Versatile High-Throughput Screening for Transcriptional Repressor Domains Using a Single 300mer Oligo Pool

Explore a series of high-throughput screening applications, made possible by 300mer Twist Oligos

ABSTRACT

Improvements in the quality and scale of oligonucleotide (oligo) synthesis have enabled a wide variety of high-throughput applications, ranging from deep mutational scans to protein tiling screens. Here, a novel pooled recruitment assay—known as high-throughput recruit (HT-recruit)—was used for high-throughput identification of transcriptional repressor and activator domains in both annotated and unannotated human nuclear protein sequences. A single Twist Oligo Pool was used to generate three individual libraries: an annotated nuclear protein domain library, a tiling library of unannotated nuclear protein domains, and a deep mutational scanning library of the ZNF10 (KOX1) KRAB domain. The ability to generate 300mer oligos was a critical tool for researchers, enabling them to screen more than 70% of sequences identified in PFAM as human nuclear protein domains, which are potential effectors. Coupled with the exceptional precision, length, and uniformity of Twist Oligos, HT-recruit opens the door to protein domain characterization on a genome-wide scale.

INTRODUCTION

Of the roughly 20,000 proteins coded in the human genome, approximately 4,000 have no known function. Among those with known functions, very few are characterized to such a degree that researchers can link protein functions to specific protein domains. Uncovering the functional anatomy of proteins can significantly help elucidate mechanisms of disease and inspire better molecular technologies, such as CRISPRi/a.

Of particular interest are protein domains that influence gene expression, so-called effector domains, or simply effectors. Effector function has historically been measured using recruitment assays in which candidate effector domains are fused to a canonical DNA-binding domain that is targeted to the promoter upstream of a reporter gene. If the candidate functions as a transcriptional effector, it should alter reporter gene expression when recruited.

While useful for studying individual proteins, traditional recruitment assays are limited in flexibility and throughput, in large part because of their reliance on cDNA isolation. When the source material must be harvested from a cell, it's not practical to do point-by-point mutational analyses or to tile walk along an amino acid sequence as would be needed to study effector sequences with amino acid resolution. Additionally, isolating cDNA means researchers are limited to studying species-specific proteins.

To gain high-throughput characterization of protein effector domains, researchers need to be able to generate candidate effector sequences without the pains and limitations of isolating cDNA. In short, they need synthetic DNA.

Large-scale DNA synthesis technology can produce screening libraries containing thousands of unique effector sequences. However, candidate effector domains can come in many lengths, from as small as 10 amino acids to longer than 300 amino acids. Historically, robust oligonucleotide synthesis was limited to approximately 150 nucleotides in length, or 50 amino acids. This limitation prevented the study of many potential repressor domains which tend to be more structured and longer in length.

Recent advances in synthesis technology have made it possible to now produce oligos up to 300 nucleotides in length (300mer), enabling researchers to study significantly more candidate effector domains. With longer oligos, researchers can include more content in each oligonucleotide which may be critical when expressing highly structured protein domains. Twist Bioscience's DNA synthesis platform has made it possible to generate highly uniform oligo pools up to 300 nucleotides in length.

Using Twist's platform, researchers at Stanford University recently developed a HT-recruit novel high-throughput recruitment assay capable of screening thousands of nuclear protein domains for transcriptional repressor activity (Tycko et al, 2020). The method combines massively parallel 300mer oligo synthesis, magnetic cell selection, and next-generation sequencing to iden-tify repressor function in annotated, unannotated, and mutant nuclear protein domains of human cells.

This application note showcases three high-throughput screening applications that use the HT-recruit method. In the first screen, thousands of PFAM-annotated nuclear protein domains were screened for repressor function. In the second, HT-recruit was applied to discover repressor domains in unannotated nuclear protein domains using a protein tiling approach. Finally, a deep mutational scan was performed to identify amino acid variants that alter the repressor activity of a well-known KRAB domain from the CRISPRi system. The collective results highlight the versatility of 300mer Twist Oligo libraries in resolving repressor functionality in human nuclear protein sequences in this and other applications.

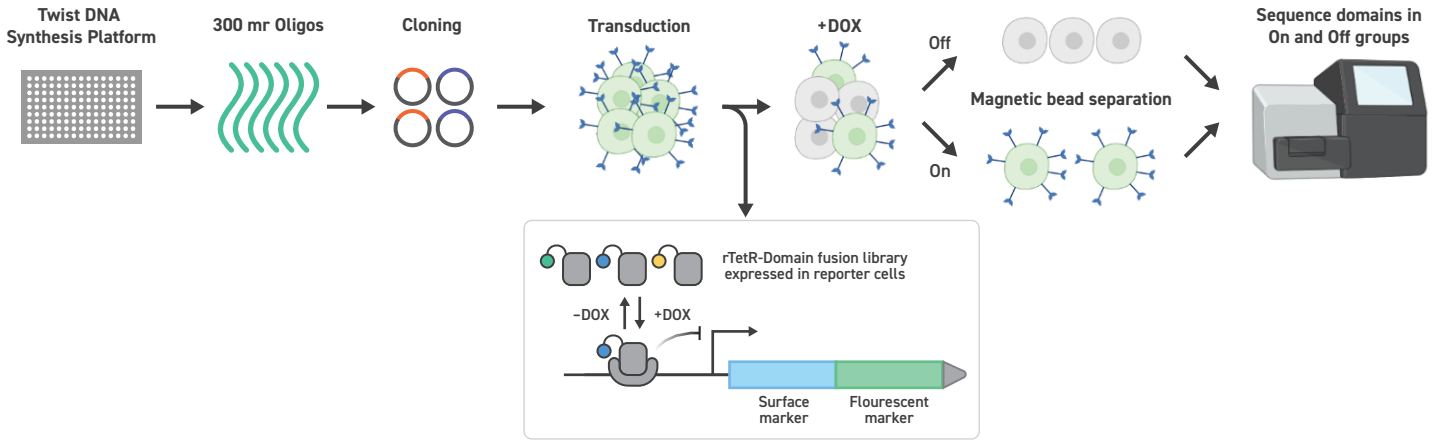

Figure 1. HT-recruit workflow. Schematic diagram of the HT-recruit workflow used to identify repressor domains in annotated, unannotated, and mutated human nuclear protein sequences. 300mer oligonucleotides coding for different potential effector domains are synthesized using Twist Bioscience's silicon-based platform. These are then cloned into rTetR-vectors, forming rTetR-fusions. Vectors are transduced into cells containing a reporter surface marker and citrine report genes. Doxycycline (dox) treatment induces rTetR binding to the TetO motif upstream of the reporter, enabling effector domains to repress reporter expression. After 5 days of dox treatment, cells are sorted based on reporter expression and sequenced to determine which domains had repressive activity (by a lack of reporter expression). Figure created with Biorender.com

HT-RECRUIT WORKFLOW

HT-recruit uses K562 cells that express a constitutively active synthetic reporter construct. The synthetic construct is the human IgG1 Fc region linked to an Igκ leader and PDGFRB trans-membrane domain, enabling the selection of cells with magnetic beads. From library synthesis to sequencing, the HT-recruit workflow involves five steps (Figure 2):

1. **Library generation:** 300mer oligos are synthesized as a single, massive oligonucleotide pool. Three sets of sublibrary-specific adapters were encoded in the oligonucleotides to enable libraries to be amplified individually from the master oligo pool and then cloned into lentiviral vectors as fusions with a reverse Tet Repressor (rTetR), a doxycycline-inducible DNA binding protein.
2. **Spinfection:** Reporter K562 cells, which are amenable to large-scale suspension culture in spinner flasks, are transduced at low multiplicity of infection (0.3) to ensure that cells only receive a single rTetR-domain fusion.
3. **Recruitment:** Successfully transduced cells are exposed to doxycycline to induce recruitment of rTetR domains to a promoter upstream of the constitutively active synthetic reporter gene. Cells expressing repressive rTetR-domain fusions inhibit the expression of the synthetic reporter gene (OFF cells), whereas those with domains lacking function continue to express it (ON cells).
4. **Selection:** OFF and ON cells are separated using magnetic beads that bind the surface-expressed, synthetic reporter gene.
5. **Sequencing:** Genomic DNA from OFF and ON cells is extracted and sequenced to identify rTetR-domain fusions enriched in OFF cells (i.e., those with repressor function).

This workflow was applied to all three screens described below with minor customizations as noted below.

SCREEN 1: DISCOVERY OF REPRESSOR FUNCTION IN PFAM-ANNOTATED DOMAINS

Goal: To comprehensively screen for repressor function in annotated nuclear protein domains from the *PFAM* protein family database (El-Gebali et al, 2019).

Library Design & Quality: PFAM-annotated domains were retrieved from all human nuclear proteins of the UniProt database. Being shorter or equal to 80 AA long, an overwhelming majority of these domains (72% at the time of publication) were compatible with pooled synthesis as 300mers (Figure 1); these domains were chosen for screening. Shorter domains were lengthened to 80 AA to mitigate PCR biases during library preparation and cloning. Negative controls included 499 randomly generated 240-bp (80-AA) sequences lacking stop codons and 361 sequences of the same length tiling the DMD protein, a protein without transcriptional regulatory function. The total library size was 5,955 oligos.

Results: Only 3,014 domains (66%) expressed well in K562 cells. Of these, 446 functioned as repressors, as evidenced by a clear reduction in reporter expression. Repressor hits included members of a wide variety of domain families, including well-characterized ones like KRAB and Homeodomain as well as previously uncharacterized domains (i.e., domains of unknown function [DUF] like DUF3669 and DUF1087). Replicate screens demonstrated high reproducibility ($r^2 = 0.96$).

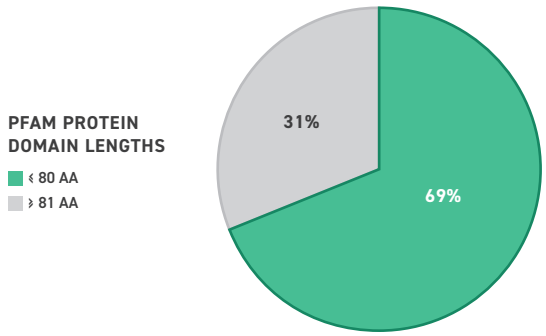

Figure 2. PFAM domain lengths for human nuclear proteins. Most human nuclear protein domains are ≤80 AAs and therefore compatible with pooled oligo synthesis as 300mers. Data was collected from the UniProt database in May of 2022. A total of 15,277 domains were analyzed. This number slightly differs from those reported by Tycko et al., owing to the addition of new entries to PFAM after the publication of their work.

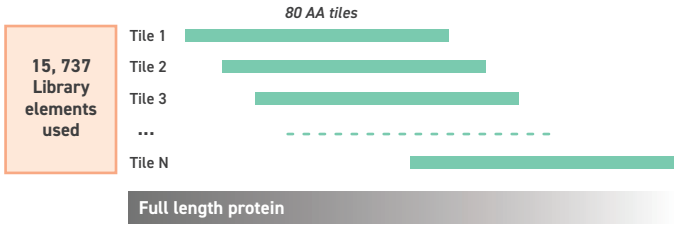

Figure 3. Illustration of tiling approach from Screen 2. 80 amino acid (AA) tiles were staggered in an overlapping pattern every 10 amino acids for the full length of 238 nuclear protein sequences. A total of 15,737 tiles were ordered.

SCREEN 2: ASCRIBING FUNCTION TO UNANNOTATED DOMAINS USING PROTEIN TILING

Goal: To discover repressor function in unannotated nuclear protein domains.

Library Design & Quality: A list of 238 unannotated nuclear proteins was curated manually from a database of transcriptional regulators (Lambert et al, 2018). To determine any functional repressor domains the genomic sequences were tiled with 240-bp (80-AA) oligos separated by a 30-bp (10-AA) sliding window (Figure 3). After adding 361 negative control tiles (spanning DMD proteins), the resulting library size was 15,737 oligos.

Results: The protein tiling screen identified repressor hits in 141 of 238 proteins. Of the tiles used in this study, 4.3% showed a reproducible hit as a repressor, with replicate experiments showing concordant results. Among the hits were unannotated domains in BAZ2A, a component of the nuclear remodeling complex (NoRC); the ten eleven translocation (TET) DNA demethylases TET1, TET2, and TET3; and the transcription factor MGA. Although the repressor function of MGA was already known (Blackledge et al, 2014), the protein tiling screen identified the two domains—AAs 341–420 and 2,381–2,460—that specifically mediate this function. By investigating the overlap between hit tiles for MGA, a minimal, 10 AA-long repressor domain was discovered.

SCREEN 3: DEEP MUTATIONAL ANALYSIS OF ZNF10 KRAB DOMAIN

Goal: To clarify sequence-function relationships in the ZNF10 KRAB domain, the first KRAB domain discovered (Bellefroid et al, 1991) and the most popular repressor domain used for CRISPRi (Gilbert et al, 2013).

Library Design & Quality: The ZNF10 KRAB domain sequence was mutated with all possible single and consecutive double and triple mutations to generate the mutational library. In addition, all PFAM-annotated human KRAB sequences and 80-AA tiling sequences for five KRAB Zinc Finger genes were included. Negative controls included 300 randomly generated sequences and 200 sequences tiling the DMD protein. The total library size was 5,731 oligos.

Results: The deep mutational scan found 13 residues important for the ZNF10 KRAB domain's repressor function. These mutations were linked to the domain's A-box and B-box regions, which participate in the interaction between KRAB and the co-repressor KAP1. The reproducibility of the screen was high (r^2 = 0.93), and stabilizing mutations were found which could enhance CRISPRi activity.

DISCUSSION

This application note showcases a series of screens made possible by the HT-recruit assay and Twist Oligos. However, this is only the tip of the iceberg in terms of what can be achieved with HT-recruit. Indeed, Tycko et al (2020) also adapted HT-recruit to screen for transcriptional activator domains in PFAM-annotated nuclear protein domains. Similar to the screens described here, the activator screen identified numerous functional domains with high reproducibility.

Continued innovations in oligo synthesis technologies have enabled the synthesis of high-quality libraries of longer and longer oligos (LeProust et al, 2010). Oligonucleotide pool quality features such as uniformity and sequence fidelity contribute directly to screening sensitivity and success by maximizing the number of scorable library elements. These features also ensure that a screen is reproducible, as highlighted by the high correlation between replicates across the screens described here (r^2 values ≥ 0.72). The more reproducible a screen is, the more likely its results reflect true biological phenomena.

Whereas oligo quality impacts the sensitivity of a screen, oligo length ultimately determines how much content can be screened per oligo. This is particularly relevant for protein tiling applications like the one described here (Screen 2). Although a minimal repressor domain of 10 AAs was identified in MGA, only a minority of effector nuclear protein domains exceed 80 AAs in length (Figure 1). The upper and lower resolution bounds of a tiling library are determined by oligo length and the offset between adjacent tiles, respectively.

The cost of high-throughput functional screening has been dramatically reduced by advances in oligo synthesis. These costs can be further reduced by combining multiple sublibraries into a master library for synthesis. The 300mer length of Twist Oligos accommodates features like primer sequences that allow sublibraries to be deconvoluted from a single master oligo pool. The Oligonucleotide pools used in this application note were synthesized together in a single oligo pool that also included sub-libraries outside the scope of this work, totaling tens of thousands of oligos.

Overall, this application note demonstrates the versatility and performance of Twist Oligo Pools in high-throughput screening applications.

REFERENCES

Bellefroid EJ, Poncelet DA, Lecocq PJ, Revelant O, & Martial JA. (1991). The evolutionarily conserved Krüppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. Proceedings of the National Academy of Sciences of the United States of America, 88(9), 3608–3612.

Blackledge NP, Farcas AM, Kondo T, King HW, McGouran JF, Hanssen LLP, Ito S, Cooper S, Kondo K, Koseki Y, Ishikura T, Long HK, Sheahan TW, Brockdorff N, Kessler BM, Koseki H, & Klose RJ. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell 157, 1445–1459.

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, & Finn RD. (2019). The PFAM protein families database in 2019. Nucleic acids research, 47(D1), D427–D432.

Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, & Weirauch MT. (2018). The Human Transcription Factors. Cell, 172(4), 650–665.

LeProust EM, Peck BJ, Spirin K, McCuen HB, Moore B, Namsaraev E, & Caruthers MH. (2010). Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. Nucleic acids research, 38(8), 2522–2540.

Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, Kawli T, Davis CA, Dobin A, Kaul R, Halow J, Van Nostrand EL, Freese P, Gorkin DU, Shen Y, He Y, Mackiewicz M, Pauli-Behn F, Williams BA, Mortazavi A, Keller CA, Zhang XO, Elhajjajy SI, Huey J, Dickel DE, Snetkova V, Wei X, Wang X, Rivera-Mulia JC, Rozowsky J, Zhang J, Chhetri SB, Zhang J, Victorsen A, White KP, Visel A, Yeo GW, Burge CB, Lécuyer E, Gilbert DM, Dekker J, Rinn J, Mendenhall EM, Ecker JR, Kellis M, Klein RJ, Noble WS, Kundaje A, Guigó R, Farnham PJ, Cherry JM, Myers RM, Ren B, Graveley BR, Gerstein MB, Pennacchio LA, Snyder MP, Bernstein BE, Wold B, Hardison RC, Gingeras TR, Stamatoyannopoulos JA, Weng Z, ENCODE Project Consortium. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature, 583(7818), 699–710.

Tycko J, DelRosso N, Hess GT, Aradhana, Banerjee A, Mukund A, Van MV, Ego BK, Yao D, Spees K, Suzuki P, Marinov GK, Kundaje A, Bassik MC, & Bintu L. (2020). High-Throughput Discovery and Characterization of Human Transcriptional Effectors. Cell, 183(7), 2020–2035.e16.