ability to analyse large-sized samples of complex diseases is still limited by cost and complexity. We expect that AuNPs-enhanced allele-specific sequencing will become an important laboratory tool for performing large-scale association studies to provide an in-depth understanding of complex genetic diseases.

Methods

Gold nanoparticles (5 nm) were purchased from Sigma-Aldrich. All primers were synthesized in Invitrogen, and the sequences are shown in Supplementary Tables S11 and S12. Lambda DNA, Taq, LA Taq polymerase and related reagents were from TaKaRa. Sequencing-related reagents and SYBR Green I were from ABI.

Quantitative real-time PCR was performed on an ABI 7900 instrument. All the resulting C\textsubscript{T} values were calculated using SDS software (version 2.2, ABI). PCR reactions were initiated from a hot-start at 95 °C for 10 min, and each PCR cycle consisted of two stages, 95 °C for 15 s and 60 °C for 60 s, with 40 cycles for lambda DNA and 50 cycles for the human genome. Dissociation curves were measured to verify the purity and length of products. The AS-PCR reaction system consisted of 0.5 μl of buffer, 0.25 μl of 20× SYBR Green I, 0.25 μl of template (1 ng μl\(^{-1}\)), 0.1 μl of primers (10 μM), 0.1 μl of Taq polymerase (5 U μl\(^{-1}\)), 0.7 μl of MgCl\(_2\) (25 mM) and 0.4 μl of dNTP (100 mM). Either AuNPs of appropriate concentrations or deionized water was added to the system to reach a final volume of 5 μl.

All the SNP loci were genotyped by sequencing on an ABI 3730. The long-range AS-PCR was initiated from 98 °C (3 min), and each PCR cycle consisted of three stages, 98 °C for 15 s, 57 °C for 30 s and 68 °C for several minutes (depending on the length of PCR product). Each PCR reaction solution (15 μl) consisted of 0.75 U of LA Taq polymerase, 1 × PCR buffer, 240 pmol dNTP each, 3 μM of each primer, 2 μl of 10 mM AuNPs and deionized water. PCR products were qualified with gel and purified by precipitation, followed by a standard sequencing protocol provided by ABI.

Received 18 April 2011; accepted 26 July 2011; published online 4 September 2011

References


Acknowledgements
This work was supported by the National 863 Program (2009AA022701), the National 973 Program (2010CB529602, 2007CB936000, 2012CB932600), the Natural Science Foundation of China (20725516, 31000553, 90913014, 20902096), the Foundation for the Author of National Excellent Doctoral Dissertation of PR China (201026), the Program for New Century Excellent Talents in University (NCET-09-0550), the Shanghai Changning Health Bureau Program (2008406082), the Shanghai Municipal Health Bureau Program (20080895), the Shanghai Leading Academic Discipline Project (B205), CAS (KJCX2-EW-N03) and the Major S&T Program (2009ZX10004-301).

Author contributions
Y.S. and C.F. conceived and designed the experiments. P.C., D.P., J.C., D.W. and H.Z. performed the experiments. J.C. and P.C. analysed the data. K.H., Y.L., P.L. and L.H. contributed materials/analysis tools. Y.S. and G.F. collected DNA samples. P.C., D.P., C.F. and Y.S. co-wrote the paper. P.C. and D.P. contributed equally to this work. All authors discussed the results and commented on the manuscript.

Additional information
The authors declare no competing financial interests. Supplementary information accompanies this paper at www.nature.com/naturenanotechnology. Reprints and permission information is available online at http://www.nature.com/reprints. Correspondence and requests for materials should be addressed to Y.S. and C.F.