



OPEN

## Enrichment of low abundance DNA/RNA by oligonucleotide-clicked iron oxide nanoparticles

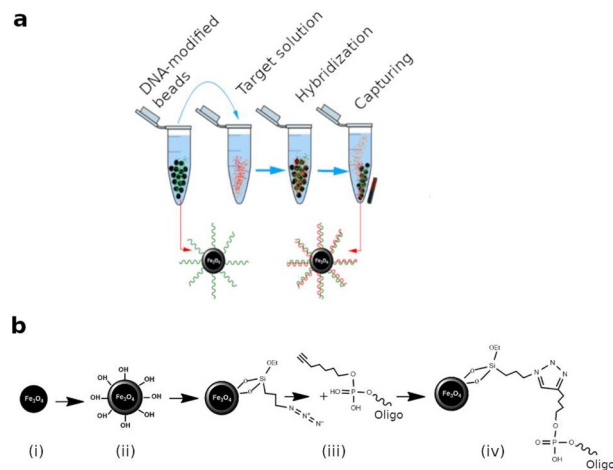
Fereshte Damavandi<sup>1,2,7</sup>, Weiwei Wang<sup>3,5,7</sup>, Wei-Zheng Shen<sup>1,2</sup>, Sibel Cetinel<sup>1,2,6</sup>, Tracy Jordan<sup>3</sup>, Juan Jovel<sup>3</sup>, Carlo Montemagno<sup>1,2,8</sup> & Gane Ka-Shu Wong<sup>3,4</sup>✉

Detection of low abundance target DNA/RNA for clinical or research purposes is challenging because the target sequences can be hidden under a large background of human genomic or non-human metagenomic sequences. We describe a probe-based capture method to enrich for target sequences with DNA-clicked iron oxide nanoparticles. Our method was tested against commercial capture assays using streptavidin beads, on a set of probes derived from a common genotype of the hepatitis C virus. We showed that our method is more specific and sensitive, most likely due to the combination of an inert silica coating and a high density of DNA probes clicked to the nanoparticles. This facilitates target capture below the limits of detection for TaqMan qPCR, and we believe that this method has the potential to transform management of infectious diseases.

In many samples that are sequenced for clinical or research purposes, the targets of interest are found in mixtures of asymmetrical abundance, i.e. the target sequences are obscured under a massive background of human genomic or non-human metagenomic sequences. Next generation sequencing (NGS) is effective for surveying medium-to-high abundance targets, but it is more problematic for low abundance targets, e.g. viruses and bacteria at the initial pre-symptomatic or the later chronic stages of infection, non-abundant viruses and bacteria in dirty metagenomics samples, or subsets of gene transcripts that are expressed at low levels. Several approaches have been developed for target enrichment, most notably PCR pre-amplification<sup>1</sup> and nucleic acids capture<sup>2</sup>. Initial efforts that combined hybridization capture with NGS were onerous, expensive, and time consuming. For instance, the first human gene capture arrays used solid support and expensive instruments like the MAUI hybridization station (NimbleGen). They had run times of 65 h<sup>3,4</sup>. Magnetic particles for separation of nucleic acids have also been developed<sup>5</sup>, a system referred to as solution hybridization capture. DNA probes are biotinylated to enable the isolation of complementary molecules by capturing biotin with high-affinity streptavidin-coated paramagnetic beads<sup>6</sup>.

The coding regions of the human genome (the exome) comprise about 1% of the genome and have arguably been the paramount subject of study for hybridization-based capture and NGS<sup>6–10</sup>. Some examples are diagnosis of Mendelian disorders<sup>11,12</sup>, development of precision oncology approaches for monitoring cancer evolution and the response to treatment<sup>13,14</sup>, and discovery of naturally occurring loss-of-function mutations that correlate with medically desirable traits to guide drug development<sup>15,16</sup>. Capture assays have also been used for recovery of gene transcripts in degraded archival samples that are not amenable to poly(A) selection<sup>17</sup>. Mitochondrial and nuclear genes share homology, so it is also possible to assemble the mitochondrial genome using off-target sequences captured during whole-exome enrichment<sup>18</sup>. This at the same time reflects the limited specificity of capture systems currently used. These methods have also been used for genotyping of bacteria in natural environments<sup>19</sup>, bacterial transcriptome profiling during host cell infection<sup>20</sup>, enrichment of viromes and individual viruses<sup>21–24</sup>, genotyping of flowering-time genes in morphotypes of oilseed rape<sup>25</sup>, disentangling of octoploid strawberry subgenomes<sup>26</sup>, identification of insecticide-resistance mutations from the mosquito genome<sup>27</sup>, and mapping of resistant NB-LRR loci in *Solanum*<sup>28</sup>.

<sup>1</sup>Ingenuity Lab, 1-070C, 11421 Saskatchewan Drive NW, Edmonton, AB T6G 2M9, Canada. <sup>2</sup>Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada. <sup>3</sup>Department of Medicine, University of Alberta, Edmonton, AB T6G 2E1, Canada. <sup>4</sup>Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada. <sup>5</sup>Present address: Geneis Inc., Bldg A, 5 Guangshun North Street, Beijing, China. <sup>6</sup>Present address: Nanotechnology Research and Application Center (SUNUM), Sabanci University, Istanbul 34956, Turkey. <sup>7</sup>These authors contributed equally: Fereshte Damavandi and Weiwei Wang. <sup>8</sup>Carlo Montemagno is deceased. ✉email: gane@ualberta.ca



**Figure 1.** Description of InBeads capture system. **(a)** DNA-conjugated IONPs are incubated with a target solution to promote hybridization. Non-specifically hybridized DNA/RNA are removed by washes. Specifically hybridized DNA/RNA are retained on the magnetically immobilized beads. **(b)** Preparation of InBeads. (i) IONPs synthesis, (ii) silica coating, (iii) azide functionalization, and (iv) conjugation with DNA probes through click chemistry.

However, all commonly used capture systems have a limited recovery rate for low abundance target sequences. Commercial streptavidin-coupled magnetic beads, in particular, are surprisingly inefficient<sup>21,23,29</sup>. Although they provide some enrichment of the target molecules, non-specific binding to human genomic (or non-human metagenomic) background sequences can overwhelm the low abundance target sequences. To overcome this limitation, we designed a novel capture system to minimize non-specific interaction with nucleic acids, which uses a magnetic iron oxide nanoparticle (IONP) coated with an inert silica layer. Silica coating protects the magnetic core and prepares the surface of the nanoparticle for subsequent functionalization. To attach DNA (or alternatively RNA), a layer of azide functionalized silane is deposited on the silica shell, after which the nucleic acid probes are attached with a click chemistry reaction using Cu[I]-catalyzed azide-alkyne Huisgen cycloadditions (CuAAC), which gives stable and rigid triazole linkages. Reaction conditions are designed to optimize the density of probes on the surface, further reducing the likelihood of non-specific interactions through negative charges and steric hindrance<sup>30–33</sup>. Performance of this novel capture system, relative to commercial streptavidin-coupled magnetic beads, is demonstrated on simulated and real clinical samples containing the hepatitis C virus (HCV) at varying concentrations.

## Results

In most capture protocols, target sequences are hybridized to biotin-labeled probes in solution, and only after that are streptavidin-coated beads used to “pull down” on-target sequences by virtue of the high affinity of streptavidin for biotin. However, we do the reverse. Our probes are conjugated to the beads prior to hybridizing with the target sequences. We do not believe this difference has any meaningful impact on capture performance, but it is something to note as it is not the conventional protocol.

**InBeads preparation and characterization.** Iron oxide nanoparticles are silica-coated to provide an inert surface that reduces off-target hybridization (Fig. 1a). They are prepared by a solvothermal reaction, followed by silica coating, azide functionalization, and conjugation of DNA probes with click chemistry (Fig. 1b). Click chemistry was implemented with a 5'-hexynyl modification that introduces a 5'-terminal alkyne group into the DNA. This readily reacts with azide in the presence of copper, forming stable 1,2,3-triazole bonds. Cutler et al.<sup>34</sup> used a different click chemistry reaction to conjugate DNA to iron oxide nanoparticles and tested for cellular uptake. They used 10 nm aminated iron oxide nanoparticles (Ocean Nanotech) coated with an amphiphilic polymer, resulting in a high cellular uptake. However, polymer coatings are prone to non-specific interactions with biomolecules, which limits their applicability for bead-based capture. We chose a silica shell precisely to reduce non-specific interactions. Our magnetic iron-oxide-nanoparticles (IONPs) also have an average diameter of  $235 \pm 20$  nm (Fig. S1a–b), with an additional 40 nm shell for the silica coating (Fig. S1c–d). IONPs conjugated to DNA are hereafter referred to as InBeads.

To confirm that probe-target hybridization occurred, we used 20-bp 3'-fluorescein-labeled 5'-hexynyl oligonucleotides (green-oligos) and 5'-Cy5-labeled complementary oligonucleotides (red-oligos). Click chemistry was performed to conjugate azide-functionalized IONPs to the green-oligos. Hybridization against the red-oligos was promoted by heating the reaction to 65 °C and slowly cooling to room temperature. Green fluorescence (from the fluorescein-labeled probes; Fig. S2b) and red fluorescence (from the Cy5-labeled targets; Fig. S2c) could be detected before and after hybridization, respectively. This procedure was repeated using eight 90-bp probes complementary to several regions of the HCV genome. X-ray photoelectron spectroscopy (XPS) revealed that DNA grafting was successful (Fig. S2d, peaks N1s and P2p).

To optimize how DNA is immobilized on the beads surface, we performed the click-chemistry reaction under different conditions. Amount of free DNA in solution before and after the reaction was quantified with the Quant-iT OliGreen ssDNA Assay. The difference was interpreted as the amount of immobilized DNA (iDNA). Reaction conditions were assessed with probes (C and F) that bind to different regions of the HCV genome. Figure S3a shows the amount of iDNA at different concentrations of DNA in the reaction (0.1, 1, 2.4, 5, and 7.5  $\mu\text{M}$ ). Both probes showed an increase in iDNA with increases in DNA concentration. Yields peaked at 2.4  $\mu\text{M}$  with both probes (Fig. S3a), arguing that the immobilization was most efficient at this DNA concentration. Amount of iDNA increased over time (Fig. S3b). Finally, while no DNA was immobilized on our beads without Cu(I), the reaction reached a saturation point at around 0.5 mM of Cu(I) due to the steric hindrance between DNA strands (Fig. S3c).

**Demonstrations of InBeads performance.** *Enrichment for targets and reduction in non-specific capture.* The hepatitis C virus (HCV) is a well-studied pathogen that causes liver inflammation, cirrhosis, and cancer. In the early stages of infection, most patients are asymptomatic, but approximately half of infected subjects develop a chronic infection. The disease often remains undiagnosed until serious liver damage has occurred<sup>35</sup>. To evaluate the efficiency of our capture system, two sets of InBeads with probes for two adjacent regions of the HCV genome (labeled A and B) were prepared. The targets used were synthetic genomic DNAs (a single fragment of length 270 bp covering A and B), which we called gblocks and incorporated into Illumina TruSeq libraries. In parallel, a library from human genomic DNA was also constructed. A ten-fold dilution series containing 10 to  $10^6$  copies of the HCV gblocks library was prepared, each mixed with 100 ng of the human library. This simulated a DNA mixture extracted from human cells infected with relatively small amounts of HCV. We used real time qPCR to measure absolute abundances (copy numbers), before and after capture, with the HCV probes A and B, and a probe for the human housekeeping gene B2M. Before capture, B2M copy numbers were out of range (Fig. 2a), while the HCV gblocks could only be detected in dilutions containing at least  $10^4$  or  $10^3$  molecules, using probes A or B, respectively (Fig. 2a). After capture however, recovery of the HCV gblocks was substantially enhanced (Fig. 2b,c). Stable target signals were obtained in dilutions containing at least  $10^2$  or  $10^3$  molecules, using probes A or B, respectively. Importantly, although B2M copy numbers were out of range before capture (Fig. 2a), they were undetectable in all instances after capture (Fig. 2b), showing the huge reduction in off-target effects associated with our system.

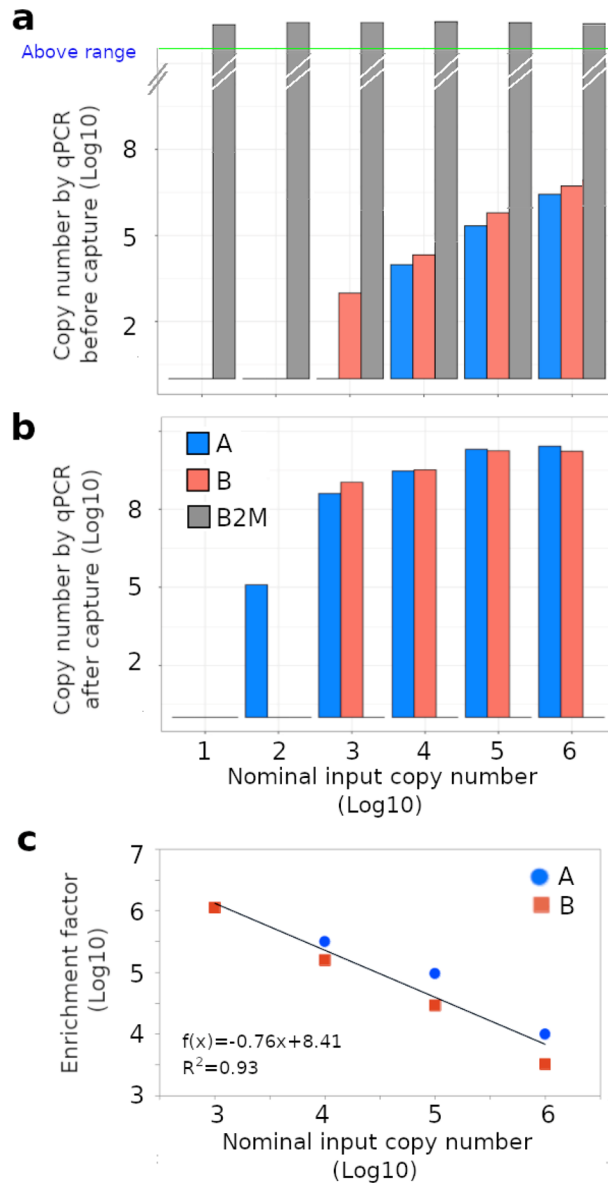
*Comparison of InBeads versus streptavidin-based capture.* To compare the performance of InBeads with streptavidin-conjugated beads (Dynabeads MyOne Streptavidin T1), eight sets of capture probes (A-H) spanning the HCV genome were designed and synthesized. These included the 5' untranslated region (UTR) and regions encoding the core (coat) protein, the envelope glycoprotein E2, and three non-structural proteins associated with viral replication (NS4B, NS5A and NS5B). NGS libraries were constructed for an HCV-positive patient (Pt339-1) showing the highest viral load for all our patient samples, with an estimated titer of  $5.6 \times 10^6$  RNA genomes per ml of plasma. Results for both capture systems were variable over different HCV regions (Fig. 3a). At 24 h of hybridization, InBeads were on average 154X more effective than Dynabeads, most notably for regions A, C, D, E and H, corresponding to 5'UTR, E2, E2, NS4B and NS5B, respectively (Fig. 3b). We also developed a fast InBeads protocol, requiring only 4 h of hybridization. There was some loss in performance (Fig. 3b), but on average it was still 12.5X more effective than Dynabeads. Importantly, the Dynabeads system failed to detect region D, so the improvement is shown as 'out of range'. For our target averaged improvements, we could not use infinity for region D, and so we adopted the best number from the other target regions. As such, the actual improvements are better than what we have indicated.

It should also be noted that qPCR saturates at copy numbers of roughly  $10^9$  or  $10^{10}$ . However, the relationship between the measured and the actual copy number is monotonic. So, if one signal is larger than another, it will remain larger when qPCR measurements are compared. Moreover, the actual ratio of these two signals will be larger than what is measured through qPCR. As such, the results described here are valid despite the potential for saturation, and quantitatively even better than what we have indicated.

*InBeads capture of HCV from plasma of infected patients.* We assessed InBeads performance on plasma samples from HCV patients, with viral titers starting at below 100 copies and ending at nearly  $10^7$  copies per ml. The fast (4 h hybridization) protocol was utilized. Despite its reduced performance, relative to our 24 h protocol, we still observed consistently stable signals from all 8 HCV target fragments (Fig. 4a), even for the one sample (Pt804-2) where HCV was not detectable by qPCR but where we knew the virus was nonetheless present as the patient eventually suffered a relapse. Reassuringly, we did not find a HCV-like signal in our negative control (Pt555), and we could not detect B2M in any of our samples (Fig. 4a). Enrichment factors varied with the target region but the probe-averaged trend versus target abundance was remarkably similar to what was seen in Fig. 2. Sample Pt339-1, with the highest viral titer in this data set, had an enrichment factor of 19x (Fig. 4b). Conversely for sample Pt804-2, where we arbitrarily set the copy number to 100 as the virus was below qPCR limits of detection, the enrichment factor was  $9.5 \times 10^5$  (Fig. 4b).

## Discussion

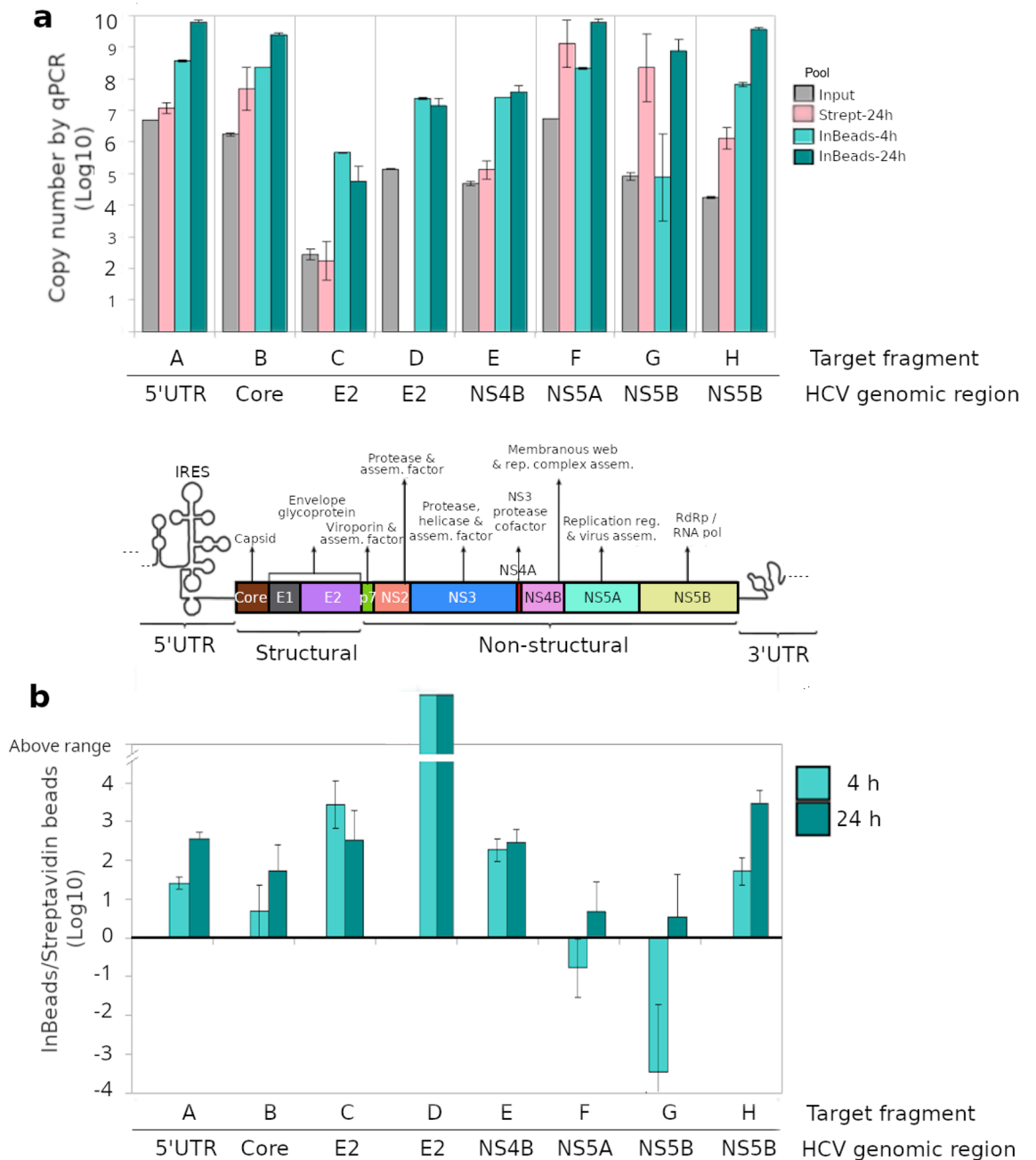
There are many ways to improve a capture system. First, sensitivity can be increased to handle samples with ultra-low target sequence abundances. Second, off-target hybridization can be reduced to increase the amount of useful sequence recovered. Third, hybridization times can be minimized without overly compromising the above considerations. InBeads excelled in all aspects: sensitivity, specificity, and speed.



**Figure 2.** InBeads evaluation with synthetic HCV targets. Capture specificity was evaluated with a synthetic DNA gblock representing two adjacent regions of the HCV genome (A and B), mixed with 100 ng of human genomic DNA to simulate a clinical sample background. Copy number was measured by real-time qPCR, for probes A and B as well as for a human housekeeping gene B2M, both before (a) and after (b) capture. Notably, B2M copy numbers were above-range before capture and undetectable after capture. (c) Enrichment factor is the ratio of copy numbers before and after capture.

Our initial experiments demonstrated that InBeads exhibit very high specificity. Before capture, copy numbers for the B2M housekeeping gene were out of range for qPCR, but copy numbers became undetectable after InBeads capture with two probes complementary to the HCV genome. In contrast for conventional biotin-streptavidin systems, off-target effects are evidenced by the fact that mitochondrial genomes can be recovered in off-target reads from nuclear whole-exome capture assays<sup>18,36</sup>. The superiority of InBeads over the commercial capture assay (Dynabeads MyOne Streptavidin T1) was evident across all target regions when both used 24 h hybridization. When the hybridization time was reduced to 4 h for InBeads, it still outperformed Dynabeads (over 6 of the 8 HCV fragments) despite the advantage of a 24 h hybridization for the latter. On average, performance of InBeads at 24 h or 4 h of hybridization was at least 154 or 12.5-fold superior to Dynabeads at the full 24 h hybridization time, respectively. As turnaround times are often critical in clinical applications<sup>37,38</sup>, this difference has practical implications.

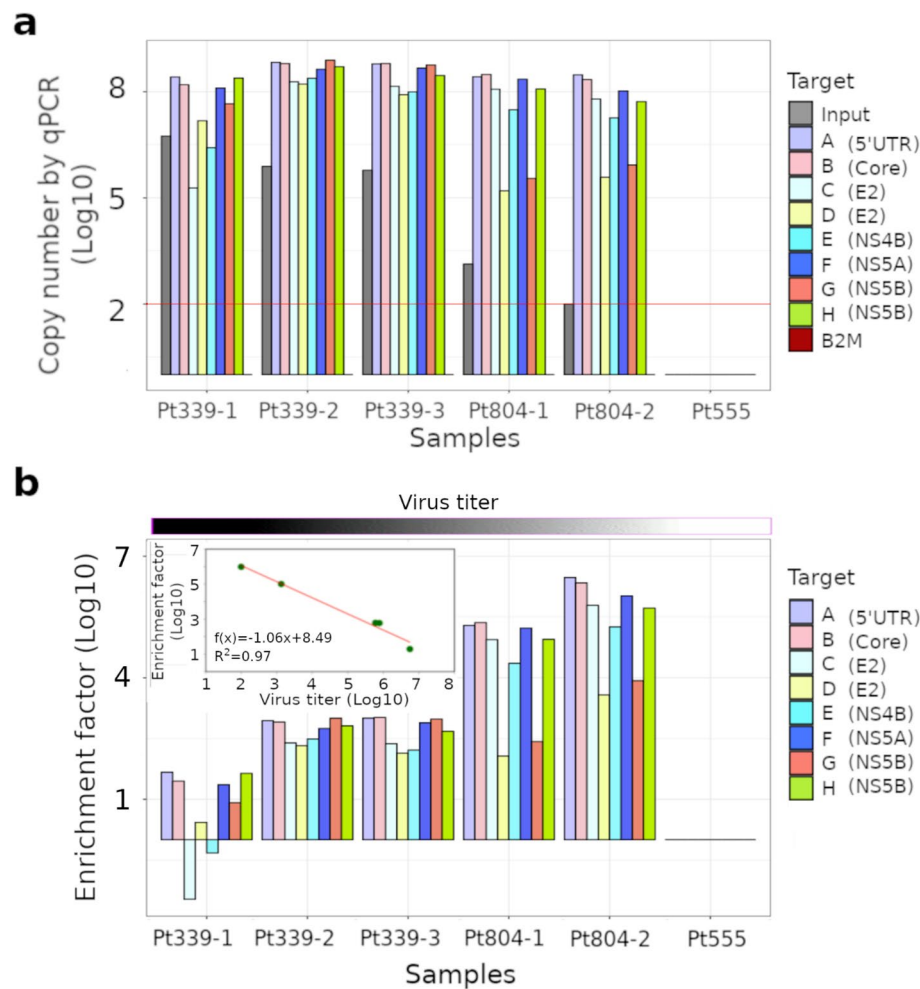
InBeads sensitivity and specificity with only 4 h hybridization was evaluated in plasma from patients with different titers of HCV. We observed a strong and consistent signal even in the one sample where HCV abundance was below the limits of detection for qPCR. Conversely, no signal was detected for the B2M human housekeeping gene. In a comparative study for several commercial capture systems, Thomson reported HCV enrichments in the



**Figure 3.** Comparing InBeads with streptavidin beads. Eight regions of the HCV genome (labels A to H) were targeted on one HCV patient (Pt339-1) sample. For InBeads, 4 h and 24 h hybridization times were used. For streptavidin, only 24 h was used. The capture experiments were done 1, 2, and 3 times for InBeads-4 h, InBeads-24 h, and streptavidin-24 h, respectively. (a) The copy numbers before and after capture were measured by qPCR, twice for each capture experiment. Data averaging was done on the log transformed values, i.e. geometric means and not arithmetic means. A schematic of the HCV genome shows which genes are in the targeted regions. (b) Ratio of InBeads to streptavidin performance. At 24 h, InBeads outperformed streptavidin for all regions, especially for D where streptavidin failed. Even at reduced 4 h hybridization, InBeads outperformed streptavidin in all but two regions (where curiously the error bars seemed larger than other regions).

2 to 168-fold range, when compared to metagenomics sequencing<sup>39</sup>. Bonsall observed enrichments of up to 1000-fold for an IDT biotinylated probe capture array<sup>40</sup>. Briese used a whole-virome capture array and reported enrichments of 100 to 10,000-fold in comparisons with conventional viral enrichment procedures using filtration<sup>21</sup>. InBeads showed the highest enrichment of all, especially at the lowest viral titer, where it almost reached 10<sup>6</sup>.

Beyond the demonstrated superiority of InBeads for capture hybridization, which is of particular importance for detection of low abundance pathogens, we would suggest that it may be time to rethink the near-universal use of streptavidin–biotin interactions in biomedical assays. Biologists like to use biomolecules because that is



**Figure 4.** Evaluation on plasma of HCV-infected patients. The same 8 regions from Fig. 3 were targeted on five HCV patient samples and one negative control. (a) Measured copy numbers were zero for all target regions in the HCV-negative (Pt555) sample, and the B2M gene in all HCV-positive samples. Red line indicates the detection limit of qPCR. (b) Enrichment factor is the ratio of copy numbers before and after capture. This is shown as a function of sample (main plot) and viral titer (inset). One sample (Pt804-2) was below the qPCR limits of detection, but known to be HCV-positive because the patient had a relapse. Nonetheless, InBeads capture was able to detect all 8 regions.

what they are comfortable with, but there have been many advances in nanotechnology and inorganic chemistry. Only by collaborating with people outside our comfort zones will the field move forward.

## Materials and methods

**Synthesis of superparamagnetic IONPs ( $\text{Fe}_3\text{O}_4$ ).** Superparamagnetic iron oxide nanoparticles were synthesized as per Tian et al.<sup>41</sup>, with some modifications. The precursor solution was prepared by dissolving iron chloride ( $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$ ; Sigma-Aldrich) in ethylene glycol (EG; Sigma-Aldrich). NaAc (Sigma-Aldrich) was added as a pH adjuster, under stirring for 20 min. The resultant solution was autoclaved at 200 °C for 8 h. After allowing for cooling, the black product was collected using a magnet and washed 3X with water and ethanol. The washed product was freeze-dried.

**Adding a silica coating to the IONPs ( $\text{Fe}_3\text{O}_4@\text{SiO}_2$ ).** IONPs were coated with a ~40 nm silica shell through a hydrolysis-condensation reaction of tetraethyl orthosilicate in a sol-gel process<sup>42</sup>. Briefly, 1 mg/ml of IONPs suspension was prepared in a solution of water, ethanol, and ammonium hydroxide (Sigma-Aldrich). The mixed solution was sonicated to produce well-dispersed magnetic particles. During sonication, 0.05 mg TEOS (Sigma-Aldrich) was added dropwise, to promote reaction with water and initiate the coating process. The reaction was continued under vigorous stirring for 6 h. Finally, the silica-coated IONPs were separated from the solution with a small magnet and freeze-dried.



**Azide functionalization of IONPs ( $\text{Fe}_3\text{O}_4@\text{SiO}_2@\text{N}_3$ ).** Azide functionalization groups were added to the silica-coated surface through a 3-azidopropyl triethoxysilane deposition<sup>43</sup>. Following a typical protocol, 5 mg of dried silica iron oxide nanoparticles were suspended in tetrahydrofuran (THF; Sigma-Aldrich) under  $\text{N}_2$  protection, followed by addition of 4 mg 3-azidopropyl triethoxysilane under continuous stirring. This mixture was kept stirred two days at room temperature. Functionalized IONPs were collected by magnetic separation and then washed 3X with THF and ethanol to remove surface residues. Once again, the washed product was freeze dried.

**Attaching DNA by click chemistry ( $\text{Fe}_3\text{O}_4@\text{SiO}_2@\text{DNA}$ ).** DNA-probe-clicked IONPs were prepared via a copper(I)-catalyzed azide-alkyne cyclo-addition (CuAAC) reaction employing azide-functionalized IONPs and alkyne-modified oligonucleotides. Briefly, in a 500  $\mu\text{L}$  test tube, 10  $\mu\text{L}$  of 24  $\mu\text{M}$  5'-hexynyl modified DNA probe (IDT) in water was mixed with 3  $\mu\text{L}$  of a 2 M triethylammonium acetate buffer (pH 7.0) and 10  $\mu\text{L}$  DMSO. We then added 10  $\mu\text{L}$  of 9.6 mg/mL azide functionalized IONPs in DMSO and water (1:1) as solvent and 3  $\mu\text{L}$  of 5 mM ascorbic acid in water, after which we bubbled the reaction with argon gas for 30 s. 3.5  $\mu\text{L}$  of 10 mM Cu(II)-TBTA solution in 55% DMSO (Lumiprobe) was added under argon gas protection. The reaction was incubated at room temperature overnight under rotation. The next day, the supernatant was removed, and particles were resuspended in water. This was repeated five times to eliminate residual reagents. DNA probe-clicked IONPs in water were stored at 4 °C and found to be stable for a year.

**Nanoparticle characterization procedures.** To determine the density of DNA probes on the IONPs surface, the reaction was conducted as above, and the supernatant as well as three previous washes were collected right after the reaction was completed. DNA concentration was estimated using Quant-iT OliGreen ssDNA Assay Kit (Thermo Fisher Scientific).

IONP size and morphology were determined by field emission high-resolution scanning electron microscopy (SEM) with a Hitachi-S4800 HR instrument set at 30 kV and by transmission electron microscopy (TEM) imaging with a JEOL TEM-2200FS instrument. Samples were prepared by dropping the nanoparticle suspension on a 400-mesh carbon grid and drying it in a vacuum oven for 2 h. X-ray powder diffraction (XRD) patterns were collected using a Rigaku XRD Ultima IV instrument to study the structural properties of IONPs. X-ray photoelectron spectra (XPS) were taken on a Kratos AXIS 165 electron spectrometer with 150 W monochromatized Al K $\alpha$  radiation (1486.6 eV), whereby all peaks were referred to the signature C1s peak for adventitious carbon at 284.8 eV.

**Plasma samples from HCV patients (and controls).** Plasma samples in this study were collected from: (1) patient Pt804 (two time-points), positive for both HCC (Hepatocellular carcinoma) and HCV (Hepatitis C virus); (2) patient Pt339 (three time-points), positive for HCV; (3) a normal individual (Pt555). Viral copy numbers were determined by a procedure described in Steenbergen et al. 2010<sup>44</sup>. The Ethics Committee for the University of Alberta approved this investigation. Informed consent was obtained, and all experiments were performed in accordance with relevant guidelines/regulations.

**RNA extraction, reverse transcription, and library construction.** Total RNA was extracted with QIAamp viral RNA kit (Qiagen) from 1 mL plasma and cleaned up by RNeasy Mini kit with RNase-free DNase set (Qiagen). This was reverse transcribed to cDNA with Superscript II Reverse Transcriptase (Thermo Fisher Scientific) and random primers following standard protocols. Second strand cDNA was synthesized with a reaction mix (Tris, pH 7.8; 50 mM MgCl<sub>2</sub>; dNTP 10 mM; DTT 0.1 M; RNase H 2U/ $\mu\text{L}$ ; DNA Polymerase I 10U/ $\mu\text{L}$ ) at 16 °C for 2.5 h. Double-stranded (ds) cDNA was cleaned up by QIAquick PCR Purification kit (Qiagen) and eluted in 40  $\mu\text{L}$  ultrapure water (Gibco). The ds cDNA was sheared on a Covaris-S2 to an average size of 400 bp, followed by library construction with NEBNext Ultra II DNA library prep kit (NEB). HCV copy numbers were measured with TaqMan qPCR according to standard protocols.

**Design of probes and dsDNA target sequences.** Eight probes for different HCV regions (A to H) were selected to evaluate our InBeads against a commercial capture assay (Dynabeads MyOne streptavidin T1; Thermo Fisher Scientific). These 90-bp probe sequences (Tab. S1) were designed from the reference genome sequence (gi: 22,129,792) for HCV genotype 1 (excluding the repetitive U3 region) under the following criteria: (1) melting temperature 15–25°C higher than the 65°C hybridization temperature, i.e. 80–90°C; (2) GC content within 40% to 65%; (3) no significant similarity to the human genome sequence as assessed by blastn; (4) no stable secondary structure ( $\Delta\text{G}$  value higher than –9.0 kcal/mole) under hybridization conditions. We synthesized two sets of these 8 probes (IDT); one was modified with 5'-hexynyl (for InBeads use) and the other with 5'-biotin (for Dynabeads use). In addition, 270-bp dsDNA of HCV sequences (gblocks, covering both A and B in a single fragment) (Tab. S1) were synthesized by IDT and ligated with Illumina adaptors to be used as target fragments. These target fragments were serially diluted from 10<sup>6</sup> to 10 copies and supplemented with 100 ng of a non-HCV human DNA library.

**Target sequence enrichment using InBeads.** 100 ng of library DNA were mixed with blocking reagents (2.5  $\mu\text{g}$  of human Cot-1 DNA, 2.5  $\mu\text{g}$  salmon sperm DNA, 300 pmol blocking oligos complementary to library adapter sequences) and denatured at 95 °C for 5 min, then maintained at 65 °C. Hybridization buffer (10X SSPE, 10X Denhardt's solution, 10 mM EDTA, 0.2% SDS and 20U RNase Block Ribonuclease inhibitor) was freshly prepared and mixed with InBeads conjugated to equimolar amounts of each probe (20  $\mu\text{L}$  InBeads from

stock solution, storage solution removed before use), pre-warmed to 65 °C and transferred to the DNA library mix to create the hybridization mix. The final volume of the hybridization mix was 30 µL, which was then incubated at 65 °C for 24 h. The supernatant was removed and InBeads were washed 3X with buffer (0.1X SSC, 0.1% SDS) at 65 °C for 10 min each. Finally, InBeads were re-suspended in 30 µL ultra-pure water.

**Target sequence enrichment by streptavidin beads.** The pre-capture library mix (final volume 9 µL) was prepared by combining 100 ng DNA library with blocking reagents and denatured at 95 °C for 5 min, then maintained at 65 °C. Capture probe mix was prepared by mixing the 500 ng biotin-labeled probe set (pool of eight 5'-biotin-labeled probes at equal concentrations) with 20U RNase Block Ribonuclease Inhibitor and ultra-pure water to a final volume of 7 µL, pre-warmed to 65 °C for at least 2 min. Hybridization buffer (10X SSPE, 10X Denhardt's solution, 10 mM EDTA, 0.2% SDS and 20U RNase block) was freshly prepared and pre-warmed at 65 °C for at least 5 min. 13 µL hybridization buffer and 7 µL capture probe mix were rapidly added to the pre-capture library mix to make the hybridization mix in a final volume of 29 µL, and kept at 65 °C for 24 h. 50 µL of Dynabeads MyOne streptavidin T1 beads (Thermo Fisher Scientific) were washed twice with 200 µL binding buffer and re-suspended with 200 µL binding buffer. The hybridization mix was directly transferred to the beads solution and incubated at room temperature for 30 min under rotation, and then the supernatant was discarded. The beads were washed once with 500 µL wash buffer I (1X SSC, 0.1% SDS) at room temperature for 15 min, then 4X with 500 µL wash buffer II (0.1X SSC, 0.1% SDS) at 65 °C for 10 min. After removing the wash buffer, the beads were re-suspended with 30 µL of ultra-pure water.

**Post-capture qPCR following on-beads PCR.** Post-capture signal detection was a two-step process. First, we did an on-beads PCR using KAPA HiFi PCR kit (KAPA Biosystems) with the following program: 98 °C 10 min; 98 °C 20 s, 58 °C 15 s, 72 °C 30 s, 35 cycles; 72 °C 10 min; 4 °C forever; universally primed off the library construct with P5 primer (5' ACACTCTTCCCTACACGACGCTCTTCCGATCT 3') and P7 primer (5' GATCGGAAGAGCACACGTCTGAACTCCAGTC 3'). The PCR product was purified with the QIAquick PCR Purification kit (Qiagen), after which we performed a qPCR using SYBR Green PCR Mastermix (Thermo Fisher Scientific). Eight pairs of primers were designed (Tab. S1); for each pair, one primer was located in the on-target region, with the other in the flanking region, to ensure a product length in the 150–200 bp range and to avoid amplifying probe sequences. 2 µL of purified on-beads PCR product was used for each qPCR reaction in a final volume of 25 µL. Each qPCR reaction was carried out in duplicate in an ABI 3700 real-time PCR system (Applied Biosciences). The following thermocycling program was used: 95 °C 10 min, followed by 40 cycles of 95 °C 15 s, 60 °C 1 min. A final dissociation step was performed to determine the melting curve. The copy numbers were calculated based on a standard curve that was generated from a serially diluted plasmid with HCV genomic sequence (from 10<sup>6</sup> to 10 copies per µL).

Received: 3 February 2021; Accepted: 7 June 2021

Published online: 22 June 2021

## References

1. Sjöblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006).
2. Gasc, C., Peyretailade, E. & Peyret, P. Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. *Nucleic Acids Res.* **44**, 4504–4518 (2016).
3. Hodges, E. *et al.* Genome-wide in situ exon capture for selective resequencing. *Nat. Genet.* **39**, 1522–1527 (2007).
4. Albert, T. J. *et al.* Direct selection of human genomic loci by microarray hybridization. *Nat. Methods* **4**, 903–905 (2007).
5. Berensmeier, S. Magnetic particles for the separation and purification of nucleic acids. *Appl. Microbiol. Biotechnol.* **73**, 495–504 (2006).
6. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**, 182–189 (2009).
7. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. USA* **106**, 19096–19101 (2009).
8. Sulonen, A.-M. *et al.* Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol.* **12**, R94 (2011).
9. Ross, J. P., Dion, P. A. & Rouleau, G. A. Exome sequencing in genetic disease: recent advances and considerations. *F1000Res* **9**, 336 (2020).
10. Bartha, Á. & Györfy, B. Comprehensive outline of whole exome sequencing data analysis tools available in clinical oncology. *Cancers* **11**, 1725 (2019).
11. Yang, Y. *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* **369**, 1502–1511 (2013).
12. Meng, L. *et al.* Use of Exome Sequencing for Infants in Intensive Care Units: Ascertainment of Severe Single-Gene Disorders and Effect on Medical Management. *JAMA Pediatr.* **171**(12), e173438. <https://doi.org/10.1001/jamapediatrics.2017.3438> (2017).
13. Kumar-Sinha, C. & Chinnaiyan, A. M. Precision oncology in the age of integrative genomics. *Nat. Biotechnol.* **36**, 46–60 (2018).
14. Roychowdhury, S. *et al.* Personalized oncology through integrative high-throughput sequencing: A pilot study. *Sci. Transl. Med.* **3**, 111–121 (2011).
15. Saleheen, D. *et al.* Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* **544**, 235–239 (2017).
16. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
17. Cieslik, M. *et al.* The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res.* **25**, 1372–1381 (2015).
18. Picardi, E. & Pesole, G. Mitochondrial genomes gleaned from human whole-exome sequencing. *Nat. Methods* **9**, 523–524 (2012).
19. Vezzulli, L. *et al.* Whole-genome enrichment provides deep insights into vibrio cholerae metagenome from an African River. *Microb. Ecol.* **73**, 734–738 (2017).



20. Betin, V. *et al.* Hybridization-based capture of pathogen mRNA enables paired host-pathogen transcriptional analysis. *Sci. Rep.* **9**, 19244 (2019).
21. Briese, T. *et al.* Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio* **6**, e01491-e1515 (2015).
22. Forth, J. H. *et al.* A deep-sequencing workflow for the fast and efficient generation of high-quality african swine fever virus whole-genome sequences. *Viruses* **11**, 846 (2019).
23. Wylie, T. N., Wylie, K. M., Herter, B. N. & Storch, G. A. Enhanced virome sequencing using targeted sequence capture. *Genome Res.* **25**, 1910–1920 (2015).
24. Yeh, Y.-T. *et al.* A rapid and label-free platform for virus capture and identification from clinical samples. *Proc. Natl. Acad. Sci. USA* **117**, 895–901 (2020).
25. Schiessl, S., Samans, B., Hüttel, B., Reinhard, R. & Snowdon, R. J. Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. *Front. Plant Sci.* **5**, 404 (2014).
26. Tennessen, J. A., Govindarajulu, R., Ashman, T.-L. & Liston, A. Evolutionary origins and dynamics of octoploid strawberry sub-genomes revealed by dense targeted capture linkage maps. *Genome Biol. Evol.* **6**, 3295–3313 (2014).
27. Faucon, F. *et al.* Identifying genomic changes associated with insecticide resistance in the dengue mosquito *Aedes aegypti* by deep targeted sequencing. *Genome Res.* **25**, 1347–1359 (2015).
28. Jube, F. *et al.* Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* **76**, 530–544 (2013).
29. Paul, A., Avci-Adali, M., Ziemer, G. & Wendel, H. P. Streptavidin-coated magnetic beads for DNA strand separation implicate a multitude of problems during cell-SELEX. *Oligonucleotides* **19**, 243–254 (2009).
30. Dederich, D. A. *et al.* Glass bead purification of plasmid template DNA for high throughput sequencing of mammalian genomes. *Nucleic Acids Res.* **30**, e32–e32 (2002).
31. Bruce, I. J. & Sen, T. Surface modification of magnetic nanoparticles with alkoxysilanes and their application in magnetic bioseparations. *Langmuir* **21**, 7029–7035 (2005).
32. Tan SC, Yiap BC. DNA, RNA, and protein extraction: the past and the present. *J Biomed Biotechnol.* 2009:574398. <https://doi.org/10.1155/2009/574398> (2009).
33. Walsh, M. K., Wang, X. & Weimer, B. C. Optimizing the immobilization of single-stranded DNA onto glass beads. *J. Biochem. Biophys. Methods* **47**, 221–231 (2001).
34. Cutler, J. I., Zheng, D., Xu, X., Giljohann, D. A. & Mirkin, C. A. Polyvalent oligonucleotide iron oxide nanoparticle ‘click’ conjugates. *Nano Lett.* **10**, 1477–1480 (2010).
35. Messina, J. P. *et al.* Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology* **61**, 77–87 (2015).
36. Griffin, H. R. *et al.* Accurate mitochondrial DNA sequencing using off-target reads provides a single test to identify pathogenic point mutations. *Genet. Med.* **16**, 962–971 (2014).
37. Nasir, J.A. *et al.* Rapid Design of a Bait Capture Platform for Culture- and Amplification-Free Next-Generation Sequencing of SARS-CoV-2. Preprints 2020, 2020020385. <https://doi.org/10.20944/preprints202002.0385.v1>.
38. Li, B. *et al.* Discovery of bat coronaviruses through surveillance and probe capture-based next-generation sequencing. *mSphere* **5**, 170 (2020).
39. Thomson, E. *et al.* Comparison of next-generation sequencing technologies for comprehensive assessment of full-length hepatitis C viral genomes. *J. Clin. Microbiol.* **54**, 2470–2484 (2016).
40. Bonsall, D. *et al.* ve-SEQ: Robust, unbiased enrichment for streamlined detection and whole-genome sequencing of HCV and other highly diverse pathogens. *Research* **4**, 1062 (2015).
41. Tian, Y., Yu, B., Li, X. & Li, K. Facile solvothermal synthesis of monodisperse Fe<sub>3</sub>O<sub>4</sub> nanocrystals with precise size control of one nanometre as potential MRI contrast agents. *J. Mater. Chem.* **21**, 2476–2481 (2011).
42. Stöber, W., Fink, A. & Bohn, E. Controlled growth of monodisperse silica spheres in the micron size range. *J. Colloid Interface Sci.* **26**, 62–69 (1968).
43. Schatz, A., Hager, M. & Reiser, O. Cu(II)-Azabis(oxazoline)-complexes immobilized on superparamagnetic magnetite@Silica-nanoparticles: A highly selective and recyclable catalyst for the kinetic resolution of 1,2-diols. *Adv. Funct. Mater.* **19**, 2109–2115 (2009).
44. Steenbergen, R. H. G. *et al.* Lipoprotein profiles in SCID/uPA mice transplanted with human hepatocytes become human-like and correlate with HCV infection success. *Am. J. Physiol. Gastrointest. Liver Physiol.* **299**, G844–G854 (2010).

## Acknowledgements

This work was financially supported by grants from the Government of Alberta through Ingenuity Lab, and through an iCORE Strategic Chair (RES0010334) to G.K.S.W. We thank Hui Qian for providing us the HR-TEM pictures in this work.

## Author contributions

W.W., C.M., and G.K.S.W. conceived the project. F.D., W.Z.S., and S.C. synthesized the InBeads. W.W. and T.J. performed the capture experiments. F.D., W.W., W.Z.S., J.J., and G.K.S.W. wrote the paper. All authors read and approved the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-92376-9>.

**Correspondence** and requests for materials should be addressed to G.K.-S.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021