A smart polymer for sequence-selective binding, pulldown and release of DNA targets

Elisha Krieg,[1-5,*] Krishna Gupta,[4,6] Andreas Dahl,[7] Mathias Lesche,[7] Susanne Boye,[4] Albena Lederer,[4,5] William M. Shih[1-3,*]

[1]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, USA; [2]Wyss Institute for Biologically Inspired Engineering at Harvard University, Boston, USA; [3]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, USA; [4]Leibniz-Institut für Polymerforschung Dresden e.V., Germany; [5]School of Science, Technische Universität Dresden, Germany; [6]Biotechnology Center (BIOTEC), Technische Universität Dresden, Germany; [7]DRESDEN-concept Genome Center, Center for Molecular and Cellular Bioengineering, Technische Universität Dresden, Germany.

* Correspondence and requests for materials should be addressed to E.K. (krieg@ipfdd.de) or to W.M.S. (william.shih@wyss.harvard.edu).

Abstract

Selective isolation of DNA is crucial for applications in biology, bionanotechnology, clinical diagnostics and forensics. We herein report a smart methanol-responsive polymer (MeRPy) that can be programmed to bind and separate single- as well as double-stranded DNA targets. Captured targets are quickly isolated and released back into solution by denaturation (sequence-agnostic) or toehold-mediated strand displacement (sequence-selective). The latter mode allows 99.8% efficient removal of unwanted sequences and 79% recovery of highly pure target sequences. We applied MeRPy for depletion of *insulin* cDNA from clinical next-generation sequencing (NGS) libraries. This step improved data quality for low-abundance transcripts in expression profiles of pancreatic tissues. Its low cost, scalability, high stability and ease of use make MeRPy suitable for diverse applications in research and clinical laboratories, including enhancement of NGS libraries, extraction of DNA from biological samples, preparative-scale DNA isolations, and sorting of DNA-labeled non-nucleic acid targets.

Introduction

Materials that enable selective separation of DNA sequences are crucial for many life science applications. Isolation of high-purity DNA is required across a wide range of scales, for analytical and preparative purposes alike.[1–5] Next-generation sequencing (NGS), for instance, necessitates extraction of DNA from biological samples, enrichment of a subset of target sequences or depletion of interfering library components.[1,6–8] Despite rapid advancements, high reagent costs and time-consuming sample preparation remain a significant obstacle for the full implementation of NGS in clinical practice.[9]

Several approaches are available to isolate and purify nucleic acids (NA).[1,8,10] Target sequences can be pulled down from solution via biotinylated probes that are captured by streptavidin-coated solid beads (e.g. *Thermo Scientific* Dynabeads). Enzymatic

approaches can be used to selectively amplify desired NA species or digest undesired ones (e.g. *New England Biolabs* NEBNext rRNA Depletion Kit, or *Evrogen* Duplex-Specific Nuclease[11]). Multiple research groups have recently reported innovative solutions that complement and improve on existing technologies.[2–4,7]

Despite the diverse variety of approaches, current methods have critical shortcomings that require trade-offs between material cost, ease of use, versatility and performance. Commercial kits are expensive,[9,10] due to high costs of recombinant proteins and enzymes. Solid beads are susceptible to non-specific interfacial adsorption of macromolecules, leakage of surface-attached streptavidin, degradation in presence of reducing agents and chelators, or incomplete release of target molecules. Most assays have limited stability, shelf life, and a narrow range of compatible experimental conditions. It is therefore crucial to engineer less expensive and more robust materials that allow fast and efficient binding, manipulation and release of selectable target sequences.

To address this challenge, we have developed an oligonucleotide-grafted methanol-responsive polymer (*MeRPy*). MeRPy's development was inspired by SNAPCAR, a recently reported method for scalable production of kilobase-long single-stranded DNA (ssDNA).[3] MeRPy acts as a ready-to-use macroligand[12] for affinity precipitation. We show that the polymer can bind one or multiple DNA targets, isolate, and release selected targets back into the medium. MeRPy pulldown is directly applicable to unlabeled DNA, requires only most basic lab techniques and can be completed within a few minutes experimental time.

Results and Discussion

MeRPy is a poly(acrylamide-*co*-acrylic acid)-*graft*-oligo(nucleic acid) copolymer (Figure 1c). It can be selectively precipitated by addition of methanol (Figure 1a,b). The polymer's carboxylate groups (1 wt%) are crucial to suppress nonspecific binding of free DNA. Grafted oligonucleotides serve as universal *anchor strands* that provide high binding capacity and activity. To define sequences for target capture and release, MeRPy is programmed with *catcher strand* probes that consist of three domains (Figure 1d): (i) an *adapter site*, (ii) a *target binding site, and* (iii) an (optional) *release site*. After targets hybridize to the binding site, the polymer is precipitated. The pellet is then redispersed in water and targets are released, either non-selectively by thermal or basic denaturation, or selectively via toehold-mediated strand displacement (TMSD)[13].

We synthesized two variants of MeRPy (Supplementary Procedure 1): (i) **MeRPy-10** carries ~10 anchor strands per polymer chain. It can bind up to 2 nmol ssDNA per milligram polymer. (ii) **MeRPy-100** was synthesized for applications that demand increased binding capacity and kinetics. It is endowed with ~100 anchor strands per chain, providing 20 nmol hybridization sites per milligram polymer. Its binding capacity was found to be 15 nmol per milligram polymer, corresponding to ~75% of its theoretical limit (Supplementary Figure S1). Both MeRPy variants have significantly higher binding capacities than widely-used magnetic beads, which are limited by molecular crowding at the solid-liquid interface (typically max. 200–500 pmol ssDNA per milligram substrate).[14]

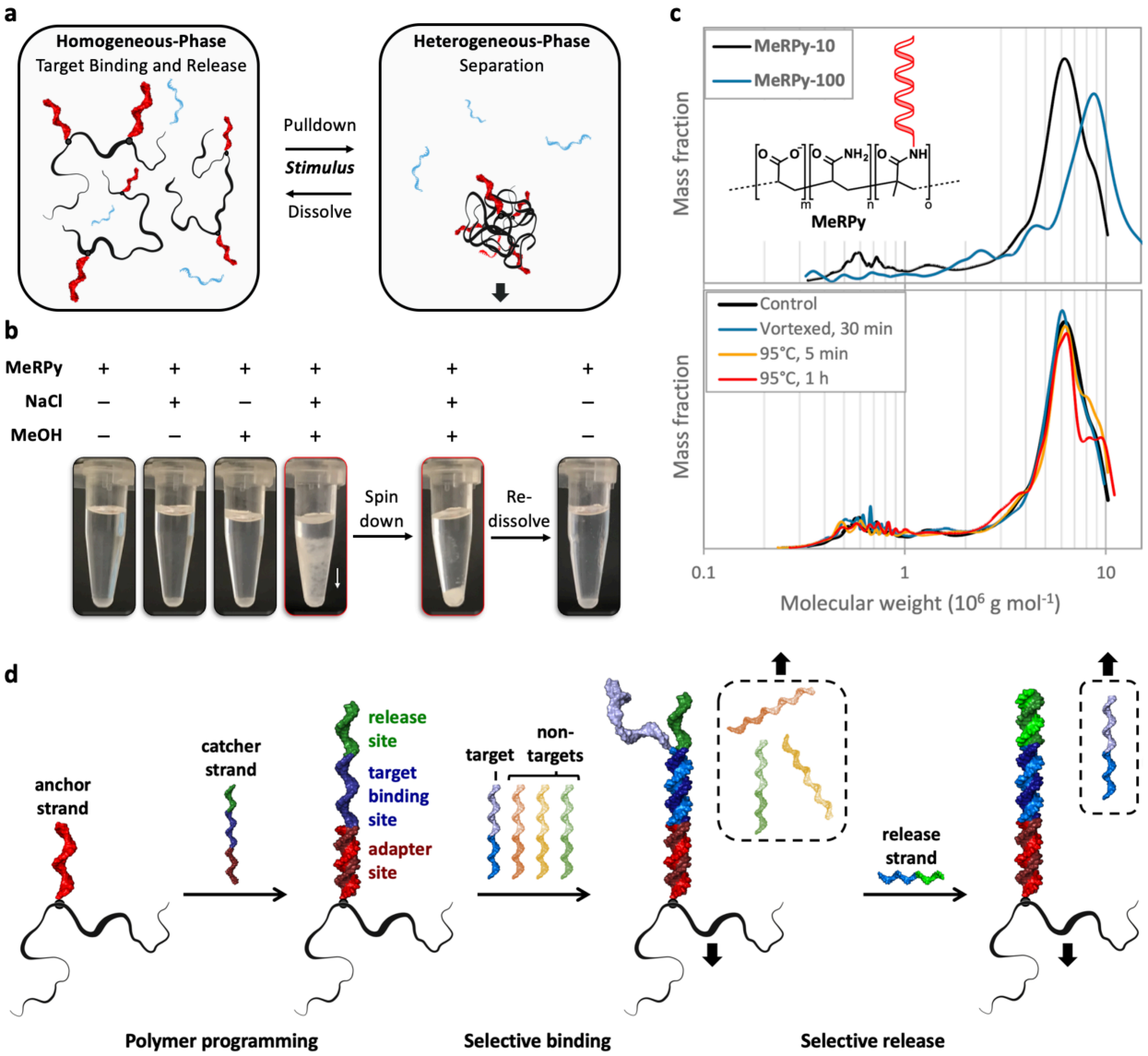

Figure 1. Basic features of MeRPy. a) Due to its methanol-responsiveness MeRPy combines rapid binding/release in homogeneous phase with facile separation in heterogeneous phase. **b)** MeRPy is soluble in aqueous buffers but precipitates in solutions containing ≥30 mM NaCl when methanol is added. The precipitate can be redissolved in water and triggered to release targets on demand. **c)** Chemical structure and molecular weight distributions. Top: Molecular weight distributions of **MeRPy-10** (m~740, n~74,000, o~10) and **MeRPy-100** (m~900, n~90,000, o~130), obtained by AF4-LS. Bottom: Molecular weight distributions of **MeRPy-10** as synthesized (control), after extensive vortexing and heating to 95°C. **d)** Scheme of MeRPy programming, target binding and release.

MeRPy-10 and **MeRPy-100** are soluble in aqueous media. Methanol-induced precipitation requires prior adjustment of the ionic strength, as negative charges of the polymer's carboxylate groups and anchor strands must be sufficiently shielded by counterions (Figure 1b). **MeRPy-10** requires 30–150 mM NaCl and 1 volume of

methanol for complete precipitation. In contrast, **MeRPy-100** requires 100–300 mM NaCl and 1.5 volumes of methanol for this process.

MeRPy's high molecular weight is crucial for its robust and quantitative responsiveness. Asymmetrical flow-field flow fractionation measurements in combination with static and dynamic light scattering (AF4-LS) indicate that **MeRPy-10** and **MeRPy-100** have weight average molecular weights (M_w) of 5.35 MDa and 7.36 MDa, respectively. (Figure 1c, Supplementary Figure S3, Supplementary Table S1). The respective dispersities (Đ) are 1.87 and 2.00. MeRPy chains are up to six times heavier than chains produced in SNAPCAR experiments (M_w~1.2 MDa).[3] The increased molar mass, which improves the efficiency of methanol-induced precipitation, is enabled by its different synthesis approach: in SNAPCAR, target molecules and other reagents can interfere with the *in-situ* radical polymerization; in contrast, MeRPy synthesis is independent of the target capture step, and therefore takes place under highly controlled conditions.

MeRPy-10 chains assume a typical random coil conformation in TE buffer at pH 8.0, as indicated by an average scaling exponent (ν) of 0.54.[15] In contrast, ν is 0.34 for **MeRPy-100**, which—in agreement with its design—indicates a high degree of branching.[16] Gyration and hydrodynamic radii support these findings (Supplementary Table S1). The apparent volume (V_h) occupied by individual **MeRPy-10** (V_h = 2.1x10^{-3} µm^3) and **MeRPy-100** molecules (V_h = 7.1x10^{-3} µm^3) in solution comprises ≥99.5% water, thus leaving the polymer highly penetrable and anchor strands well accessible for hybridization.

We tested the structural stability of MeRPy under mechanical stress and at high temperature. MeRPy chains remain fully intact when vortexed for 30 minutes or heated to 95°C for 5 minutes (Figure 1c). These exposure times exceed those used in typical pulldown experiments (see below). Heating to 95°C for 1 hour lead to minor decomposition of the upper molecular weight fraction of MeRPy, and partial depurination of anchor strand bases can be expected to occur under these conditions.[17]

To demonstrate that MeRPy is applicable for separation of DNA and DNA-labeled target molecules, we applied **MeRPy-10** to a mixture of ssDNA-labeled cyanine dyes (T_1 = Cy5; T_2 = Cy3) (Figure 2). The polymer was programmed with either of two catcher strands. Fluorescence images show that the targeted dye was selectively pulled down, leaving non-target in solution. After separating pellet from solution, release of the captured target into clean buffer was triggered by addition of a release strand in presence of 100 mM NaCl. A second MeRPy pulldown then removed the polymer from the released target.

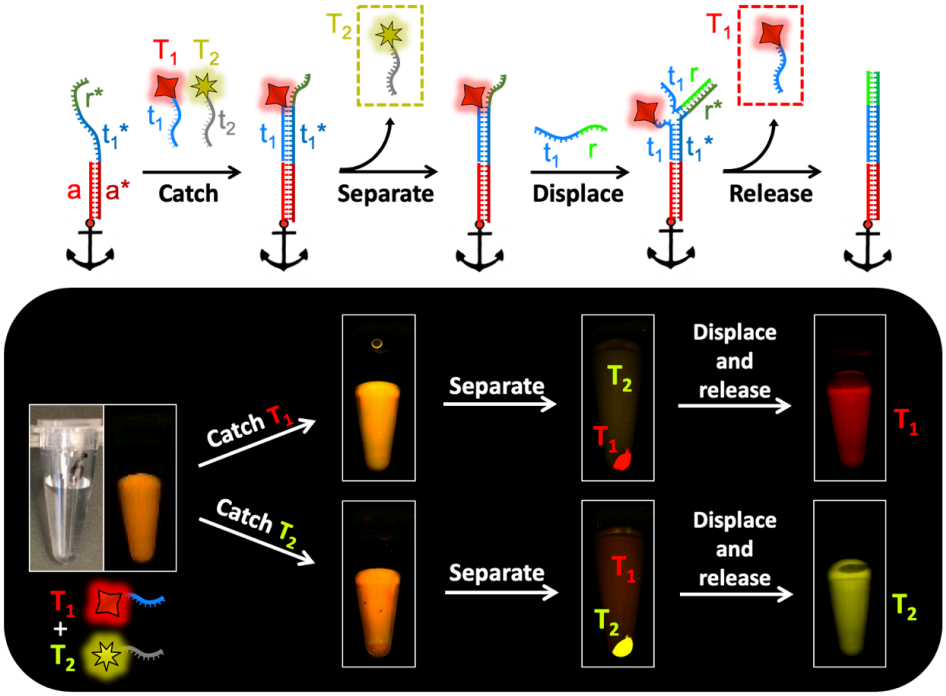

Figure 2. Sequence-specific catch-and-release of ssDNA-labeled target molecules. Photographs of tubes containing **MeRPy-10** and a mixture of two DNA-labeled fluorescent dye molecules. T_1: *Cy5*; T_2: *Cy3*. **MeRPy-10** was programmed to bind either T_1 (upper path) or T_2 (lower path). The target was then pulled down and isolated by addition of methanol and a short spin-down. After separation, the target was released back into solution by TMSD.

As MeRPy provides high anchor strand concentration, it can be used to capture many targets simultaneously at a fast rate. Figure 3 shows the manipulation of a 10-member ssDNA library (strands designated **A**–**J**) in the length range of 20–190 nt (Supplementary Table S4). Multiple targets were selected by addition of catcher strand libraries (CSL) of different compositions (Figure 3a). After short annealing of **MeRPy-10** with the target mixture and CSL, the targeted members were depleted from the supernatant with 88±4% efficiency. 98% of non-target strands remained in solution, on average. Nonspecific binding was undetectable for the majority of library components (Figure 3c). Low levels of nonspecific binding were significant only for one library member (strand **H**). After redispersing the pellet in clean buffer solution, the selected library subsets were released via TMSD by addition of either all or only a subset of corresponding release strands. The release efficiency was 90±12%, and the resulting sub-libraries contained the desired strands with a total yield of 79±13%. The recovered target strands were free from non-target contaminations (including strand **H**) within the precision of the measurement (99.8±0.5%). We attribute the high purity to the *dual selection* of target capture and release: target binding first requires the correct binding site sequences (i.e., non-targets stay in the supernatant); TMSD then requires the correct release site sequences (i.e., nonspecifically adsorbed non-targets remain in the pellet).

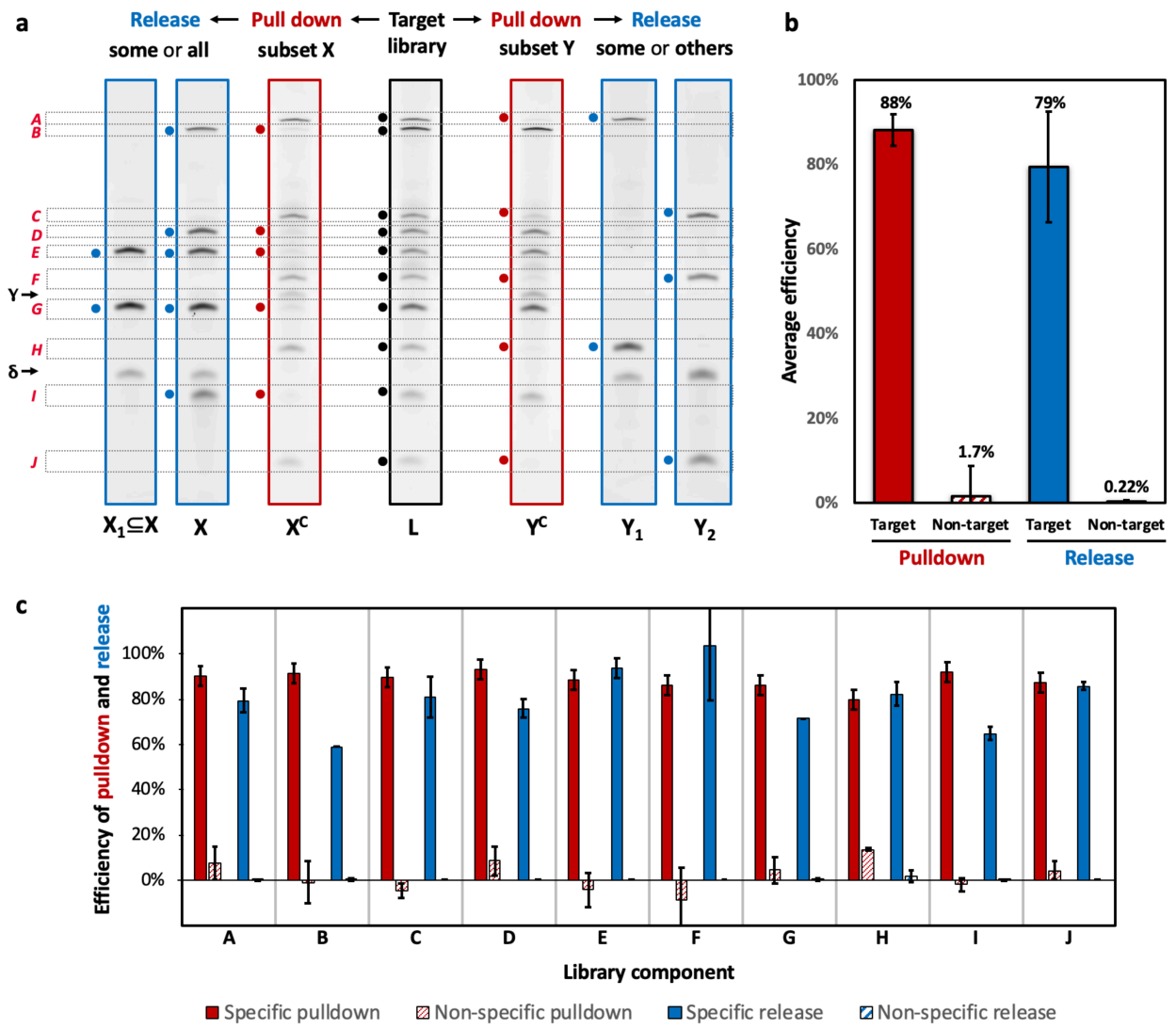

Figure 3. Multiplexed capture and release of selectable targets in a 10-component ssDNA library. **a)** Denaturing polyacrylamide gel electrophoresis (dPAGE) of the library (L) before pulldown (black), after pulldown of selected strands (red), and after release of targeted library subsets (blue). Red and blue circles indicate strands that were targeted by catcher and release strands, respectively. γ = catcher strand band δ = release strand band. **b)** Average binding efficiency and specificity. **c)** Efficiencies and specificities for individual members of the library.

In addition to ssDNA, we sought to apply MeRPy to double-stranded DNA (dsDNA) targets. This aim was motivated by the high demand for tools that allow for sequence-selective depletion of complementary DNA (cDNA) from NGS libraries to enhance efficiency and sensitivity in gene expression profiling.[18,7]

Binding dsDNA targets via hybridization is generally challenging, since rapid re-hybridization of target strands after denaturation hampers the sustained attachment of catcher strands. Our approach involved thermal target denaturation, annealing with MeRPy and a CSL, followed by MeRPy pulldown. To slow down the rate of sense-

antisense rehybridization the CSL was designed to tile the entire target sequence in an alternating manner (Figure 4a). The redundancy of catcher strands also ensures that not only full-length targets but also target fragments are depleted.

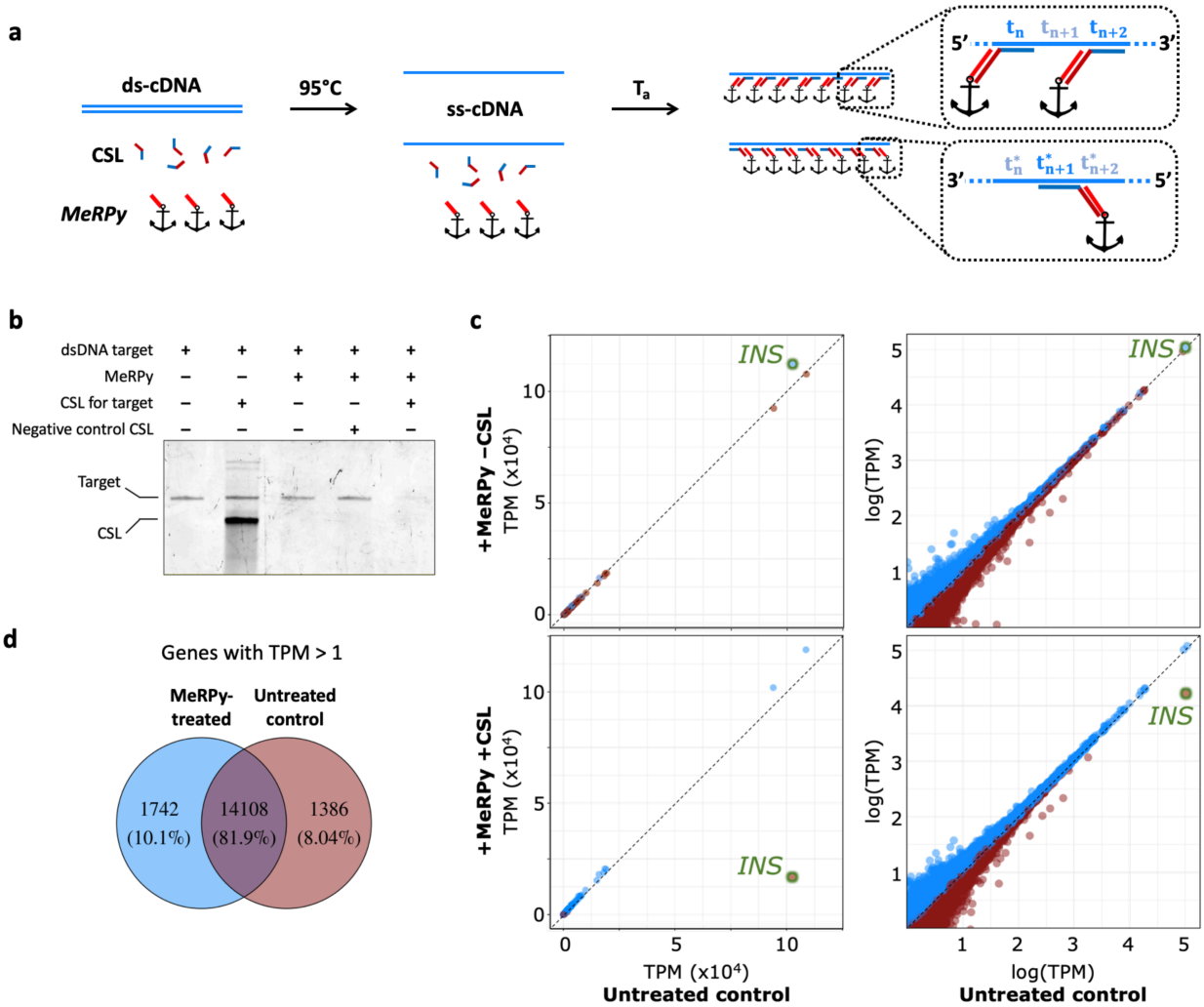

Figure 4. Selective capture of double-stranded DNA (dsDNA) targets. a) The target, a catcher strand library (CSL), and MeRPy are mixed and heated to 95°C. Subsequently, the sample is cooled to bind CSL to the target and MeRPy. MeRPy is quickly precipitated to deplete the target from solution. **b)** Denaturing PAGE after pulldown of a dsDNA target (150 bp) with **MeRPy-100**, showing high pulldown efficiency and specificity. **c)** Selective depletion of high-abundance *insulin* (*INS*) cDNA from a clinical NGS library by MeRPy in presence of an *INS*-specific CSL. Blue and red data points represent genes that were sequenced with higher and lower number of transcripts per million (TPM), respectively, as compared to the original sample (untreated control). **d)** Venn Diagram for all genes above a 1 TPM detection threshold in MeRPy-treated vs. untreated cDNA library.

Early attempts to capture dsDNA targets with **MeRPy-10** were still inefficient. However, **MeRPy-100** allowed up to 10x higher CSL concentrations for increased target binding rates (at the expense of higher material cost; cf. Supplementary Table S2). When using

a short annealing protocol (2 min 95°C | 2 min 4°C), followed by immediate pulldown, target binding became favored over sense-antisense rehybridization. Under these conditions, a depletion efficiency of 89.8±2.4% without detectable levels of non-specific binding was achieved for 150-nt dsDNA (Figure 4b).

To demonstrate practical application of the method, we used MeRPy for targeted depletion of *insulin* (*INS*) cDNA from clinical NGS libraries, which had been generated from patient-derived pancreatic islets.[19] Owing to its high expression level, *INS* consumes a large fraction of NGS reads, thus affecting sequencing depth (i.e., data quality) for all other transcripts in the library, some of which carry diagnostically relevant information.[20]

MeRPy pulldown depleted *INS* from the library with ~80% efficiency and high selectivity (Figure 4c). 91% of reads uniquely mapped to the human genome, independent of MeRPy treatment. Therefore, application of MeRPy and CSL did not generate interfering background for NGS. Removal of *INS* from the complex sample made available reads for an additional 80,000 transcripts per million (TPM) (Supplementary Figure S5), which increased the sequencing depth for 62% of genes in the library without introducing biases or distortions in the expression profile (Figure 4c, Supplementary Figure S6). Moreover, owing to the effective increase in average sequencing depth, a net surplus of ~350 previously undetected genes with TPM > 1 were sequenced in *INS*-depleted samples (Figure 4d).

Conclusions

In summary, MeRPy enables multiplexable sequence-selective enrichment of DNA targets and depletion of undesired sequences from complex mixtures. Ready-to-use MeRPy solutions contain up to 100 µM anchor strands, thus offering exceptionally high binding capacity and rate, combined with low non-specific adsorption. MeRPy was synthesized for $0.27–0.31 material costs per nanomole binding capacity (Supplementary Table S2). The low cost is a prerequisite for large-scale and high-throughput applications. Unlike widely used temperature-responsive poly(N-isopropylacrylamide) (PNIPAM) and other DNA-grafted polymers,[21] MeRPy's precipitation is triggered by methanol. Its insensitivity towards temperature is important, as it allows for *in-situ* thermal target denaturation and annealing steps that are crucial for achieving high capture efficiency and specificity. MeRPy pulldown is applicable to single- as well as double-stranded DNA targets. Successive target hybridization and TMSD release provided dual sequence-selectivity for retrieval of high-purity target libraries. Depletion of high-abundance *INS* cDNA from NGS libraries was simple and fast (~10 minutes experimental time). The assay enabled sequencing of low-abundance transcripts in clinical samples that had not been detected in untreated samples at identical sequencing depth. We anticipate that MeRPy will be a useful tool for enhancing the quality and diagnostic value of transcriptomic signatures,[8] sorting of DNA-encoded chemical libraries,[22] purification of components for DNA nanotechnology,[23] as well as isolation of DNA from crude biological samples. Moreover, MeRPy can be used as a stimulus-responsive "macroprimer" in preparative PCR amplifications.[24]

Acknowledgements

E.K. acknowledges support from the Human Frontier Science Program (LT001077/2015-C) and the Open Topic Postdoc Program of TU Dresden. E.K. would like to thank Prof. Dr. Carsten Werner for support and helpful discussions. W.M.S. acknowledges support from the Wyss Institute at Harvard Core Faculty Award. We thank Prof. Dr. Michele Solimena for sharing cDNA library samples and the Deep Sequencing Group (SFB 655/BIOTEC) at the Biotechnology Center TU Dresden for NGS services.

Literature

1. Psifidi, A. *et al.* Comparison of Eleven Methods for Genomic DNA Extraction Suitable for Large-Scale Whole-Genome Genotyping and Long-Term DNA Banking Using Blood Samples. *PLoS ONE* **10**, (2015).
2. Praetorius, F. *et al.* Biotechnological mass production of DNA origami. *Nature* **552**, 84–87 (2017).
3. Krieg, E. & Shih, W. M. Selective Nascent Polymer Catch-and-Release Enables Scalable Isolation of Multi-Kilobase Single-Stranded DNA. *Angew. Chem. Int. Ed.* **57**, 714–718 (2018).
4. Pinto, A., Chen, S. X. & Zhang, D. Y. Simultaneous and stoichiometric purification of hundreds of oligonucleotides. *Nat. Commun.* **9**, 2467 (2018).
5. Svobodová, M., Pinto, A., Nadal, P. & Sullivan, C. K. O. Comparison of different methods for generation of single-stranded DNA for SELEX processes. *Anal. Bioanal. Chem.* **404**, 835–842 (2012).
6. Patanjali, S. R., Parimoo, S. & Weissman, S. M. Construction of a uniform-abundance (normalized) cDNA library. *Proc. Natl. Acad. Sci.* **88**, 1943–1947 (1991).
7. Gu, W. *et al.* Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol.* **17**, (2016).
8. Boone, M., De Koker, A. & Callewaert, N. Capturing the 'ome': the expanding molecular toolbox for RNA and DNA library construction. *Nucleic Acids Res.* **46**, 2701–2721 (2018).
9. Nimwegen, K. J. M. van *et al.* Is the $1000 Genome as Near as We Think? A Cost Analysis of Next-Generation Sequencing. *Clin. Chem.* **62**, 1458–1464 (2016).
10. Adiconis, X. *et al.* Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat. Methods* **10**, 623–629 (2013).
11. Yi, H. *et al.* Duplex-specific nuclease efficiently removes rRNA for prokaryotic RNA-seq. *Nucleic Acids Res.* **39**, e140–e140 (2011).
12. Freitag, R., Schumacher, I. & Hilbrig, F. Affinity precipitation an option for early capture in bioprocessing. *Biotechnol. J.* **2**, 685–690 (2007).

13. Yurke, B. & Jr, A. P. M. Using DNA to Power Nanostructures. *Genet. Program. Evolvable Mach.* **4**, 111–122 (2003).

14. *Technical specification of Thermo Scientific Dynabeads M-270, M-280, MyOne Streptavidin C1, and MyOne Streptavidin T1*.

15. Nilsson, L. Separation and characterization of food macromolecules using field-flow fractionation: A review. *Food Hydrocoll.* **30**, 1–11 (2013).

16. Burchard, W. Solution Properties of Branched Macromolecules. in *Branched Polymers II* (ed. Roovers, J.) 113–194 (Springer Berlin Heidelberg, 1999). doi:10.1007/3-540-49780-3_3

17. Lindahl, T. & Nyberg, B. Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **11**, 3610–3618 (1972).

18. Archer, S. K., Shirokikh, N. E. & Preiss, T. Selective and flexible depletion of problematic sequences from RNA-seq libraries at the cDNA stage. *BMC Genomics* **15**, (2014).

19. Sturm, D. *et al.* Improved Protocol For Laser Microdissection Of Human Pancreatic Islets From Surgical Specimens. *J. Vis. Exp. JoVE* (2013). doi:10.3791/50231

20. Solimena, M. *et al.* Systems biology of the IMIDIA biobank from organ donors and pancreatectomised patients defines a novel transcriptomic signature of islets from individuals with type 2 diabetes. *Diabetologia* **61**, 641–657 (2018).

21. Peng, L. *et al.* Engineering and Applications of DNA-Grafting Polymer Materials. *Chem. Sci. R. Soc. Chem. 2010* **4**, 1928–1938 (2013).

22. Brenner, S. & Lerner, R. A. Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci.* **89**, 5381–5383 (1992).

23. Pinheiro, A. V., Han, D., Shih, W. M. & Yan, H. Challenges and opportunities for structural DNA nanotechnology. *Nat. Nanotechnol.* **6**, 763–772 (2011).

24. Minev, D. *et al.* Rapid and scalable in vitro production of single-stranded DNA. *bioRxiv* 558429 (2019). doi:10.1101/558429