Automated platform for high-throughput screening of base-modified aptamers for affinity and specificity

Diana Wu[†1], Trevor Feagin[†2], Peter Mage[†3], Alexandra Rangel[2], Leighton Wan[1], Anping Li[2], John Coller[4], Michael Eisenstein[2, 3], Sharon Pitteri[2], H. Tom Soh[2, 3]

Affiliations:

[1]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA.

[2]Department of Radiology, Stanford University, Stanford, CA 94305, USA.

[3]Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA.

[4]Stanford Functional Genomics Facility, School of Medicine, Stanford University, Stanford, CA 94305, USA.

[†]*authors contributed equally*

*Correspondence to: tsoh@stanford.edu

Summary

Aptamers incorporating chemically modified bases can achieve superior affinity and specificity compared to natural aptamers, but their discovery remains a labor-intensive, low-throughput task. Here we describe the 'non-natural aptamer array' (N2A2) system, which enables fully automated, high-throughput screening of base-modified aptamers using a minimally modified Illumina MiSeq instrument. We demonstrate the capability to screen multiple different base modifications to identify the optimal choice for high-affinity target binding. We further use N2A2 to generate aptamers that specifically recognize protein posttranslational modifications, and which maintain strong target affinity in serum. Finally, we demonstrate comprehensive profiling of single- and double-base aptamer mutations to rapidly identify key sequence motifs responsible for binding activity in a single run. N2A2 requires only minor mechanical modifications to the MiSeq and a software suite for automation that we have made freely available. As such, we believe our platform offers a broadly accessible and user-friendly tool for generating custom reagents on-demand.

Introduction

Aptamers made of DNA and RNA have shown the capability to bind many types of biomolecules with excellent affinity and specificity.[1,2] However, due to the limited chemical repertoire of natural nucleic acids, the range of biomolecules that aptamers can recognize pales in comparison to protein-based reagents. Non-natural aptamers that incorporate chemically modified bases can potentially overcome this problem. For example, the pioneering work of Gold and coworkers has shown that the addition of hydrophobic groups to the aptamer structure greatly enhances the affinity and specificity of aptamers for targeting proteins. They showed that such base-modified aptamers could be generated with high affinities (K_d < 10 nM) for ~84% of the human proteome, versus the < 30% that were suitable targets for natural aptamers.[3] In principle, base-modified aptamers can incorporate a myriad of chemical functional groups beyond those found in nature. This offers the exciting potential for generating affinity reagents for a broader spectrum of proteins as well as non-protein analytes that would be difficult or impossible to target using conventional antibodies, including small molecules, post-translational modifications, and complex carbohydrate structures.[4]

To date, there have been two main approaches for generating base-modified aptamers. The first uses engineered polymerase enzymes to incorporate chemically-modified nucleotides directly into the aptamer sequence.[5] Although several groups have achieved success on this front,[6,7] this strategy can result in compromises in terms of the fidelity, processivity or stability of the resulting enzyme. Furthermore, every new chemical modification is likely to require a new round of polymerase engineering, and some modifications will simply be too bulky to be accommodated by any polymerase enzyme. An alternative strategy, first described by Tolle *et al*, uses a copper-catalyzed azide-alkyne cycloaddition (CuAAC) 'click chemistry' reaction[8,9] to couple azide-modified functional groups onto alkyne-modified nucleobases (or vice versa).[10] This approach is compatible with commercially available enzymes and allows researchers to incorporate modified bases bearing virtually any functional group. However, it still represents a considerable time investment to identify the appropriate base modification for a given target, and the synthesis and screening of base-modified aptamer libraries can be a daunting and labor-intensive task, particularly for laboratories that are not already specialized in aptamer discovery. This difficulty is reflected in the fact that only a few base-modified aptamers have been reported in the

literature,[11-13] and represents a serious constraint on the research community's capacity to efficiently generate high-quality affinity reagents for a truly broad spectrum of targets.

As a solution to this problem, we have developed an integrated system for the automated generation, screening, and characterization of base-modified aptamers. Our 'non-natural aptamer array' (N2A2) system uses a modified version of a benchtop high-throughput DNA sequencing instrument – the Illumina MiSeq – to characterize binding of ~10^7 base-modified aptamers in parallel. N2A2 employs a click chemistry-based modification approach that enables efficient incorporation of virtually any chemical functional group with commercially available polymerase enzymes. In each experimental run, sequence-defined libraries of base-modified aptamers are directly synthesized at assigned locations within the flow-cell, and then incubated with different concentrations of labeled target, with binding measured directly from each base-modified aptamer cluster. This approach allows users to readily screen a variety of different functional groups in order to identify the best chemical modification for the target. Although other groups have repurposed sequencing instruments for the screening of natural DNA and RNA aptamers,[14,15,16] N2A2 represents the first automated system that enables screening of base-modified aptamers. And critically, N2A2 is straightforward to implement for any lab with access to a MiSeq, requiring only minor mechanical modifications to the instrument (*i.e.,* a few tubing changes) and a suite of custom software that we have made freely available.

As an initial demonstration of this platform's capabilities, we used N2A2 to rapidly identify which chemical modification confers superior binding to aptamers for vascular endothelial growth factor (VEGF). This enabled us to isolate a tryptophan-modified aptamer with significantly better affinity than the best VEGF aptamer reported to date. N2A2 also enables simultaneous screening for specificity as well as affinity by performing runs in which base-modified aptamers are challenged with either a counter-target for negative selection or screened in a complex and interferent-rich sample matrix. We first exploit this capability to isolate base-modified aptamers that can discriminate targets based on post-translational modifications—in this case, glycosylation. Such modifications can profoundly alter the functional characteristics of a protein, and the selection of modification-specific antibodies generally remains a difficult task. Specifically, we screened a tryptophan-modified library to isolate a base-modified aptamer that can bind to sialic acid-modified fetuin but not to asialofetuin, which lacks this post-translational modification. We also generated a highly specific phenylalanine-modified aptamer for the hormone insulin by

screening in diluted serum, obtaining molecules that exhibit superior binding performance to a previously reported aptamer, which no longer exhibits measurable target affinity in this complex background. Finally, we show that N2A2 can efficiently characterize the contributions of individual nucleotides to secondary/tertiary structure and target binding for base-modified aptamers by generating interaction 'landscapes' for comprehensive sets of single- and double-mutant sequences. Our results suggest that it should be possible to use this data to further tune and optimize the binding of newly identified base-modified aptamers. We believe that N2A2 offers a simple, user-friendly approach for rapidly generating high-quality base-modified aptamers without specialized expertise or infrastructure, and thus has the potential to greatly accelerate biomedical research by enabling many research labs to generate custom reagents on demand.

Results and Discussion

Overview of the non-natural aptamer array (N2A2) platform

N2A2 offers an integrated automated workflow for the parallel screening of ~10^7 unique base-modified aptamer candidates in terms of both affinity and specificity. The N2A2 was fabricated by converting the Illumina MiSeq instrument such that base-modified aptamer clusters can be synthesized and characterized directly on the sequencing flow-cell (**Figure 1A**). The workflow entails three main steps: sequencing, conversion, and *in situ* screening (**Figure 1B)**. It begins with sequencing of a DNA library; during the paired-end turnaround and second read of the sequencing process,[17,18] the DNA clusters are converted into base-modified aptamer clusters. The flow-cell is then incubated with fluorescently labeled protein to screen for affinity and specificity, with intensity information collected for each cluster. These sequencing and screening data are processed to generate a phenotype-genotype linked map of all the clusters.

The initial libraries comprise a randomized region flanked by a forward primer (FP) and reverse primer complementary (RPc) sequence. We pre-enriched our initial library for binding to the target using one or two rounds of screening with conventional selection methods in order to reduce the pool's diversity to a scale that can effectively be screened on the N2A2. The pre-enriched aptamer pool is amplified with a pair of adaptor sequences (**Table S1**), producing a library of aptamers that feature an EcoRI recognition sequence adjacent to the RPc sequence (**Figure 1C**).

In a separate amplification step, Nextera XT indices that include the Illumina-defined flow-cell primers are also added onto the 5' and 3' ends of each library molecule. Once the library is prepared, we perform paired-end sequencing of the aptamer pool with a 150-cycle V3 MiSeq kit. During this process, the aptamer pool is displayed on the flow-cell as antisense strands, with a 5' reverse primer sequence and 3' forward primer complementary sequence. Sequencing produces a FASTQ file that represents these sequences as sense strands.

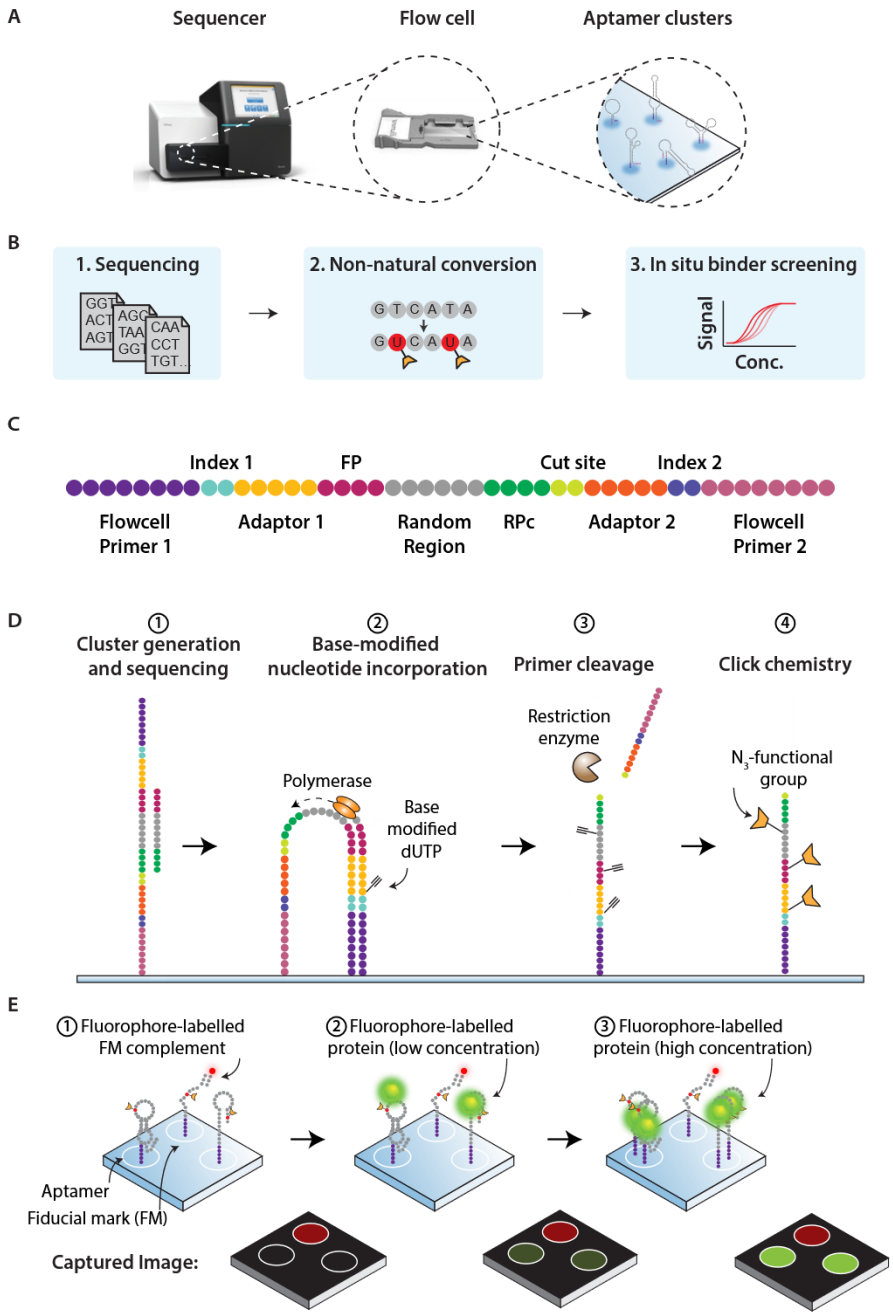

Figure 1. Overview of N2A2. **A**) N2A2 uses a MiSeq flow-cell to generate base-modified aptamer clusters and screen their performance. **B**) The workflow begins with sequencing of a DNA library. During this process, the DNA clusters are converted into base-modified aptamer clusters, which are then incubated with fluorescently-labeled protein to screen for affinity and specificity, with intensity information collected for each cluster. **C**) The initial DNA pool, which comprises a randomized region (gray) flanked by primer-binding sequences (magenta, dark green), is subjected to two amplification steps to introduce Illumina flow-cell primer sites (pink, purple), Illumina 'adaptor' sequencing primers (gold, orange), Illumina-defined indices (teal, dark blue), and an EcoRI cut site (light green). **D**) The base-modified aptamer synthesis process. (1) DNA is covalently attached to the flow-cell, natural DNA clusters are formed, and sequencing by synthesis is performed. (2) During the paired-end turnaround, C8-alkyne dUTP is incorporated into the DNA by KOD-XL polymerase. (3) Flow-cell primer 2, index 2, and adaptor 2 sequences are cleaved from the clusters by EcoRI. (4) An azide-tagged functional group is conjugated to the alkyne handle via click chemistry. **E**) Assessing binding on the flow-cell surface. (1) Base-modified aptamer clusters are incubated with a fluorescently-labeled sequence complementary to a fiducial mark DNA, which is included in the assay as a positive control, and then imaged. (2) The aptamers are first incubated with a low concentration of fluorescently-labeled protein and imaged, and then (3) this process is repeated with higher concentrations of fluorescently-labeled protein.

The conversion of the DNA into base-modified aptamers is achieved through a four-step process (**Figure 1D**). The first read of sequencing generates the natural DNA template and FASTQ data (step 1). Second, during the paired-end turnaround step, C8-alkyne-dUTP is substituted for native dTTP during bridge PCR using the KOD-XL polymerase, which can incorporate this modification with high fidelity (step 2).[19,20] This produces aptamer clusters with alkyne click handles on the flow-cell, which are compatible with post-synthesis modification via a click chemistry reaction.[8,9] This enables us to modify every T with virtually any chemical functional group that we wish to incorporate. Third, we use EcoRI digestion to remove the adaptor and flow-cell primer sequences that were added to the 3' end of the aptamers during cluster formation in order to prevent potential steric hindrance between the aptamer and the target (step 3). As mentioned above, our library incorporates an EcoRI recognition sequence between the aptamer and the adaptor 2 primer; by hybridizing a complementary strand to this primer and the EcoRI recognition sequence, we form a double-stranded cut-site that enables enzymatic excision of the 3' sequencing primer, index, and flow-cell primer sequences. Finally, during the second read of the paired-end sequencing process, the desired chemical modifications are conjugated to the aptamer through click chemistry (step 4).

Importantly, the entire process of converting DNA into base-modified aptamers is completely automated, with no manual intervention. This was achieved by editing the XML files to direct the instrument to incorporate additional reagents (*i.e.*, non-natural dNTPs, KOD-XL

enzyme, EcoRI enzyme, and buffer) stored in the custom primer tubes in the sequencing cartridge. Details of the modifications to the XML files are provided in the SI. Only modest hardware modifications are required, as shown in **Figure S1.** Control experiments for converting DNA into base-modified aptamers are shown in **Figure S2**.

Once the conversion process is complete, we perform affinity screening *in situ* within the flow-cell. A single fiducial mark sequence is included in the library of DNA that has been sequenced. We incubate the flow-cell with a fluorescently-labeled fiducial mark complement oligo, which serves as a positive binding control in all steps of the experiment (**Figure 1E**, step 1). Next, we incubate the aptamer clusters on the flow-cell with fluorescently-labeled target proteins (**Figure 1E**, step 2), wash with buffer, and use the built-in imaging system of the MiSeq instrument to capture images of the flow-cell. These images provide the fluorescence intensity of all clusters simultaneously, which is representative of the amount of target bound. Next, we increase the concentration of the target protein, wash, and image, repeating this process until we have a binding curve at multiple concentrations, which represents the binding affinity (**Figure 1E**, step 3). As we demonstrate in later sections, it is also possible to perform this process in complex matrices such as serum, which allows us to measure the specificity as well as the affinity of individual base-modified aptamers on the flow-cell.

We have developed custom software tools to directly link the sequence of every base-modified aptamer on the flow-cell with its affinity measurements. This is achieved by creating a 1:1 mapping between the aptamer sequence and the fluorescence intensity profile of its cluster across a range of target concentrations. Briefly, we extract the internal binary files produced by the MiSeq software during image analysis. The MiSeq's internal algorithm performs background normalization and Gaussian intensity fitting to convert raw cluster images to mapped cluster intensities. These are organized into .locs files, which contain the unique physical address of each cluster (expressed as the cluster's tile on the flow-cell and *x*/*y*-coordinates on that tile), and .cif files, which contain the extracted fluorescence intensity of each cluster in each of the four fluorescence channels. Together, these files provide an intensity map of every cluster on the flow-cell. Next, we extract the tile and *x*/*y*-coordinates for each sequence as provided in the output files using the .fastq format. In this way, we obtain both sequence and location information for each cluster on the flow-cell. Finally, we cross-reference the intensity map from the .locs and .cif files

with the sequence map obtained from the .fastq file in order to create a linked 1:1 sequence-intensity map. This yields a direct link between an aptamer's sequence and its intensity at a given target concentration.

N2A2 enables screening of multiple chemical modifications

The automated design of N2A2 makes it straightforward to rapidly compare different chemical modifications in order to identify the optimal choice for a given target molecule. As a demonstration, we assessed how two different chemical modifications— the amino acids tyrosine (Y) and tryptophan (W)—impact the affinity of base-modified aptamers for VEGF. We conducted three separate screens, assessing VEGF binding to natural DNA, Y-modified, and W-modified aptamers. We chose Y and W modifications because they are known to critically contribute to epitope recognition in antibodies.[21] It should be noted that although thymine was replaced with Y- or W-conjugated uracil during our screen, the modified bases will be referred to here as 'T' for ease of reading.

We first pre-enriched a random library of 10^{14} sequences by performing a single round of traditional SELEX followed by a single round of particle display[22] (**Figure S3**). The resulting nucleic acid pool was then screened on N2A2, either as natural DNA or base-modified DNA. In each experiment, we also included a positive control sequence – the published VEGF aptamer SL2B (K_d = 0.50 ± 0.32 nM)[23,24] – and a negative control sequence that does not bind VEGF (**Table S1**). To minimize the effect of measurement noise, we averaged the fluorescence intensity of aptamers represented by more than 10 clusters with identical sequences on the flow-cell. The fluorescence measurements for each sequence are shown as a function of VEGF concentration in **Figure 2A–C** for natural, Y-modified, and W-modified DNA, respectively.

W-modified aptamers exhibited considerably higher affinity for VEGF compared to natural DNA or Y-modified aptamers. For example, we clearly observed binding at 100 pM VEGF with the W-modified aptamers (**Figure 2C**), whereas binding was not detected at concentrations below 10 nM VEGF for the natural DNA aptamers (**Figure 2A)**. We also observed increased background binding with the two modified libraries, as indicated by the intensity of the modified negative control sequence (orange regions in **Figure 2B, C**), which is a consequence of the general affinity-enhancing characteristics of these modifications in T-rich sequences (**Figure S4)**.

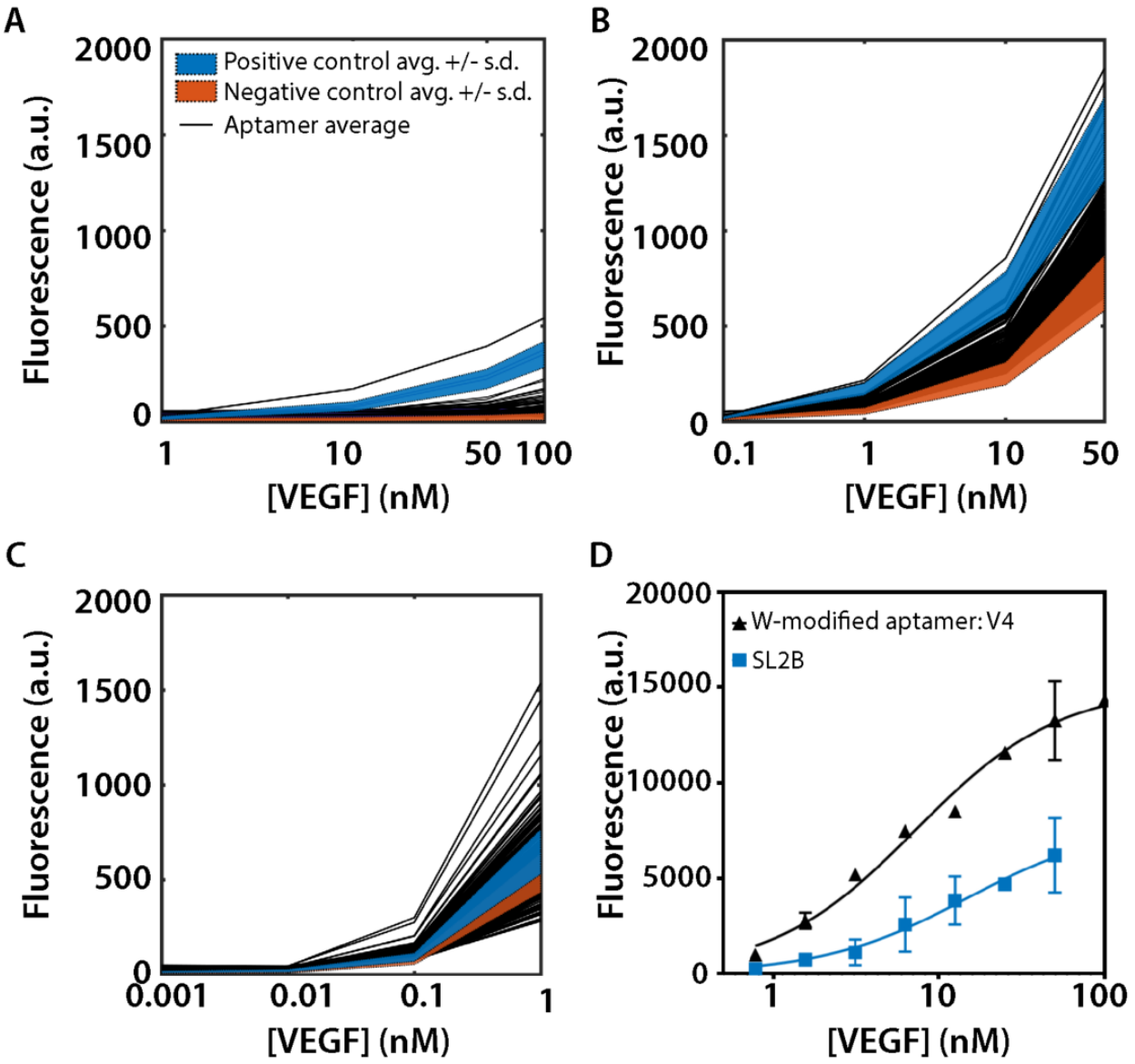

Figure 2. N2A2 screening of aptamer library binding to fluorescently-labeled VEGF. Binding curves are shown for **A**) natural DNA, **B**) tyrosine (Y)-modified DNA and **C**) tryptophan (W)-modified DNA. Black lines show mean binding curves for all unique aptamer candidates. Blue shaded region indicates intensity from the positive control aptamer ± 1 S.D, and orange shaded region shows background intensity from the negative control sequence ± 1 S.D. **D**) Flow cytometry characterization of a strong VEGF binder identified by N2A2. W-modified aptamer V4 (black) shows greater affinity for VEGF than the positive control aptamer, SL2B (blue).

We next examined the performance of base-modified aptamer families and individual rare sequences. We identified unique sequences that appeared in all three N2A2 experiments, and used the FASTAptamer software[25] to group sequences with a Levenshtein edit distance of 6 or less into families. We characterized sequences from families 1, 3, 10, 14 and 341, since they represent

sequences that exhibit either weak (10), intermediate (1 and 3), or strong (14 and 341) binding at 1 nM VEGF relative to the negative control. The intensity data for all individual sequences and families in the tryptophan experiment are plotted in **Figure S5A** and **B**. We then synthesized consensus sequences for each of the five families to test as aptamer candidates.

To study rare individual aptamers, we applied stringent filtering to our aptamer sequences. Starting again from the unique sequences that appeared in all three N2A2 experiments, we filtered according to two rules: 1) the binding intensity at 0 nM VEGF had to be < 50 a.u. to eliminate false positives, and 2) the binding intensity at the highest concentration of VEGF had to be greater than the average intensity of the positive control aptamer at the same concentration. We further narrowed these candidates down to aptamers for which the binding intensity at the highest VEGF concentration exceeded the average fluorescence intensity of the positive control plus one standard deviation. This extremely stringent rule left us with a high-quality aptamer subset representing well below 1% of the total number of unique sequences, a subpopulation that most likely would not have been found without the screening capabilities of the N2A2 (**Table S2**). We synthesized three rare sequences with the highest fluorescence intensity, and which did not belong to one of the five families already being characterized.

We measured the affinity of each of these eight W-modified aptamer candidates via flow cytometry at 1 pM–100 nM VEGF to create full binding curves. Curve fitting was performed using nonlinear regression in Prism, and K_d measurements are reported in **Table S3** for each candidate. The highest-affinity aptamers from the family (V4) and rare sequence (V8) subsets respectively exhibited a calculated K_d of 2.8 nM ± 0.63 and K_d of 4.6 nM ± 0.6. In comparison, positive control aptamer SL2B exhibited a K_d of 18.7 nM in the same assay (**Figure 2D**). We hypothesize that our K_d measurement of the published aptamer is ~40 fold higher than the reported K_d due to differences in steric hindrance associated with our measurement techniques. In the initial publication describing this aptamer, the authors used an immobilized protein and surface plasmon resonance, whereas we used an immobilized aptamer and flow cytometry. These findings demonstrate that N2A2 can efficiently identify the optimal aptamer base modification for a given target, and isolate multiple aptamer candidates with low nanomolar target affinity in just a single automated run.

N2A2 enables screening for specificity

N2A2 can also be used to screen base-modified aptamers for specificity. In contrast to conventional aptamer screening procedures, which require counter-selection steps that add considerable time and labor and introduce further opportunities for bias,[26] N2A2 enables us to simultaneously screen for affinity and specificity directly in a single experiment. As an initial demonstration of this capability, we developed a base-modified aptamer that can sensitively discriminate between two glycoforms of the same protein. Post-translational modifications such as glycosylation can profoundly alter the function of a protein, but it remains challenging to isolate affinity reagents that exhibit robust affinity and specificity for a particular proteoform. However, there is considerable value in developing such reagents—for example, a number of tumor-associated proteins display aberrant glycosylation profiles that could be exploited as diagnostic biomarkers or therapeutic targets.[27] To demonstrate that our platform can achieve the specificity needed to discriminate between glycoforms, we used N2A2 to screen for tryptophan-modified aptamers that recognize fetuin but not asialofetuin. Fetuin is a glycoprotein found in the blood that is involved in the transport of diverse molecular cargos, including free fatty acids. It is decorated with sialylated N-linked and O-linked glycans[28] , whereas asialofetuin is an alternative glycoform of this protein that is structurally similar to fetuin but does not contain sialic acid groups. We first performed a pre-enrichment procedure to expand the starting population of aptamer candidates. Briefly, this consisted of four rounds of positive (with fetuin) and negative SELEX (with asialofetuin) and two rounds of particle display—one with natural DNA, and one with W-modified DNA. We chose this modification because there is evidence that the aromatic ring of this amino acid can interact with saccharides.[29]

We then used N2A2 to screen aptamer binding to 100 pM, 1 nM, 10 nM, 100 nM and 1 μM fluorescently labelled fetuin and asialofetuin in series. Then we ranked the affinities of the aptamers on the array based on fluorescence intensity (**Figure 3A**) and selected top seven candidates (**Table S4**). Next, we tested the specificity of these candidates by flow cytometry and found that three of our seven candidates showed binding for fetuin but not for asialofetuin **(Figure S6).** One of these three aptamers, fet4, showed excellent specificity, with a K_d of 3 μM for fetuin but minimal binding to asialofetuin, such that a meaningful K_d could not be determined (**Figure 3B**).

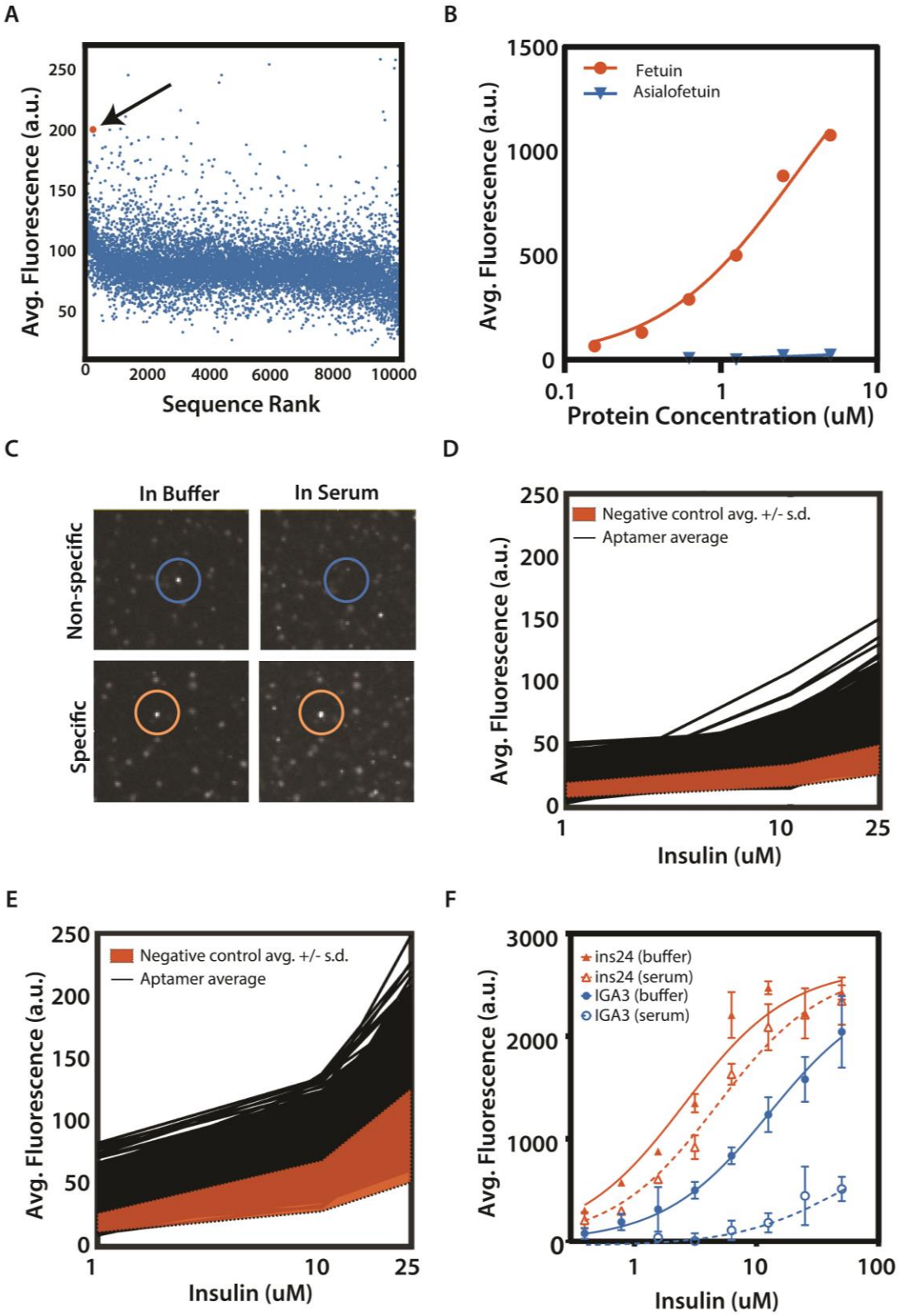

Figure 3 N2A2 can generate base-modified aptamers with excellent specificity. **A**) N2A2 measurements of fluorescence intensities of individual aptamer candidates with 1 μM fetuin. Sequences are ranked by

their fluorescence intensity at 100 nM. Arrow indicates fet4. **B**) Flow cytometry binding analysis of fet4 displayed on particles and incubated with fluorescently-labeled fetuin and asialofetuin. Average fluorescence intensity at each protein concentration is shown. **C**) N2A2 enables direct differentiation of highly target-specific aptamer clusters. A non-specific aptamer (top) shows strong insulin binding in buffer (left), but this signal disappears in serum (right). In contrast, a high-specificity aptamer (bottom) shows strong insulin binding in both buffer and serum. **D**) N2A2 screening of phenylalanine-modified aptamer library binding to fluorescently labeled insulin in buffer and **E**) 1% human serum. Black lines show the average fluorescence of each unique aptamer sequence. Orange region shows the negative control +/- 1 S.D. **F**) A flow cytometry-based bead-binding assay for the previously published aptamer IGA3 (blue) and the N2A2-identified aptamer ins24 (red) in buffer (dotted lines) and 1% human serum (solid lines). Only ins24 retains target binding in serum.

Next, we isolated a base-modified aptamer that can specifically bind to insulin in diluted human serum, a characteristic that would be essential in the context of a clinical detection assay. We selected a phenylalanine modification based on the crystal structure of the insulin receptor,[21] which reveals that phenylalanine plays an important role in the hydrophobic pocket of the insulin-binding site. To identify high-specificity insulin aptamers, we screened a phenylalanine-modified aptamer library in buffer and 1% human serum on the N2A2. We performed pre-enrichment (one round each of positive and negative SELEX) to ensure we had multiple copies of each aptamer on the flow-cell. **Figure 3C** shows how nonspecific and specific aptamers can readily be discriminated with N2A2; when aptamers are non-specific, sequence clusters that fluoresce in buffer conditions disappear in serum, whereas target-specific clusters retain the same binding intensity in both conditions. We then measured the fluorescence intensity of the resulting base-modified aptamer clusters when incubated with 1, 10, and 25 μM insulin in buffer and 1% human serum (**Figure 3D** and **E**).

Our phenylalanine-modified library showed binding to fluorophore-labeled insulin in both buffer (**Figure 3D**) and 1% human serum (**Figure 3E**). Unexpectedly, the fluorescence intensity for both the aptamer library and the negative control sequence were higher when serum was included in the experiment. We hypothesize that the serum is acting as a blocking agent on the flow-cell. Under buffer conditions, the insulin could bind to other areas on the flow-cell and raise the background fluorescence, while under serum conditions, the flow-cell is coated with unlabeled serum proteins, such that the aptamer clusters appear brighter. As with our VEGF aptamers, we grouped our sequences into families based on a Levenshtein edit distance of 2 or less, which we then ordered by highest copy number (**Figure S7**). We then identified sequence families with greater relative mean intensities for 1 μM insulin in both buffer and 1% serum. Of the 120 families

with at least 10 members, families 1, 10, 17, and 43, showed high mean intensities relative to the other families in both conditions. We also defined 'specificity fold-change' as a heuristic to identify families that have high specificity (**Figure S8**) and determined that families 1 and 10 stood out by this metric. The consensus sequences of families 1, 10, 17, and 43 were therefore chosen as aptamer candidates (**Table S5).**

The resulting four aptamer candidates—ins19, ins20, ins22 and ins24—showed strong affinity for insulin in a flow cytometry-based fluorescence assay under serum conditions (**Figure S9**). As a positive control in our flow cytometry experiments, we employed a previously published DNA aptamer for insulin (IGA3) described by Yoshida *et al.*[30] We compared the affinities of our base-modified aptamer and IGA3 via flow cytometry, and measured a roughly comparable K_d of 2.56 μM and 12.7 μM for the two aptamers, respectively (**Figure 3F**). However, only ins24 retained this strong affinity for its target when we performed the same assay in serum. Our aptamer exhibited nearly unchanged insulin binding in these conditions, with a K_d of 4.8 μM, whereas we were unable to obtain a meaningful K_d measurement for IGA3 in the same conditions (**Figure 3F,** dotted blue line). Insulin is a polypeptide hormone made up of 51 amino acids (molecular weight 5.8 kDa),[31,32] and the affinity of ins24 is typical for aptamers to polypeptide targets, which generally falls within the high nM to low μM range.[33,34] However, it remains notable that this aptamer achieves such affinity in the context of human serum. These results therefore confirm our capacity to efficiently identify base-modified aptamers that exhibit both excellent affinity and specificity for their target in complex biological specimens.

Mutational analysis with N2A2 offers insights into binding mechanism

Aptamer binding site characterization can be a challenging task, typically requiring extensive and laborious truncation and point-mutant screens.[35] Less than ten crystal structures have been resolved for aptamer-target complexes,[36] and it has proven even more difficult to generate structural information about base-modified aptamers. N2A2 offers a highly efficient alternative to conventional mutagenesis assays, enabling the generation of comprehensive functional sequence-fitness landscapes for a myriad of base-modified aptamer variants. The deep characterization of mutants in this way allows us greater insight into how a base-modified aptamer interacts with its target. Drawing inspiration from chemical accessibility profiles[37] and the mutate-and-map

strategy,[38,39] we can use the resulting fitness heatmaps to infer areas of structural and functional importance.

As a demonstration, we explored the full affinity landscape of all single- and double-nucleotide mutants of ins24 in a single N2A2 run. We sequenced and characterized the binding performance of a custom synthesized pool of 7,141 unique aptamer sequences towards insulin in serum. **Figure 4A** shows a heatmap of the normalized fluorescence intensity readings for every single-nucleotide variant of ins24 with 1.6 μM insulin in 1% serum. Combining this information with secondary structure prediction[40] for canonical base-pairing, we note that these mutations fall into four categories: 1) no change in secondary structure but improved binding (**Figure 4B**, red), 2) no change in secondary structure but decreased binding (**Figure 4B**, blue), 3) change in secondary structure and increased binding (**Figure 4C**, red), and 4) change in secondary structure and decreased binding (**Figure 4D**, blue).

The genotype-phenotype linkage can offer insights into why certain mutations improve or decrease insulin binding. For example, when position 18 is changed from G to T, the heatmap shows that the aptamer binds insulin with higher affinity. We noted that this mutation is predicted to result in a change in the aptamer secondary structure from two stem loops within the aptamer's variable sequence (**Figure 4B**) to a single stem-loop (**Figure 4C**). On the other hand, when position 13 is mutated from a C to a G, we see a decrease in binding. In the predicted secondary structure, we see that one of the two stem loops in the variable region has shifted closer to the other, migrating towards the 3' end of the aptamer (**Figure 4D**). These results strongly indicate that that these two bases play an important role in the secondary and tertiary structure of the aptamer, and may even have direct contact with insulin.

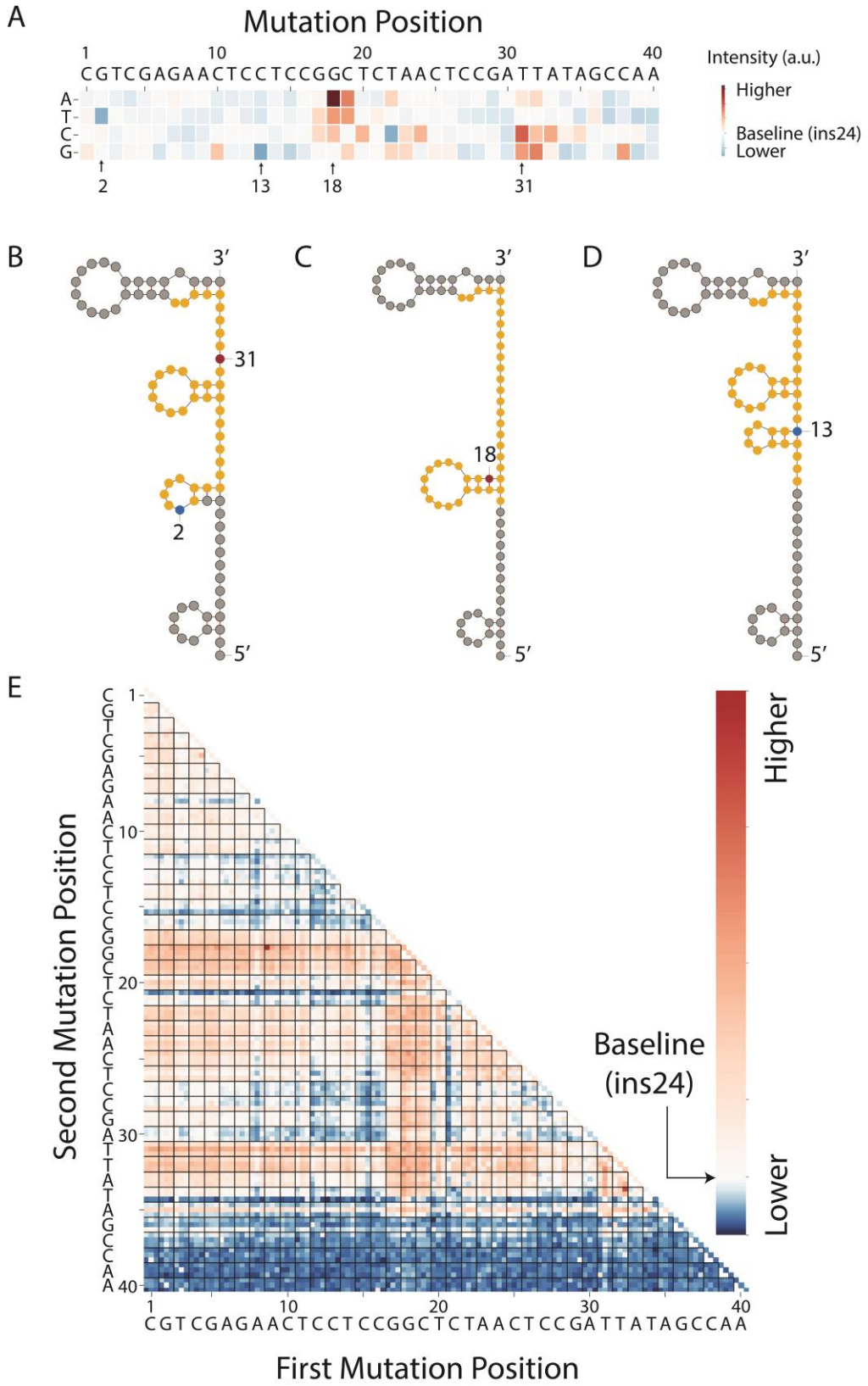
A
Mutation Position
1
10
20
30
40
CGTCGAGAACTCCTCCGGCTCTAACTCCGATTATAGCCAA
A
T
C
G
2
13
18
31
Intensity (a.u.)
Higher
Baseline (ins24)
Lower
B
C
D
3′
5′
31
2
18
13
E
Second Mutation Position
First Mutation Position
Higher
Lower
Baseline
(ins24)

Figure 4 A) Heatmap of the impact of single-nucleotide mutations on ins24 binding to 1.6 μM insulin in 1% insulin. Intensities are normalized relative to ins24. Red indicates increased intensity; blue indicates decreased intensity. Arrows indicate mutations discussed in the main text. **B**) Predicted secondary structure of ins24. Forward primer and reverse primer complement are colored in gray; N40 variable region is colored in yellow. Positions 2 (red) and 31 (blue) are highlighted as examples of mutations that respectively improve or weaken insulin binding without changing the structure. **C**) Predicted secondary structure of ins24 with a G-to-T mutation at position 18 (red) that enhances insulin binding. **D**) Predicted secondary structure of ins24 with a C-to-G mutation at position 13 (blue) that decreases insulin binding. **E**) Heatmap of the impact of all double mutants on binding of ins24 to 1.6 μM insulin in 1% serum, with data represented as in A. Each 3 x 3 square shows the nine possible mutants for a given position. Nucleotide substitutions are ordered A-T-G-C, not including the nucleotide in the original ins24 sequence.

N2A2 also offers an effective strategy for rapidly identifying mutations that contribute to an increase or decrease in binding even if no secondary structure change is predicted. For example, a G-to-T mutation at position 2 results in a decrease in binding intensity but no change in secondary structure (**Figure 4B**). This results in an additional phenylalanine modification, and we hypothesize that this may interfere in some way with the binding mechanism of the ins24 aptamer. This demonstrates that the arbitrary addition of more chemically-modified bases is insufficient to enhance target binding, and that a poorly-positioned modification can instead impede aptamer function.

We were likewise able to evaluate the binding characteristics every double-mutant of ins24 in the same N2A2 run (**Figure 4E)**. Our data show a general increase in fluorescence intensity when positions 17–20 and 31–33 are mutated to other bases, which seems to be largely independent of the second mutation location. Mutations at position 15 from C to G, at position 21 from C to A, and at position 34 from T to G broadly resulted in a decrease in fluorescence intensity that was likewise largely independent of the second mutation. These results indicate that these bases may potentially represent key binding positions. We also observed that any double mutants that include positions 37–40 consistently exhibited a strong reduction in fluorescence intensity (**Figure 4E**). As shown in **Figure 4B**, these bases form part of a stem-loop with the reverse primer complement sequence. This stem loop appears to be key to target binding, based on the disruptive effects of these mutations and the fact that no second mutation can rescue insulin binding.

These results thus reinforce and expand on the interaction insights gleaned from single-mutant analysis. Such mutational landscapes could be deeply informative for assay design or the introduction of further modifications to base-modified aptamers, and in many cases could even yield aptamers with further optimized binding properties. In the case of ins24, the improvements

we observed were relatively modest but clearly reproducible in other assays. For example, flow cytometry analysis of the double mutant Ins24_18A_31C (**Table S5**), in which G18 is replaced with A and T31 is replaced with C (**Figure S10**), showed a ~60% improvement in affinity in both buffer and serum. Thus, we believe N2A2 should generally offer an effective means for the further optimization of base-modified aptamer sequences to achieve superior binding characteristics.

Conclusion

We have demonstrated a fully automated platform that allows for the rapid and efficient screening and characterization of millions of base-modified aptamers that achieve superior affinity and specificity relative to their natural DNA or RNA counterparts. Our N2A2 platform is designed around a lightly-modified Illumina MiSeq instrument, and exploits that technology's sophisticated flow-cell and imaging apparatus to analyze vast arrays of base-modified aptamer candidates *in situ*, in a manner that directly couples binding data to defined sequences based on their position within the flow-cell. Since our approach relies on a click chemistry-based modification strategy, one can employ commercially available polymerase enzymes to incorporate virtually any base modification without meaningfully changing the experimental workflow.

As an initial demonstration, we showed that N2A2 can be used to compare the affinity of DNA aptamer libraries incorporating different base modifications in order to identify which confers superior affinity for the target protein VEGF. Our results show that tryptophan modification greatly enhances aptamer affinity for this target, enabling us to isolate a novel base-modified aptamer with four-fold greater affinity than a previously published natural DNA aptamer.

N2A2 can also be used to simultaneously screen base-modified aptamers on the basis of specificity as well as affinity. First, we demonstrated the capability to generate aptamers that can sensitively discriminate target from non-target molecules based entirely on the presence or absence of specific post-translational modifications. Our tryptophan-modified fet4 aptamer exhibited low-micromolar affinity for the glycoprotein fetuin, with negligible binding to asialofetuin—a variant of this glycoprotein that lacks sialic acid modifications. We also screened a library of phenylalanine-modified aptamers against insulin in both buffer and diluted serum, obtaining an aptamer that exhibits low micromolar affinity even in this complex medium, whereas a previously

published aptamer with comparable affinity in buffer essentially loses its capacity to bind insulin in serum.

Finally, we show that N2A2 can be a powerful tool for characterizing the critical nucleotides that contribute to target binding in base-modified aptamer sequences. After performing a screen of every possible single- or double-mutant variant of one of our aptamers from the insulin experiment, we were able to home in on nucleotides that play a prominent role in determining affinity. Several mutants even exhibited modest improvements in affinity in serum, as confirmed by flow cytometry experiments, indicating that N2A2 could also facilitate further optimization of base-modified aptamers isolated after a first round of screening.

Although many studies have demonstrated the potential of base modification to extend the utility and performance of aptamers as affinity reagents, the adoption of base-modified aptamers has been limited to date by the technical challenges associated with their generation and characterization. By essentially automating the screening and characterization process, N2A2 promises to greatly simplify the generation of a broad range of high-performance base-modified aptamers, thereby expanding access to these powerful and promising reagents. Our approach requires only simple and straightforward modifications to a widely available sequencing instrument, and as such, we believe that N2A2 will accelerate the development of a highly optimized toolbox of affinity reagents for both clinical and basic research applications.

References

(1) Tuerk, C., and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science 249*, 505–10.
(2) Ellington, A., and Szostak, J. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature 346*.
(3) Gold, L., Ayers, D., Bertino, J., Bock, C., Bock, A., Brody, E. N., Carter, J., Dalby, A. B., Eaton, B. E., Fitzwater, T., Flather, D., Forbes, A., Foreman, T., Fowler, C., Gawande, B., Goss, M., Gunn, M., Gupta, S., Halladay, D., Heil, J., Heilig, J., Hicke, B., Husar, G., Janjic, N., Jarvis, T., Jennings, S., Katilius, E., Keeney, T. R., Kim, N., Koch, T. H., Kraemer, S., Kroiss, L., Le, N., Levine, D., Lindsey, W., Lollo, B., Mayfield, W., Mehan, M., Mehler, R., Nelson, S. K., Nelson, M., Nieuwlandt, D., Nikrad, M., Ochsner, U., Ostroff, R. M., Otis, M., Parker, T., Pietrasiewicz, S., Resnicow, D. I., Rohloff, J., Sanders, G., Sattin, S., Schneider, D., Singer, B., Stanton, M., Sterkel, A., Stewart, A., Stratford, S., Vaught, J. D., Vrkljan, M., Walker, J. J., Watrobka, M., Waugh, S., Weiss, A., Wilcox, S. K., Wolfson, A., Wolk, S. K., Zhang, C., and Zichi, D. (2010) Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One 5*, e15004.
(4) Pfeiffer, F., Rosenthal, M., and Siegl, J. (2017) Customised nucleic acid libraries for enhanced aptamer selection and performance. *Curr. Opin. Biotechnol. 48*, 111–118.
(5) Anosova, I., Kowal, E. A., Dunn, M. R., Chaput, J. C., Horn, W. D. V., and Egli, M. (2016) The structural diversity of artificial genetic polymers. *Nucleic Acids Res. 44*, 1007–1021.
(6) Pinheiro, V. B., and Holliger, P. (2012) The XNA world: Progress towards replication and evolution of synthetic genetic polymers. *Curr. Opin. Chem. Biol. 16*, 245–252.
(7) Yang, Z., Hutter, D., Sheng, P., Sismour, A. M., and Benner, S. A. (2006) Artificially expanded genetic information system: A new base pair with an alternative hydrogen bonding pattern. *Nucleic Acids Res. 34*, 6095–6101.
(8) Tolle, F., Rosenthal, M., Pfeiffer, F., and Mayer, G. (2016) Click Reaction on Solid Phase Enables High Fidelity Synthesis of Nucleobase-Modified DNA. *Bioconjug. Chem. 27*, 500–503.
(9) Agard, N. J., Prescher, J. A., and Bertozzi, C. R. (2004) A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc. 126*, 15046–15047.
(10) Tolle, F., Brändle, G. M., Matzner, D., and Mayer, G. (2015) A Versatile Approach Towards Nucleobase-Modified Aptamers. *Angew. Chemie - Int. Ed. 54*, 10971–10974.
(11) Latham, J. A., Johnson, R., and Toole, J. J. (1994) The application of a modified nucleotide in aptamer selection: Novel thrombin aptamers containing -(1 -pentynyl)-2′-deoxyuridine. *Nucleic Acids Res. 22*, 2817–2822.
(12) Gawande, B. N., Rohloff, J. C., Carter, J. D., Von Carlowitz, I., Zhang, C., Schneider, D. J., and Janjic, N. (2017) Selection of DNA aptamers with two modified bases. *Proc. Natl. Acad. Sci. U. S. A. 114*, 2898–2903.
(13) Gordon, C. K. L., Wu, D., Pusuluri, A., Feagin, T. A., Andrew, T., Eisenstein, M. S., Hawker, C. J., Niu, J., and Soh, H. T. (2019) Click-Particle Display for Base-Modified Aptamer Discovery. *ACS Chem. Biol. 14*, 2652–2662.
(14) Nutiu, R., Friedman, R. C., Luo, S., Khrebtukova, I., Silva, D., Li, R., Zhang, L., Schroth, G. P., and Burge, C. B. (2011) Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nat. Biotechnol. 29*, 659–664.

(15) Tome, J. M., Ozer, A., Pagano, J. M., Gheba, D., Schroth, G. P., and Lis, J. T. (2014) Comprehensive analysis of RNA-protein interactions by high-throughput sequencing-RNA affinity profiling. *Nat. Methods 11*, 683–8.
(16) Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., Chang, H. Y., and Greenleaf, W. J. (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature 523*, 486–490.
(17) David R. Bentley, Shankar Balasubramanian, Harold P. Swerdlow, Geoffrey P. Smith, John Milton, Clive G. Brown, Kevin P. Hall, Dirk J. Evers, Colin L. Barnes, Helen R. Bignell, Jonathan M. Boutell, Jason Bryant, Richard J. Carter, R. Keira Cheetham, Anthon, R. D. & A. J. S. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature 456*, 53–59.
(18) Metzker, M. L. (2010) Sequencing technologies the next generation. *Nat. Rev. Genet. 11*, 31–46.
(19) Kornreich-Leshem, H., Jäger, S., Thum, O., Rasched, G., Famulok, M., and Engeser, M. (2005) A Versatile Toolbox for Variable DNA Functionalization at High Density. *J. Am. Chem. Soc. 127*, 15071–15082.
(20) Hili, R., Niu, J., and Liu, D. R. (2013) DNA Ligase-Mediated Translation of DNA Into Densely Functionalized Nucleic Acid Polymers. *J. Am. Chem. Soc. 135*, 98–101.
(21) Birtalan, S., Zhang, Y., Fellouse, F. A., Shao, L., Schaefer, G., and Sidhu, S. S. (2008) The Intrinsic Contributions of Tyrosine, Serine, Glycine and Arginine to the Affinity and Specificity of Antibodies. *J. Mol. Biol. 377*, 1518–1528.
(22) Wang, J., Gong, Q., Maheshwari, N., Eisenstein, M., Arcila, M. L., Kosik, K. S., and Soh, H. T. (2014) Particle Display : A Quantitative Screening Method for Generating High Affinity Aptamers. *Angew. Chemie 126*, 1–6.
(23) Kaur, H., and Yung, L. Y. L. (2012) Probing high affinity sequences of DNA aptamer against VEGF 165. *PLoS One 7*, 19–26.
(24) Jellinek, D., Green, L. S., Bell, C., and Janjić, N. (1994) Inhibition of receptor binding by high-affinity RNA ligands to vascular endothelial growth factor. *Biochemistry 33*, 10450–10456.
(25) Alam, K. K., Chang, J. L., and Burke, D. H. (2015) FASTAptamer: A Bioinformatic Toolkit for High-throughput Sequence Analysis of Combinatorial Selections. *Mol. Ther. Acids 4*, e230.
(26) Wang, J., Yu, J., Yang, Q., McDermott, J., Scott, A., Vukovich, M., Lagrois, R., Gong, Q., Greenleaf, W., Eisenstein, M., Ferguson, B. S., and Soh, H. T. (2017) Multiparameter Particle Display (MPPD): A Quantitative Screening Method for the Discovery of Highly Specific Aptamers. *Angew. Chemie - Int. Ed. 56*, 744–747.
(27) Shriver, Z., Raguram, S., and Sasisekharan, R. (2004) Glycomics: a pathway to a class of new and improved therapeutics. *Nat. Rev. Drug Discov. 3*, 863–73.
(28) Baenziger, J. U., and Fiete, D. (1979) Structure of the complex oligosaccharides of fetuin. *J. Biol. Chem. 254*, 789–795.
(29) Sujatha, M. S., Sasidhar, Y. U., and Balaji, P. V. (2004) Energetics of galactose- and glucose-aromatic amino acid interactions: Implications for binding in galactose-specific proteins. *Protein Sci. 13*, 2502–2514.
(30) Yoshida, W., Mochizuki, E., Takase, M., Hasegawa, H., Morita, Y., Yamazaki, H., Sode, K., and Ikebukuro, K. (2009) Selection of DNA aptamers against insulin and construction of an aptameric enzyme subunit for insulin sensing. *Biosens. Bioelectron. 24*, 1116–1120.
(31) Chothia, C., Lesk, A. M., Dodson, G. G., and Hodgkin, D. C. (1983) Transmission of conformational change in insulin. *Nature 302*, 500–505.

(32) Timofeev, V. I., Chuprov-Netochin, R. N., Samigina, V. R., Bezuglov, V. V., Miroshnikov, K. A., and Kuranova, I. P. (2010) X-ray investigation of gene-engineered human insulin crystallized from a solution containing polysialic acid. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 66*, 259–263.
(33) Xu, W., and Ellington, A. D. (1996) Anti-peptide aptamers recognize amino acid sequence and bind a protein epitope. *Proc. Natl. Acad. Sci. U. S. A. 93*, 7475–7480.
(34) McKeague, M., and Derosa, M. C. (2012) Challenges and opportunities for small molecule aptamer development. *J. Nucleic Acids 2012*, 748913.
(35) Katilius, E., Flores, C., and Woodbury, N. W. (2007) Exploring the sequence space of a DNA aptamer using microarrays. *Nucleic Acids Res. 35*, 7626–7635.
(36) Choi, S., and Ban, C. (2016) Crystal structure of a DNA aptamer bound to Pv LDH elucidates novel single-stranded DNA structural elements for folding and recognition. *Sci. Rep. 6*, 34998.
(37) Peattie, D. a, and Gilbert, W. (1980) Chemical probes for higher-order structure in RNA. *Proc. Natl. Acad. Sci. U. S. A. 77*, 4679–82.
(38) Kladwang, W., VanLang, C. C., Cordero, P., and Das, R. (2011) A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat. Chem. 3*, 954–962.
(39) Kladwang, W., and Das, R. (2010) A mutate-and-map strategy for inferring base pairs in structured nucleic acids: Proof of concept on a DNA/RNA helix. *Biochemistry 49*, 7414–7416.
(40) Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res. 31*, 3406–3415.