

ARTICLE

Received 30 Nov 2016 | Accepted 27 Jan 2017 | Published 6 Apr 2017

DOI: 10.1038/ncomms14738

OPEN

# DNA origami-based shape IDs for single-molecule nanomechanical genotyping

Honglu Zhang<sup>1,2</sup>, Jie Chao<sup>3</sup>, Dun Pan<sup>1,2</sup>, Huajie Liu<sup>1</sup>, Yu Qiang<sup>2</sup>, Ke Liu<sup>2</sup>, Chengjun Cui<sup>1</sup>, Jianhua Chen<sup>2</sup>, Qing Huang<sup>1</sup>, Jun Hu<sup>1</sup>, Lianhui Wang<sup>3</sup>, Wei Huang<sup>3</sup>, Yongyong Shi<sup>2</sup> & Chunhai Fan<sup>1</sup>

Variations on DNA sequences profoundly affect how we develop diseases and respond to pathogens and drugs. Atomic force microscopy (AFM) provides a nanomechanical imaging approach for genetic analysis with nanometre resolution. However, unlike fluorescence imaging that has wavelength-specific fluorophores, the lack of shape-specific labels largely hampers widespread applications of AFM imaging. Here we report the development of a set of differentially shaped, highly hybridizable self-assembled DNA origami nanostructures serving as shape IDs for magnified nanomechanical imaging of single-nucleotide polymorphisms. Using these origami shape IDs, we directly genotype single molecules of human genomic DNA with an ultrahigh resolution of  $\sim 10$  nm and the multiplexing ability. Further, we determine three types of disease-associated, long-range haplotypes in samples from the Han Chinese population. Single-molecule analysis allows robust haplotyping even for samples with low labelling efficiency. We expect this generic shape ID-based nanomechanical approach to hold great potential in genetic analysis at the single-molecule level.

<sup>1</sup>Division of Physical Biology and Bioimaging Center, Shanghai Synchrotron Radiation Facility, CAS Key Laboratory of Interfacial Physics and Technology, Shanghai Institute of Applied Physics, Chinese Academy of Sciences, PO Box 800-204, Shanghai 201800, China. <sup>2</sup>Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Bio-X Institutes, Shanghai Jiao Tong University, Shanghai 200030, China.

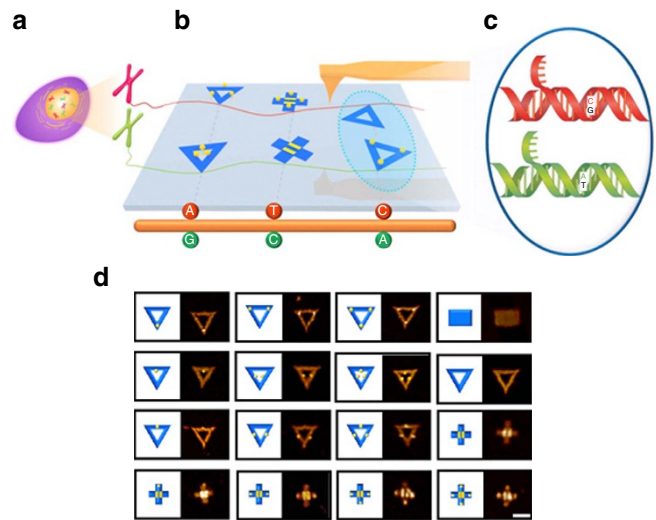
<sup>3</sup>Key Laboratory for Organic Electronics and Information Displays (KLOEID), Institute of Advanced Materials (IAM), School of Materials Science and Engineering, Nanjing University of Posts and Telecommunications, 9 Wenyuan Road, Nanjing 210046, China. Correspondence and requests for materials should be addressed to J.C. (email: iamjchao@njupt.edu.cn) or to Y.S. (email: shiyongyong@gmail.com) or to C.F. (email: fchh@sinap.ac.cn).

State-of-the-art genetic analysis is predominantly based on fluorescence imaging, which nevertheless has an optical resolution limit<sup>1–6</sup>. As an alternative yet powerful approach, atomic force microscopy (AFM)-based nanomechanical imaging is distinct in its high-resolution power in ambient conditions<sup>7</sup>. Direct reading of genetic information using AFM has long been a dream since its invention<sup>8</sup>. Although AFM in principle has much higher resolution than optical microscopy and even superresolution microscopy<sup>9</sup>, its application in genetic analysis remains to be limited. Unlike fluorescence imaging that has a spectrum of wavelength-specific fluorophores or fluorescent nanoparticles, the lack of shape-specific labels for AFM largely restricts site-specific labelling for distinctive visualization<sup>10–12</sup>. Previous studies have well documented that self-assembled DNA origami nanostructures with arbitrary shapes can be reliably fabricated by folding a long viral M13 genomic DNA with ~200 short complementary staple strands<sup>13–17</sup>. By using DNA origami as the soluble substrate, Yan and colleagues<sup>17</sup> developed nanoscale chips for hybridization detection of nucleic acids. We were thus motivated to repurpose differentially shaped origami nanostructures as unique nanomechanical shape IDs for ‘multi-colour’ labelling of genomic DNA. As a proof-of-concept, we explored high-resolution genetic phasing of haplotypes using this nanomechanical imaging method.

The human genome consists of two copies of homologous chromosomes deriving from the father and the mother, respectively. As a haplotype is the combination of alleles at multiple loci along one chromosome, the context of variations occurring on a haplotype has profound effects on the expression and regulation of genes, and even the aetiology of human diseases<sup>18,19</sup>. Especially, revealing long-range haplotype information holds great promise for identifying genetic causal variants of complex disorders<sup>20–22</sup>. However, despite rapid advances in next-generation sequencing (NGS) technologies<sup>23</sup>, along with the international collaborative efforts on constructing a haplotype map of the human genome (for example, the HapMap Project<sup>24</sup> and the 1000 Genomes Project<sup>25</sup>), our understanding of the diploid nature of the genome in most studies remains limited. Among mainstream haplotyping approaches<sup>6,21,22,26,27</sup>, imaging-based three methods are highly attractive, as their direct reading feature complements NGS in rapid assembly of short-reads of several hundred base pairs and provides feasibility and effectiveness in targeted genomic region studies. Several elegant studies have demonstrated the effectiveness of fluorescence imaging in haplotyping<sup>28,29</sup>. Nevertheless, the diffraction limit of 200–300 nm restricted the resolution to be of ~1,000 bp. By taking genetic phasing as a test bed, we herein demonstrated that the shape ID-based nanomechanical imaging provided a powerful method for haplotyping with remarkably improved resolution and at the single-molecule level.

## Results

**Design and fabrication of DNA origami-based shape IDs.** We fabricated a set of shape IDs by using DNA origami designs (Figs 1 and 2, and Supplementary Fig. 1). The basic elements are triangular, cross and rectangular shapes, which are readily distinguishable under AFM imaging. To establish that this shape ID system can specifically target gene sequences, we employed a single-stranded (ss-) bacteriophage phiX 174 DNA with a covalently closed circularity genome of 5,386 nucleotides as the test-bed for genetic analysis. The phiX 174 template was first annealed with a three-block ‘mediator’ DNA strand (M-strand) that has an M1 block for complementary hybridization with the template, an M2 spacer block and an M3 block for capturing shape IDs

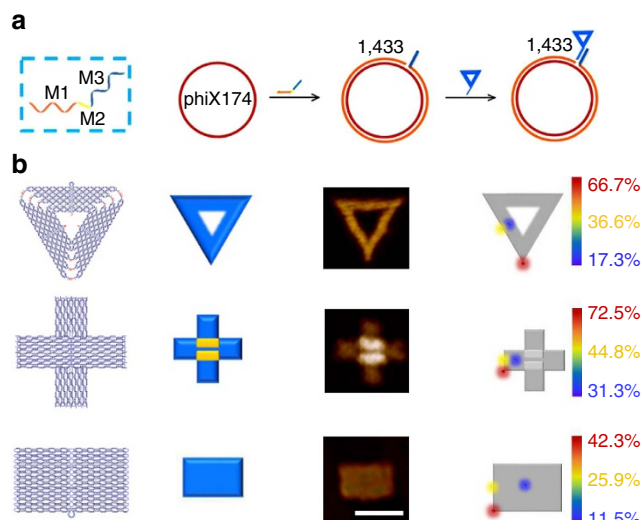


**Figure 1 | Schematic illustration of AFM-based single-molecule nanomechanical haplotyping with DNA origami shape IDs.** (a,b) Diploid genomic DNA extracted from genetic samples, which was site-specifically labeled with ‘multi-colour’ shape IDs. The two alleles of each SNP are labeled with different shape IDs. (c) Origami shape IDs serve as ‘magnifying lens’ to translate individual SNPs to origami nanostructures of several tens of nanometers. Consequently, the haplotype of this genomic DNA can be directly imaged under AFM. (d) Design and AFM images of sixteen shape IDs using DNA origami decorated with or without STV. Scale bar, 100 nm.

(Fig. 2a). Upon hybridization with the template, M1 serves as the primer to initiating DNA extension in the presence of polymerase, which turns the ssDNA template into double-stranded (ds) DNA that is more visible under AFM imaging<sup>8</sup>. The M3 block is complementary to a short strand M3’ that is carried on each corresponding shape ID.

**Genotyping using shape IDs.** To test the hybridizability of shape IDs, which is critically important for their targeting efficiency, we interrogated both the shape effect of origami and the position effect of M3’ on the labelling efficiency of shape IDs on the site 1,433 of phiX 174 (that is, hybridization efficiency between M3’ and M3). As a general trend, the labelling efficiency using the triangular and the cross origami is much higher than that using the rectangular one. The labelling efficiency is also strongly dependent on the M3’ position, which decreases in the order of corner, edge middle and inner space (Fig. 2b). Hence, both the shape of the origami and the position of M3’ on the origami greatly affect the hybridizability of shape IDs. Remarkably, we found corner positions were ‘hotspots’ for on-origami hybridization (with the efficiency of 72.5% and 66.7% for cross and triangular, respectively), which coincides with the fact that these positions have minimal steric effects, and are least perturbed by the electrostatic force field imposed by hundreds of negatively charged DNA strands in an origami (Supplementary Figs 2–10). Further studies revealed that placing multiple M3’ strands on corners of a single origami (three for triangular and four for cross) led to even greater labelling efficiency up to 85%, an effect arising from increased M3–M3’ inter-strand collision probability (Supplementary Figs 11 and 12, and Supplementary Table 1). Hence, we employed highly hybridizable triangular and cross shape IDs with multiple M3’ for subsequent genetic analysis.

Having substantiated the specific gene-targeting ability of individual shape IDs, we next explored differential labelling of a single template using multiple shape IDs. We first expanded the three-element set of shapes by using streptavidin



**Figure 2 | Origami shape IDs for 'multi-colour' nanomechanical imaging.**

(a) Schematic showing of the three-block 'mediator' DNA strand (M-strand) in labelling of phiX 174 with shape IDs. Left, M-strand is shown with three function blocks. M1 block (orange-coloured region) for site-specific hybridization and extension with the template, M2 spacer block (yellow-coloured region) and a M3 block (blue-coloured region) for capturing shape IDs. Right, the phiX 174 template is first annealed with the M-strand on the site 1,433, then M1 block serves as allele-specific primer to initiating DNA extension in the presence of polymerase, which turns the ssDNA template into dsDNA that is more visible under AFM imaging. M3 block is complementary to a short strand M3' on each corresponding shape ID. (b) Schematic and AFM images with the labeling efficiency of individual shape IDs. Left to right: design (column 1), schematic (column 2) and AFM images (column 3) of triangular-, cross- and rectangular- shaped DNA origami IDs. Scale bar, 100 nm. Position effects on the hybridization efficiency of shape IDs on the site 1,433 of phiX174 (column 4, see also in Supplementary Figs 2-10). The efficiency is the highest at the point position ('hotspot', marked in red). The edge middle and inner sites are marked in yellow and blue, respectively.

(STV)-decorated origami nanostructures (Supplementary Fig. 13a). By exploiting the addressability of DNA origami, we site-specifically anchored different numbers of STV on prescribed positions carrying biotin tags. STV of  $\sim 4$  nm is readily visible as small dots on the origami, which remarkably distinguishes STV-decorated shape IDs from non-decorated ones. Three different shape IDs, triangular, triangular with STV and cross were then chosen to target the sites of 1,433, 1,529 and 4,914 on phiX 174, respectively. We employed multiple 'mediator' DNA strands to introduce the specific M3 capturing sequences. We found that multiple M1 sequences could initiate simultaneous extension from different sites of phiX 174 in the presence of polymerase, which turned the template into a dsDNA circle similar to the single M1 initiation, albeit with a lower efficiency (Supplementary Figs 13b and 16-20 and Discussion). Importantly, multiple shape IDs can be site-specifically labelled on phiX 174, as illustrated in AFM images (Fig. 3a and Supplementary Figs 18 and 19). We also note that contour lengths between adjacent sites coincide well with those predicated from their base numbers (0.34 nm for one base) for labels at either two or three sites (Supplementary Figs 14 and 15). Of note, whereas the labelling efficiency decreases from single site ( $\sim 85\%$ ) to two sites ( $\sim 56.4\%$ ) and three sites ( $\sim 30.8\%$ ), the single-molecule nature of this nanomechanical analysis allows reliable genotyping even at a low efficiency of 30.8% (Fig. 3a and Supplementary Fig. 20). Analogous to 'multi-colour' fluorescence imaging-based genetic analysis,

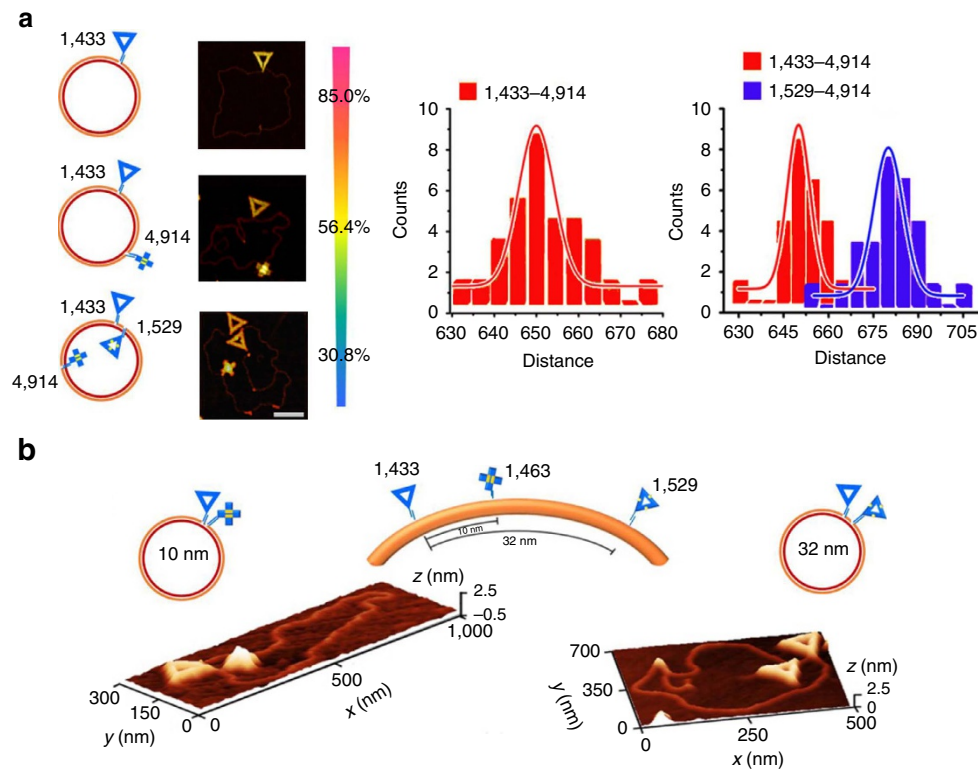
differential labelling of multiple shape IDs should greatly improve the genotyping ability in both increasing the accuracy and reducing the labour and time cost. Moreover, accurate distance information between single-nucleotide polymorphism (SNP) sites can be readily obtained from single-molecule-based statistical analysis (Fig. 3a).

We next explored the genotyping resolution of this shape ID-based single-molecule analysis. To approach the sub-100 bp resolution limit, we chose three sites (1,433, 1,463 and 1,529) on phiX 174 for labelling of shape IDs, which are separated by 30 bp ( $\sim 10$  nm) and 96 bp ( $\sim 30$  nm), respectively. Although DNA origami is  $\sim 100$  nm<sup>2</sup> in size, it is easy to resolve its centre and realize nanometre resolution given its rigid structure. Indeed, AFM imaging revealed that these sites could be readily resolved by labelled shape IDs (Fig. 3b). We note that this 10 nm resolution is the highest in imaging-based haplotyping.

We further performed genotyping of SNPs, which forms the basis for many types of genetic analysis. Our previous studies showed that gold nanoparticles (AuNPs) could effectively inhibit mismatched binding<sup>30,31</sup>. Hence, we employed AuNPs to assist discrimination of single-base mismatches in primer extension. To mimic polymorphic sites, we designed two primers with or without a single-base mismatch site at the 3'-end (Fig. 4a). Gel electrophoresis revealed that only the fully complementary primer could initiate primer extension to form the dsDNA circle, whereas the extension of single-base mismatched one was completely inhibited in the presence of AuNPs (Fig. 4b and Supplementary Fig. 23). To substantiate the specificity in AFM imaging, we chose the site 1,433 on phiX 174 to test two primers with or without a single-base mismatch, which correspond to a triangular and a cross shape ID, respectively (Fig. 4c). The site 4,914 was hybridized with a biotin-modified primer that could bind to STV, which served as the reference point. Statistical analysis of AFM images revealed that 88.0% of the phiX 174 template was correctly labeled with the cross shape ID, which confirms the high stringency of AuNPs-enhanced primer extension for SNP genotyping using shape IDs (Supplementary Fig. 25)<sup>31</sup>.

**Single-molecule haplotyping in real samples.** The high-resolution, high-specificity and 'multi-colour' imaging ability of shape IDs opens new opportunities for single-molecule genotyping and haplotyping. As a first example, we phased a 4.6 kb region on the chromosome 3 from the Han Chinese population. There are two SNP loci in this region, that is, rs17038640 (C/T) and rs4676487 (C/T). For each locus, we employed triangle with or without STV to visualize SNPs under AFM. Genomic DNA extracted from human blood samples was subjected to amplification using exonuclease-assisted and AuNP-enhanced long-range PCR (Supplementary Figs 26 and 27), generating linear dsDNA for AFM imaging. The two loci locate 1.2 and 4.19 kb from the 5'-end of the 4.6 kb fragment, which were labelled with triangular shape ID with or without streptavidin. AFM images revealed that the genomic DNA of one sample was homozygous at both loci and the genomic DNA of the other three samples was heterozygous at both loci. These direct read sequences of each allele (Fig. 5a) were confirmed using capillary sequencing<sup>31</sup>.

As a further illustration of the ability to phase haplotypes in the long range, we genotyped a 34 kb region associated with age-related macular degeneration (AMD), which is the leading cause of blindness worldwide. This region carries seven SNP alleles on the chromosome 10 from the Han Chinese population. To avoid the inefficiency in amplifying this long region, we split it into four overlapping fragments, each having two or three SNP loci (Supplementary Fig. 28). Each fragment has at least one SNP



**Figure 3 | ‘Multi-colour’ nanomechanical imaging and genotyping resolution of origami shape IDs.** (a) ‘Multi-colour’ imaging of phiX174 with multiple shape IDs and their distance analysis. Left, three different shape IDs, triangular (corresponding to site 1,433), triangular with STVs (corresponding to site 1,529) and cross (corresponding to site 4,914), are site-specifically labelled on phiX 174 with different labelling efficiency (85% for the single-site labelling, 56.4% for the two-site labelling and 30.8% for the three-site labelling). Scale bar, 200 nm. Right, histograms for the counts of shape IDs as a function of distance between site 1,433, 1,529 and 4,914 in phiX. As for two-site labelling (left), the measured distance between site 1,433 and 4,914 is 650 nm, which is in good agreement with calculated distance 648 nm. As for three-site labelling (right), the measured distances between site 1,433, 1,529 and 4,914 are 652 nm (blue) and 682 nm (red), which are in good agreement with the calculated distances 648 and 680 nm. Also see Supplementary Figs 14 and 15. (b) Schematic showing and three-dimensional AFM topographic images of triangular- (corresponding to site 1,433), cross- (corresponding to site 1,463) and STV-decorated triangular- (corresponding to site 1,529) shaped IDs for labelling phiX 174. Left, the contour length between two labelled sites (1,433 and 1,463) is  $\sim 10$  nm (30 bp). Right, the contour length between two labelled sites (1,433 and 1,529) is  $\sim 32$  nm (96 bp).

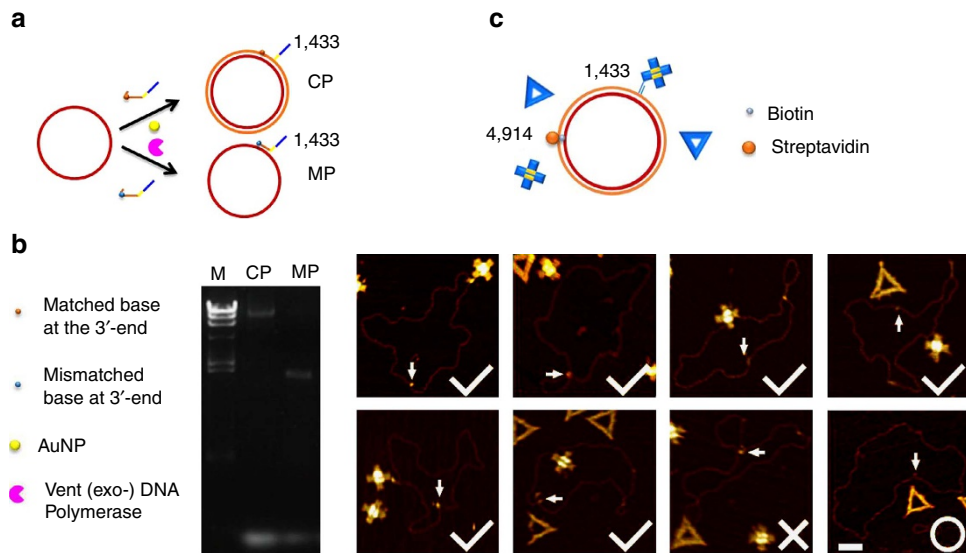
locus (that is, ‘joint locus’) that is heterozygously shared by its adjacent fragment, which can be used for subsequent fragment connection. As an example, SNP-2 (rs11200630) was used as the joint locus for fragment 1 and 2 to connect these two fragments. By using shape ID-based AFM imaging, we could genotype SNP loci on each fragment via direct reading. Connection of the four fragments reconstructs a continuous haplotype for this 34 kb region. We found that all the seven SNPs are heterozygous in the sample and two haplotypes were directly read using AFM (Supplementary Fig. 29), which were also supported by the capillary sequencing results<sup>31</sup>.

Having established the ability of shape ID-based nanomechanical imaging for reliable phasing known haplotypes of AMD, we further explored to haplotype a 12 kb region of the human *p53* gene located on the chromosome 17 using a single-blinded test (Fig. 5b and Supplementary Fig. 30). The haplotype of this region has not been shown in the literature. We first sequenced this region in a sample of Han Chinese origin by using capillary sequencing and found three heterozygous SNP sites: C/A at rs12951053, C/G at rs1042522 and C/T at rs17882227. By using six different shape IDs to label these sites, we obtained the haplotype of this region (Fig. 5b). We independently haplotyped the same region using our previously reported strategy that combines AuNP-assisted allele-specific PCR with capillary sequencing. Significantly, this single-blinded test showed consistent haplotyping information, which clearly illustrated the reliability of the new imaging-based haplotyping method. Also of

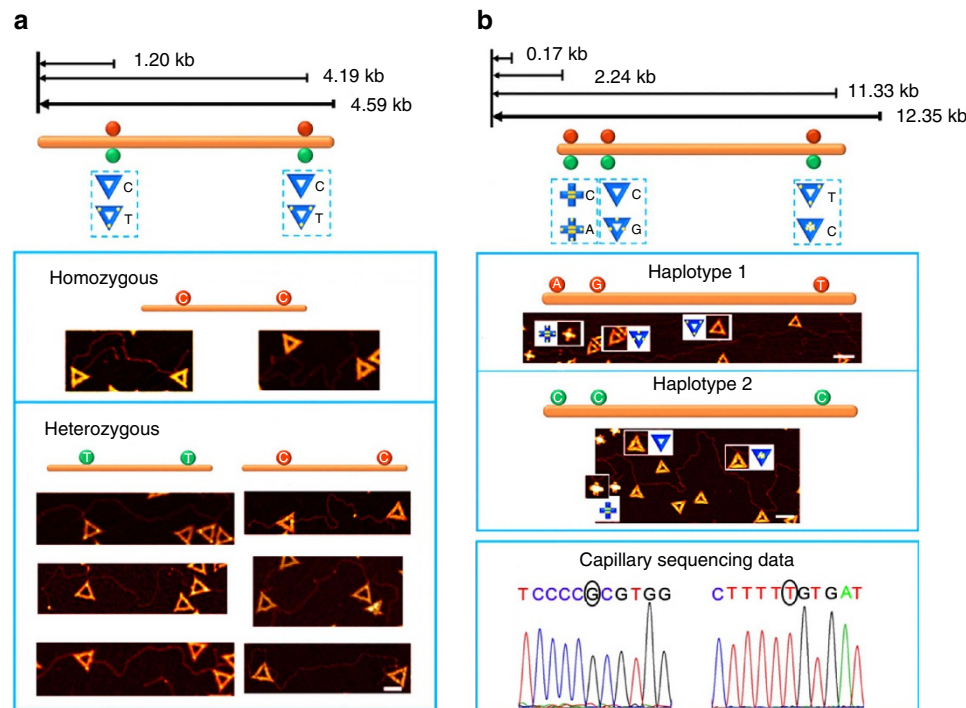
note is that the first two SNPs are spaced by 170 bp ( $\sim 55$  nm), which cannot be resolved using a fluorescence imaging-based haplotyping method due to the limit of optical diffraction<sup>28,29</sup>, which highlights the significance of developing high-resolution haplotyping methods.

## Discussion

High-resolution imaging is an ever-increasing desire for genetic analysis given the complexity of the genome, especially for resolving genetic variations in close proximity. AFM-based nanomechanical imaging is intrinsically a high-resolution method that has long been pursued as a promising tool for single-molecule genetic analysis and manipulation<sup>10–12,32,33</sup>. However, direct visualization of single DNA nucleotides in air is currently not feasible under AFM. Prior efforts to using label-based AFM imaging led to resolution limit of several tens of nanometres due to the lack of appropriate labels with high contrast<sup>12</sup>. In this work, we developed a new strategy using origami shape IDs as ‘magnifying lens’ to translate angstrom-sized individual nucleotides to origami nanostructures of several tens of nanometres. We demonstrated that this nanomechanical imaging method reached an ultrahigh haplotyping resolution of 10 nm (30 bp) that far exceeds fluorescence- and even superresolution-based imaging<sup>34–36</sup>. As this new method does not have a theoretical limit, it is possible to design smaller shape IDs ( $< 10$  nm) to further increase the resolution, which holds



**Figure 4 | Specificity of origami shape IDs.** (a) Schematic showing of AuNP-assisted allele-specific primer extension of phiX 174 by a fully complementary primer (CP) or no extension by a mismatched primer (MP), the single mismatched base of which is at the 3'-end. (b) Agarose gel electrophoresis validated that CP could initiate the extension of template ssDNA into dsDNA, whereas MP could not. M, DL 15, 000 marker. (c) Upper: schematic showing of shape ID-based haplotyping. The site 1,433 is labelled with two M-strands with or without a single-base mismatch, which correspond to a triangular- and a cross-shaped ID, respectively. The site 4,914 is hybridized with a biotin-modified primer, which serves as the reference point. Lower: AFM images show that phiX 174 is correctly labelled with cross-shaped IDs (marked by tick), wrongly labelled with triangular (marked by cross), or with no labelling (marked by circle). The site 4,914 labelled with STV is marked with an arrowhead. The yield of specific labelling of shape IDs is ~88%. Scale bar, 100 nm.



**Figure 5 | Single-molecule haplotyping.** (a) Upper, schematic showing of genotypes and relative locations of two SNPs on a 4.6 kb region. SNP loci rs17038640 (C/T) and rs4676487 (C/T) are 1.20 and 4.19 kb away from one end of the region. Lower, AFM images show that the genomic DNA of one sample is homozygous at both loci (C-C) and the genomic DNA of the other three samples is heterozygous at both loci (T-T and C-C). For each SNP locus, the triangular- and the STV-decorated triangular-shaped IDs are employed to label the alleles C and T, respectively. Scale bar, 100 nm. (b) Single-blinded test for haplotyping *P53* gene. Upper, schematic showing of a 12 kb region of the human *p53* gene located on the chromosome 17. Each origami shape ID corresponds to a specific allele. For example, for the SNP 1, A and C allele correspond to cross origami with or without STVs, respectively. Middle, AFM images for the haplotypes of these three SNPs. Haplotype 1 contains A-G-T and haplotype 2 contains C-C-C. Lower, when the first SNP was A on one sequence, the other SNPs obtained from independent capillary sequencing were G and T, which was consistent with that from the nanomechanical imaging in haplotype 1. Scale bar, 200 nm.

great promise SNP genotyping approaching single-base resolution. We also envision that this technology might provide a direct reading-based route to *de novo* assembly of genomes from short reads obtained from NGS DNA sequencing<sup>23</sup>.

DNA origami is powerful technology that can fold a long ssDNA template to virtually any prescribed shapes or patterns, which have found numerous applications in various areas. Nevertheless, the presence of intense negative electrostatic force field and significant steric effects pose intense limits for on-origami DNA hybridization. Our finding of highly hybridizable ‘hotspots’ on points of triangular and cross origami allows us to repurpose DNA origami nanostructures as nanomechanical shape IDs (with a single labelling efficiency of ~85%). Notably, by combining the shape designing and protein patterning abilities of DNA origami, it is possible to design virtually unlimited numbers of shape IDs for distinguishable AFM imaging (up to 16 as illustrated in this work). This compares favourably with fluorescence imaging (typically four colours) or super-resolution imaging (one to two colours) that are restricted by spectral overlap of fluorophores<sup>36</sup>. In addition, these shape recognition-based nanomechanical imaging has high imaging contrast and is less prone to artefacts due to the absence of bleaching or blinking of fluorophores in fluorescence imaging.

Single-molecule analysis using nanomechanical imaging is insusceptible to artefacts that are often encountered in ensemble assays. We can image individual long genomic DNA with labelled shape IDs and determine the relative locations of the latter from the visualizable contour length of the former. Non-labelled shape IDs remaining on the surface are clearly distinguishable from those labelled on the genomic DNA. More importantly, statistical analysis of multiple genomic DNA strands provides a robust means for genetic analysis, even at a relatively low labelling efficiency in multiplexing. As a demonstration of this single-molecule nanomechanical imaging method, we have phased the haplotypes of a target region associated with AMD in the Han Chinese people and a previously unknown haplotype of a fragment DNA in P53, which were all confirmed by capillary sequencing. We also note that, although the cost, speed and throughput of this single-molecule haplotyping remain to be improved, advances in high-speed and parallel AFM should provide a solution.

The real-world applications of this technology remain constrained by the cost and availability of AFM. The lab-based AFM is expensive and not easy to use or access. However, AFM is in principle a nanomechanical cantilever, which can be mass-produced with low cost and in a highly parallel manner. The development and mass-production of automated AFM should be a key step for the practical utility of single-molecule haplotyping.

In summary, we have repurposed DNA origami as highly hybridizable nanomechanical shape IDs for targeted genetic phasing of long-range haplotypes. The availability of a wide spectrum of distinguishable shape IDs leads to efficient ‘multi-colour’ labelling of genomic DNA for robust haplotyping at the single-molecule level. Owing to the magnifying effect of these shape IDs, we can reliably phase haplotypes with an ultrahigh resolution of 10 nm. These DNA origami-based shape IDs also hold great promise as biomolecular barcodes for *in-situ* probing of genetic variations on the genome. With the new advances of AFM in both resolution and throughput, it is also possible to implement direct sequencing of the genome and rapid genetic diagnosis for clinical translation.

## Methods

**Oligonucleotides.** Sequences of the oligonucleotides and primers used in this work are available in the Supplementary Information.

**Fabrication of the DNA origami shape IDs.** Triangular and cross-shaped DNA origami shape IDs were prepared by annealing the single scaffold stranded M13mp18 with staple strands and capturing strands in 1xTAE-Mg<sup>2+</sup> (40 mM Tris, 20 mM acetic acid, 2 mM EDTA and 12.5 mM magnesium acetate, pH 8.0) with a ratio of 1:10 from 95 °C to room temperature in a rate of 1 °C min<sup>-1</sup> (refs 14,37). In the experiment where STV is decorated in origami, 10 × excess of STV was added before incubation. All DNA Origami IDss were purified by agarose gel electrophoresis and subsequently extracted using Freeze ‘N Squeeze DNA Gel Extraction Spin Columns.

**PCR reaction.** A master mix contains 1.25 ng of template DNA, 0.4 μM each of the forward and reverse PCR primers, 0.25 μl LA Taq (5 units μl<sup>-1</sup>), 2.5 μl of 10 × LA PCR Buffer II (Mg<sup>2+</sup> plus), 4 μl of dNTP mixture (2.5 mM) in 25 μl reaction volume, which was prepared on ice. One of the PCR primers was modified with 5'-Cy3 label and the other was modified with 5'-phosphorylation. The mixture was annealed in a PCR thermocycler and was subjected to a 2 min hot start at 95 °C, followed by 30 cycles of amplification (98 °C for 10 s, then 68 °C for 15 min). The protocol ends with a final extension step at 72 °C for 10 min.

**Generation of long single-stranded DNA template.** To generate the single-stranded DNA, the PCR product was digested. Two microlitres of lambda exonuclease, 3 μl of 10 × lambda exonuclease reaction buffer and purified dsDNA were mixed to 30 μl and incubated in a thermocycler for 6 h at 37 °C. The exonuclease was heat inactivated at 85 °C for 8 min after the digestion.

**AuNPs-enhanced primer extension.** The sequences of mediator DNA primers were designed by Primer Premier 5.0 and NUPACK (www.nupack.org)<sup>38</sup>. ssDNA was annealed with primers in extension solution (20 μl), which consists of 1 × ThermoPol Reaction Buffer, 0.25 mM each dNTP, 0.75 μM each primer, 4 nM of AuNPs and deionized water. The allele-specific labelling reaction was performed in a PCR thermal machine. The extension was initiated from 95 °C (2 min) and annealed to 25 °C at room temperature. Vent (exo-) DNA Polymerase was added and stay at 60 °C for 45 min. The linear amplification products were purified from agarose gel to remove excess primers, nucleotides, enzymes, salts and other impurities, and ready for hybridization.

**Hybridization of target DNA with STV or origami shape IDs.** Biotinylated or mediator primers labelled target DNA was incubated with excessive amounts of STVs or origami shape IDs at 37 °C for 30 min.

**AFM imaging.** Three-microlitre solution of the samples were deposited on a freshly cleaved mica surface and left to adsorb to the surface for 3 min, then the mica surface was slowly rinsed with water three times (each time with 100 μl water) to wash away the salt. Finally, the mica surface was dried with a mild air stream by an ear-washing bulb and was imaged with a MultiMode 8 AFM with NanoScope V Controller (Bruker, Inc.) in air under tapping mode.

**Single-molecule image analysis.** The contour length of DNA molecules and the distances between every two sites labelled by origami IDs or STVs were measured by ImageJ (National Institutes of Health, Bethesda, Maryland, USA, <http://imagej.nih.gov/ij/>).

**Data availability.** The data supporting the main findings of this study are available within the main article and its Supplementary Information (Supplementary Figs 1–30 and Supplementary Tables 1–7) or from the authors upon request.

## References

- Hong, G. *et al.* Ultrafast fluorescence imaging *in vivo* with conjugated polymer fluorophores in the second near-infrared window. *Nat. Commun.* **5**, 4206 (2014).
- Wang, J. S. & Zhang, D. Y. Simulation-guided DNA probe design for consistently ultraspecific hybridization. *Nat. Chem.* **7**, 545–553 (2015).
- Zhou, X. *et al.* Phage-mediated counting by the naked eye of miRNA molecules at attomolar concentrations in a Petri dish. *Nat. Mater.* **14**, 1058–1064 (2015).
- Liong, M. *et al.* Magnetic barcode assay for genetic detection of pathogens. *Nat. Commun.* **4**, 1752 (2013).
- Mortensen, K. L., Sung, J., Flyvbjerg, H. & Spudich, J. A. Optimized measurements of separations and angles between intra-molecular fluorescent markers. *Nat. Commun.* **6**, 8621 (2015).
- Zhang, K. *et al.* Long-range polony haplotyping of individual human chromosome molecules. *Nat. Genet.* **38**, 382–387 (2006).
- Muller, D. J. & Dufrene, Y. F. Atomic force microscopy as a multifunctional molecular toolbox in nanobiotechnology. *Nat. Nanotechnol.* **3**, 261–269 (2008).
- Hansma, H. G., Sinsheimer, R. L., Li, M.-Q. & Hansma, P. K. Atomic force microscopy of single- and double-stranded DNA. *Nucleic Acids Res.* **20**, 3585–3590 (1992).

9. Gross, L., Mohn, F., Moll, N., Liljeroth, P. & Meyer, G. The chemical structure of a molecule resolved by atomic force microscopy. *Science* **325**, 1110–1114 (2009).
10. Lü, J.-H. *et al.* Positioning isolation and biochemical analysis of single DNA molecules based on nanomanipulation and single-molecule PCR. *J. Am. Chem. Soc.* **126**, 11136–11137 (2004).
11. Kufer, S. K., Puchner, E. M., Gump, H., Liedl, T. & Gaub, H. E. Single-molecule cut-and-paste surface assembly. *Science* **319**, 594–596 (2008).
12. Woolley, A. T., Guillemette, C., Cheung, C. L., Housman, D. E. & Lieber, C. M. Direct haplotyping of kilobase-size DNA using carbon nanotube probes. *Nat. Biotechnol.* **18**, 760–763 (2000).
13. Dunn, K. E. *et al.* Guiding the folding pathway of DNA origami. *Nature* **525**, 82–86 (2015).
14. Rothmund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
15. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
16. Andersen, E. S. *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459**, 73–76 (2009).
17. Ke, Y., Lindsay, S., Chang, Y., Liu, Y. & Yan, H. Self-assembled water-soluble nucleic acid probe tiles for label-free RNA hybridization assays. *Science* **319**, 180–183 (2008).
18. Douglas, J. A., Boehnke, M., Gillanders, E., Trent, J. M. & Gruber, S. B. Experimentally-derived haplotypes substantially increase the efficiency of linkage disequilibrium studies. *Nat. Genet.* **28**, 361–364 (2001).
19. Collins, F. S., Green, E. D., Guttacher, A. E. & Guyer, M. S. A vision for the future of genomics research. *Nature* **422**, 835–847 (2003).
20. Ward, L. D. & Kellis, M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat. Biotechnol.* **30**, 1095–1106 (2012).
21. Peters, B. A. *et al.* Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. *Nature* **487**, 190–195 (2012).
22. Kuleshov, V. *et al.* Whole-genome haplotyping using long reads and statistical methods. *Nat. Biotechnol.* **32**, 261–266 (2014).
23. Metzker, M. L. Sequencing technologies—the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
24. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
25. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
26. Fan, H. C., Wang, J., Potanina, A. & Quake, S. R. Whole-genome molecular haplotyping of single cells. *Nat. Biotechnol.* **29**, 51–57 (2011).
27. Selvaraj, S., Dixon, J. R., Bansal, V. & Ren, B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat. Biotechnol.* **31**, 1111–1118 (2013).
28. Xiao, M. *et al.* Direct determination of haplotypes from single DNA molecules. *Nat. Methods* **6**, 199–201 (2009).
29. Lam, E. T. *et al.* Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* **30**, 771–776 (2012).
30. Li, H. *et al.* Nanoparticle PCR: nanogold-assisted PCR with enhanced specificity. *Angew. Chem. Int. Ed.* **44**, 5100–5103 (2005).
31. Chen, P. *et al.* Gold nanoparticles for high-throughput genotyping of long-range haplotypes. *Nat. Nanotechnol.* **6**, 639–644 (2011).
32. Erdmann, M., David, R., Fornof, A. & Gaub, H. E. Electrically controlled DNA adhesion. *Nat. Nanotechnol.* **5**, 154–159 (2010).
33. Tanaka, H. & Kawai, T. Partial sequencing of a single DNA molecule with a scanning tunnelling microscope. *Nat. Nanotechnol.* **4**, 518–522 (2009).
34. Baday, M. *et al.* Multicolor super-resolution DNA imaging for genetic analysis. *Nano Lett.* **12**, 3861–3866 (2012).
35. Beliveau, B. J. *et al.* Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* **6**, 7147 (2015).
36. Hell, S. W. Far-field optical nanoscopy. *Science* **316**, 1153–1158 (2007).
37. Liu, W., Zhong, H., Wang, R. & Seeman, N. C. Crystalline two-dimensional DNA-origami arrays. *Angew. Chem. Int. Ed.* **123**, 278–281 (2011).
38. Zadeh, J. N. *et al.* NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173 (2011).

## Acknowledgements

This work was supported by the Ministry of Science and Technology of China (2013CB932803, 2013CB933802, 2016YFA0201200, 2015CB559100 and 2016YFC0903402), the National Science Foundation of China (21390414, 21227804, 21329501, 31325014, 81130022, 81272302, 81421061, 31571012, 21375139 and 21473236), the Chinese Academy of Sciences (QYZDJ-SSW-SLH031), Program for Changjiang Scholars and Innovative Research Team in University (IRT\_15R37) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (YX03002).

## Author contributions

C.F., Y.S. and J.C. directed the study. C.F., J.C. and H.Z. wrote the manuscript. H.Z., J.C. and D.P. performed the experiments, prepared the figures and revised the article. All authors contributed to the design of the study and writing the paper.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Zhang, H. *et al.* DNA origami-based shape IDs for single-molecule nanomechanical genotyping. *Nat. Commun.* **8**, 14738 doi: 10.1038/ncomms14738 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017