JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

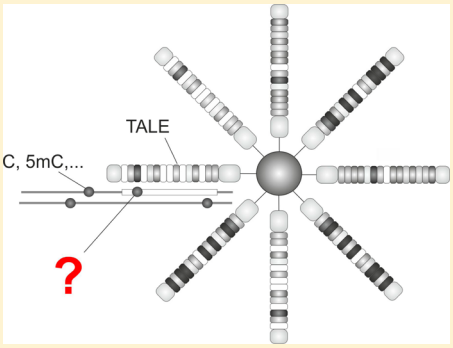
Isolation of Human Genomic DNA Sequences with Expanded Nucleobase Selectivity

Preeti Rathi, Sara Maurer, Grzegorz Kubik, and Daniel Summerer*

Department of Chemistry and Chemical Biology, Technical University of Dortmund, Otto-Hahn-Str. 6, 44227 Dortmund, Germany

Supporting Information

ABSTRACT: We report the direct isolation of user-defined DNA sequences from the human genome with programmable selectivity for both canonical and epigenetic nucleobases. This is enabled by the use of engineered transcription-activator-like effectors (TALEs) as DNA major groove-binding probes in affinity enrichment. The approach provides the direct quantification of 5-methylcytosine (5mC) levels at single genomic nucleotide positions in a strand-specific manner. We demonstrate the simple, multiplexed typing of a variety of epigenetic cancer biomarker 5mC with custom TALE mixes. Compared to antibodies as the most widely used affinity probes for 5mC analysis, i.e., employed in the methylated DNA immunoprecipitation (MeDIP) protocol, TALEs provide superior sensitivity, resolution and technical ease. We engineer a range of size-reduced TALE repeats and establish full selectivity profiles for their binding to all five human cytosine nucleobases. These provide insights into their nucleobase recognition mechanisms and reveal the ability of TALEs to isolate genomic target sequences with selectivity for single 5-hydroxymethylcytosine and, in combination with sodium borohydride reduction, single 5-formylcytosine nucleobases.



1. INTRODUCTION

The epigenetic nucleobase 5-methylcytosine (5mC, Figure 1a) exists in cytosine-guanine (CpG) dinucleotides in the human genome and acts as a dynamic regulatory element of gene expression with important roles in development and diseases, including cancer.[1,2] Cytosines with variable methylation in disease-related canonical nucleobase sequences are thus important epigenetic biomarkers with high potential for cancer diagnosis/prognosis.[3] Moreover, ten-eleven translocation (TET) enzymes oxidize 5mC to the three additional derivatives 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC, Figure 1a)[4−8] as intermediates of an active demethylation pathway.[9] Of these, 5hmC is stable, abundant, and differs between cancer and noncancer cells.[4,5,10] Moreover, biochemical studies have revealed differential interactions of individual cytosine nucleobases with methyl-CpG-binding domains (MBD)[11] and eukaryotic RNA polymerase,[12,13] and indicated an influence on nucleosome stability.[14] Finally, fishing studies have revealed different protein interaction profiles of the five nucleobases.[15,16] Further insights into the biological function of 5mC and other epigenetic cytosine nucleobases requires methods for their simple and highly resolved analysis in user-defined genomic sequences.

C and 5mC can be differentiated from one another by selective deamination with bisulfite followed by sequence analysis,[17] and this can be coupled with additional pretreatment steps to reveal oxidized derivatives of 5mC.[18−22] Alternatively, tagging reactions as well as proteins selective for 5mC or its derivatives such as antibodies, methyl-CpG-binding domains (MBD) or modifying enzymes can be used for global isolation of all modified DNA and subsequent sequence analysis.[3]

However, though these techniques are widely established and invaluable for epigenetic nucleobase analysis, they are not directly selective for epigenetic nucleobases within their canonical nucleobase sequences, that is, for actual epigenetic biomarkers. Such direct approaches bear potential for a particularly simple analysis of untreated DNA with high resolution. One attractive strategy in this direction is the use of processive DNA-reading proteins such as nanopores[23−25] or DNA polymerases[26,27] for single molecule sequencing, which enables a direct sensing of epigenetic nucleobases in a given canonical sequence. However, their routine application for the direct analysis of human genomic 5mC has not yet been demonstrated. A complementary strategy could be the development of DNA-binding affinity probes as analytical tools with expanded programmable selectivity (i.e., for both canonical and epigenetic nucleobases).[28,29] A potential scaffold for the design of such probes are transcription-activator-like effector proteins (TALEs).[30,31] These recognize one strand of double-stranded DNA via the major groove[32,33] that displays unique chemical information not only for each canonical base pair, but also for each epigenetic cytosine nucleobase.[34] TALEs consist of multiple concatenated repeats, each of which selectively recognizes one canonical nucleobase through one of two variable amino acids (repeat variable diresidue, RVD). This recognition follows a simple code with the RVDs NI, NN,

Received: May 10, 2016
Published: July 18, 2016

DOI: 10.1021/jacs.6b04807
J. Am. Chem. Soc. 2016, 138, 9910−9918

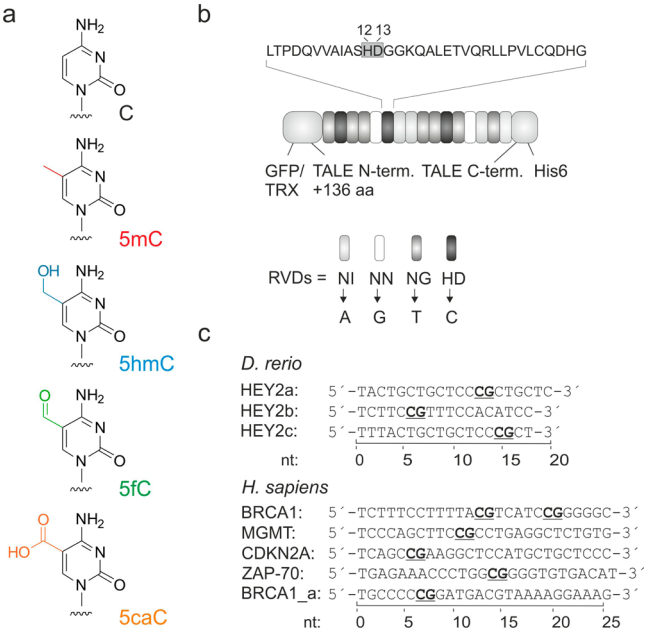


Figure 1. TALEs as DNA-binding probes with expanded programmable nucleobase selectivity. (a) Chemical structures of C and the epigenetic nucleobases 5mC, 5hmC, 5fC, and 5caC. (b) Cartoon showing the general features of employed TALEs. Amino acid sequence of one representative TALE repeat is shown with RVD amino acids 12 and 13 marked with a gray box. TALE repeats with selectivity for canonical nucleobases are shown below with RVDs specified. (c) Target sequences of employed TALEs with CpG dinucleotides targeted in enrichment experiments shown bold and underlined.

NG and HD (amino acid positions 12 and 13 within the TALE repeat) preferentially binding A, G, T, and C, respectively[35,36] (Figure 1b). Alternative RVDs with partially improved selectivities have been identified by systematic mutagenesis studies.[37,38] The simple and major-groove directed recognition mode of TALEs offers a unique potential for the design of programmable scaffolds with specific repeat codes for epigenetic nucleobases and consequently for their direct and simple analysis at user-defined sites. However, TALEs so far exhibited meaningful selectivities in vitro only in oligonucleotides and in a DNA polymerase accessibility model that merely indicated an influence of single 5mC or 5hmC on the local stability of TALE-DNA complexes at the very N-terminus (SI Figure S12).[28,29,39−41] A basic and universal application of TALEs that could be integrated with a large number of DNA analysis techniques would be the direct isolation of user-defined DNA sequences from genomes by affinity enrichment with selectivity for single epigenetic nucleobases. This could offer a resolution and quantitativeness not offered by previous affinity probes, such as antibodies. However, besides high canonical sequence selectivity, this requires a strong impact of single epigenetic nucleobases on the state of the equilibrium between fully DNA-bound and -unbound TALEs. This impact is poorly understood, and electromobility shift assays (EMSA) capable of visualizing this equilibrium so far revealed meaningful selectivity only in heavily 5mC-modified oligonucleotide models without biological relevance.[41,42] This leaves the potential of TALEs for real-world epigenetic nucleobase analyses in the human genome unexplored.

Here, we report the use of TALEs as programmable affinity probes for solid-phase isolation of user-defined DNA sequences from the human genome with direct selectivity for single epigenetic cytosine nucleobases. We characterize the analytical parameters of this approach in a variety of cancer biomarker sequences, revealing that it enables the strand-specific, quantitative analysis of 5mC levels at single nucleotide positions and that it is amenable for multiplexed analysis by the employment of custom TALE mixes. A benchmark against antibodies in the methylated DNA immunoprecipitation (MeDIP) protocol as the gold standard of 5mC affinity probes reveals superior sensitivity, resolution and technical ease of TALEs. We further engineer a range of size-reduced TALE repeats and establish full selectivity profiles for their binding to all five human cytosine nucleobases. This provides insights into the structural requirements for selective epigenetic nucleobase recognition and reveals the ability of TALEs to isolate genomic target sequences with selectivity for single 5hmC and, in combination with sodium borohydride reduction, single 5fC nucleobases.

2. EXPERIMENTAL SECTION

2.1. Construction of GFP-TALE Entry Plasmid and Golden Gate Assembly. Three DNA fragments obtained (1) by PCR on TALE entry plasmid pSaB_339[40] with primers oAnI_1366 and oAnI_1311, (2) by PCR on a MBP gene using primers oDaS_1139 and oDaS_1140 and (3) by primer extension on annealed oligonucleotides oDaS_1143 and oDaS_1144 were combined by Gibson assembly (for all oligonucleotide sequences, see the SI). This resulted in TALE entry plasmid pAnI_509 having the N-terminal TRX tag of pSaB_339 replaced by an MBP tag followed by an AvrBs3 sequence preceding the TALE DNA binding domain of pSaB_339. Two fragments obtained by PCR (1) on pAnI_509 using primers oAnI_1366 and oAnI_1311 and (2) on a GFP gene using primer pairs oAnI_1310 and oAnI_1365 were combined by Gibson assembly to replace the MBP tag with a GFP tag, resulting in the final entry plasmid pAnI_521. Plasmid pAnI_521 and the fragment generated in a PCR with the plasmid using primers oPrR_1916 and oPrR_1917 were restricted using *NdeI* and *NotI* and ligated, resulting in a final entry plasmid pPrR_773 with an N-terminal His-tag. TALEs were assembled according to a previously published protocol[43] using entry plasmid pSaB_339 or pAnI521 in golden gate 2 reactions, resulting in plasmids coding for the respective TALE proteins with TRX or GFP tag and a N- or C-terminal His6 tag.

2.2. Electromobility Shift Assays. Oligonucleotide oGrK_466 was 5′-^{32}P-end labeled using γ-^{32}P-ATP and T4 polynucleotide kinase (Fermentas) and purified using a G-25 gel filtration column (GE Healthcare). oGrK_466 was hybridized pairwise with oligonucleotides oGrK_476, oGrK_465, or oGrK_520 (containing target sequence HEY2b with a single C, 5mC, or 5hmC at position 6) at a concentration of 12.5 nM, respectively, by incubating at 95 °C for 5 min and subsequently at room temperature for 30 min in Hybridization buffer (40 mM Tris-HCl (pH = 8.0), 100 mM NaCl, 10 mM $MgCl_2$, 100 ng/μL salmon sperm DNA (Ambion), 0.2 mg/mL BSA, 10% glycerol). For binding to an off-target sequence, hybridization was performed with oGrK702 and oGrK787 (containing target sequence HEY2c). For TALE-binding, to 6 μL of the hybridized DNA duplexes, 6 μL of TALE storage buffer (20 mM Tris-HCl (pH = 7.5), 200 mM NaCl, 10% glycerol, 1 mM DTT) with varying concentrations of TALE proteins were added and mixtures were incubated at room temperature for 30 min. Mixtures were run on an analytical, nondenaturing PAGE-gel (0.5 mm, 50 cm length) in 0.25 × TB buffer at 12 W, and room temperature Gel was dried and data recorded on a phosphorimager.

2.3. Expression and Purification of TALE Proteins. TALE proteins with TRX tag were expressed and purified as described previously.[40] For expression and purification of TALEs with GFP tag, overnight cultures of *E. coli* BL21(DE3) Gold transformed with a TALE expression plasmid grown in LB media supplemented with 50 μg/mL carbenicillin at 37 °C was diluted 50-fold into LB medium

DOI: 10.1021/jacs.6b04807
J. Am. Chem. Soc. 2016, 138, 9910−9918

supplemented with the same antibiotic and was incubated at 37 °C and 200 rpm shaking until an OD_{600} of ~0.4 was reached. IPTG was added to a concentration of 0.2 mM and the culture was harvested by centrifugation after 5 h of further incubation under the same conditions. The pellet was lysed in Lysis-buffer (10 mM Tris-HCl, 300 mM NaCl, 2.5 mM $MgCl_2$, 0.1% Triton X-100, pH = 9) containing 1 mM PMSF and 50 μg/mL lysozyme (Sigma-Aldrich) by shaking at room temperature at 1400 rpm for 30 min. The suspension was pelleted by centrifugation, the supernatant was collected and extracted with Ni-NTA (Thermo Scientific). Ni-NTA was washed two times with 4 × PBS-Buffer (0.55 M NaCl, 43 mM KCl, 69 mM $Na_2HPO_4{\cdot}2H_2O$, 24 mM KH_2PO_4, pH = 8), four times with wash buffer (50 mM $NaH_2PO_4{\cdot}H2O$ and 300 mM NaCl pH = 8) containing 20 mM imidazole and once with wash buffer containing 50 mM imidazole. The protein was eluted three times with wash buffer containing 500 mM imidazole. Pooled elution fractions were added to a dialysis tube (Carl Roth) and dialyzed against TALE Storage Buffer (20 mM Tris-HCl pH = 7.5, 200 mM NaCl, 10% glycerol, 1 mM DTT). Purity of the TALE protein was analyzed on SDS PAGE stained with Gelcode blue (Thermo Scientific) and quantified by a BCA assay (Pierce). The proteins were snap-frozen and stored in aliquots at −80 °C in TALE storage buffer including 0.1 mg/mL bovine serum albumin (BSA, New England Biolabs).

2.4. Primer Extension Reactions. Reactions were performed as described previously.[40,41] Briefly, templates oPrR_1076, oPrR_1077, oPrR_1184, oPrR_1185, oPrR_1080, oPrR_1079, oPrR_1083, oPrR_1084, oPrR_1826, oPrR_1827, oPrR_1805 or oPrR_1835 and respective primers oPrR_1078, oPrR_1082, oPrR_1085, oPrR_1828 or oPrR_1803 (5′-^{32}P-end labeled) were annealed at 100 nM and 33.3 nM in 6 μL Hybridization buffer (40 mM Tris-HCl (pH = 8.0), 100 mM NaCl, 10 mM $MgCl_2$, 0.2 mg/mL BSA, 10% glycerol) by incubation at 95 °C for 5 min and then at room temperature for 30 min. TALE proteins were added in 6 μL TALE storage buffer (200 mM NaCl, 20 mM Tris-HCl (pH = 7.5), 1 mM DTT, 10% glycerol) at varying concentrations and mixtures were incubated at room temperature for 30 min. Twelve μL of a mixture of 250 mU Klenow fragment of *E. coli* DNA polymerase I (3′-5′-exo-, New England Biolabs) and 200 μM of each dNTP in binding buffer (30 mM Tris-HCl (pH = 8.0), 150 mM NaCl, 5 mM $MgCl_2$, 0.1 mg/mL BSA, 0.5 mM DTT, 7.5% glycerol) were added and the mixtures were incubated at room temperature for 15 min. Twelve μL PAGE loading buffer (80% formamide, 2 mM EDTA) were added and the mixtures were incubated at 95 °C for 5 min and cooled on ice. Mixtures were analyzed by denaturing PAGE gel (9 M urea) using a phosphorimager. For assay setup, see SI Figure S12.

2.5. Genomic DNA Samples. Human gDNA (male Yoruban individual, Encode entry NA18507) was obtained from the Coriell Institute and zebrafish gDNA was extracted from tissue of a single individual (Konstanz wild type).[41] For erasing epigenetic cytosine modifications, gDNA samples were whole genome amplified using the REPLI-g mini kit (Qiagen) and purified using the QIAamp DNA mini kit (Qiagen). gDNA samples were sheared by sonication (Branson Sonifier 250, 20% power, 20 cycles of 30 s on/30 s off), resulting in a broad fragment size distribution between 100−1000 bp. (SI Figure S3). Part of the fragmented gDNA samples (original gDNA) were methylated using 20 U *M.Sss*I CpG methyltransferase (New England Biolabs) in 200 μL Buffer NEB2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH = 7.9) supplemented with 640 μM S-adenosylmethionine (SAM) by incubation at 37 °C for 14 h (after 4 h, additional 4 μL of 32 mM SAM were added) and purified using the QIAamp DNA mini kit (Qiagen). Fractions of both original and enzymatically methylated gDNA were bisulfite-converted with the EpiTect Bisulfite Kit according to the manufacturer's protocol (Qiagen). For each sample, a representative TALE target region was amplified via nested PCR using primer pairs oDaS_754/oDaS_751 and then oDaS_756/oDaS_757 for zebrafish (targeting the HEY2 locus) and primer pairs oPrR_1307_754/oPRR_1308 and then oPrR_1255/oPrR_1256 for human (targeting the BRCA1 locus). Each one strand of the HEY2 PCR product and of the BRCA1 PCR product was sequenced by Sanger sequencing using primer oDaS_757 and primer oPrR_1257, respectively. These sequencings confirmed both the absence of CpG methylation in original gDNA and the quantitative CpG methylation in enzymatically methylated gDNA. For spike-in experiments on 5hmC (or indirect 5fC) detection, amplicons were generated by PCR on zebrafish original gDNA with primer pairs oGrK_476/oGrK_1597, oGrK_465/oGrK_1597, oGrK_520/oGrK_1597, and oGrK_1423/oGrK_1597 introducing a single C, 5mC, 5hmC or 5fC at position 6 of the target sequence HEY2b (Figure 1c) within the PCR product. For MeDIP enrichments, PCRs were conducted with primer pairs oGrK_476/oGrK_1597, oGrK_465/oGrK_1597, oGrK_1057/oGrK_1597, or oGrK_1056/oGrK_1597 to introduce zero, one, two, or three hemimethylated 5mC-positions. For a spike-in containing 15 fully methylated 5mC-positions, amplicons obtained from PCR with primers oGrK_476/oGrK_1597 were enzymatically methylated with *M.Sss*I. For experiments on strand-specificity, the amplicons were produced by PCR on human original gDNA with primer pairs oGrK_1516/oGrK_1601 or oGrK_1518/oGrK_1601 and oGrK_1529/oGrK_1599 or oGrK_1600/oGrK_1599 to introduce either a C or 5mC at position 13 of the target sequence of TALE_BRCA1 (on the parallel strand) or at position 7 of the target sequence of TALE_BRCA1_a (on the antiparallel strand) within the PCR product. All amplicons were purified by agarose gel electrophoresis and spiked into original, nonmethylated gDNA that was previously digested with *Nde*I and *ApaL*I (zebrafish) or *Sac*I (human, these recognition sites are present in the target sequences of the HEY2 and BRCA1_a qPCRs) and purified by the PCR purification kit (Thermo Scientific). For 5fC detection, zebrafish gDNA samples containing HEY2b spike-ins bearing either hmC or fC at position 6 were reduced using freshly prepared sodium borohydride solution (250 mM) for 1 h at room temperature with repeated vortexing, centrifugation and opening of the tubes to release hydrogen. Reaction were quenched with sodium acetate (250 mM, pH = 5) for 10 min at room temperature. Reactions were purified by a PCR purification kit (Thermo Scientific). Treated and nontreated samples were then used along with TALE_HEY2b_N* in enrichment assays.

2.6. Genomic Affinity Enrichment using TAL-Effectors. Before each enrichment, Ni-NTA magnetic agarose beads (Qiagen) were freshly washed/equilibrated with buffer X (150 mM NaCl, 30 mM Tris-HCl, 5 mM $MgCl_2$, 0.5 mg/mL BSA, pH = 7.9) at 4 °C using a 12-tube magnetic stand (Qiagen) with keeping the overall volume constant. For enrichment, 2 μL (100 ng/μL) of gDNA, 3 μL (1.3 μM) TALE protein and 1.2 μL 10× buffer X (with 5 mM DTT added freshly) were added to 5.8 μL of nuclease-free water. The mixture was incubated at 25 °C for 30 min and then kept on ice for 30 min. 483 μL of buffer X (with 0.5 mM DTT added freshly) and 5 μL of equilibrated beads were added to each sample and the samples were incubated in a thermo-mixer at 4 °C and 700 rpm shaking for 60 min. The tubes were placed on the magnetic stand for 2 min and the liquid was carefully removed without disturbing the beads. The beads were washed once with 500 μL of prechilled buffer X (with 0.5 mM DTT added freshly) by slowly pipetting up and down once and the tubes were placed on the magnetic stand for 2 min. The liquid was completely removed and 500 μL of nuclease-free water was added. The tube was shaken at 1400 rpm at 95 °C for 5 min. The tubes were cooled down to room temperature, spinned shortly (2−3 s) and placed on the stand. The supernatant was collected in a fresh tube, dried in a speed-vac and resuspended with 50 μL of nuclease-free water. qPCRs were conducted with 12.5 μL GoTaq qPCR Master Mix (Promega), 5 μL of primer pair (1 μM each, for oligonucleotide sequences, see the SI) and 7.5 μL of the sample as template. Copy number quantification for enriched gDNA was done by linear regression from serial dilutions of the respective gDNA with each primer pair. Original and enzymatically methylated gDNA did not differ significantly in qPCR efficiency.

2.7. Methylated DNA Immunoprecipitation (MeDIP). Enrichments were conducted with spike-in DNA samples of the zebrafish HEY2 locus varying in the number and position of 5mC (see above) using a commercially available MeDIP kit (Diagenode) according to the manufacturer's protocol. All enrichments contained internal negative and positive controls provided by the manufacturer and

DOI: 10.1021/jacs.6b04807
J. Am. Chem. Soc. 2016, 138, 9910−9918

showed enrichment efficiencies as expected from the manufacturer's protocol.

3. RESULTS AND DISCUSSION

3.1. Design of a DNA Sequence Isolation Method with Immobilized TALEs. Given the repetitive setup of the TALE scaffold and the diverse interaction sites with DNA in both the terminal and central domains,[32,33] the mechanisms of TALE-DNA complex formation are expectedly intricate,[44] and the impact of single canonical mismatches or epigenetic nucleobases on these mechanisms is poorly understood. Our previous studies revealed an influence of single epigenetic nucleobases on the accessibility of TALE-bound DNA for a DNA polymerase at the very N-terminus, indicating a local complex destabilization (see SI Figure S12 for assay setup).[39−41] However, a prerequisite for genomic DNA isolation with selectivity for single epigenetic nucleobases is a strong influence on the state of the equilibrium between fully DNA-bound and unbound TALEs under assay conditions. To get insights into this influence, we constructed TALE_HEY2b binding to a 18 nt target sequence in the zebrafish (*Danio rerio*) HEY2 gene (Figure 1c) and expressed it in *E. coli* (SI Figures S1−S2). We generally constructed TALEs by golden gate assembly[43] based on a *Xanthomonas axonopodis* scaffold[45] with an N-terminal *E. coli* thioredoxin (TRX)- or green fluorescent protein (GFP)-domain, a shortened, *AvrBs3*-type TALE N-terminus (+136 amino acids starting from canonical repeat 1) and a His6 tag.

We employed TALE_HEY2b in EMSA studies together with DNA oligonucleotide targets containing sequence HEY2b (Figure 1c) and either a C, 5mC, or 5hmC at position 6. This revealed the formation of TALE-DNA complexes with one defined electromobility in the case of C (Figure 2a, see SI Figure S10 for raw data). Reduced binding was observed for a DNA with the off-target sequence HEY2c (Figure 1c), indicating canonical sequence selectivity. However, the presence of a single 5mC or 5hmC did not result in significantly reduced DNA-binding or in visible changes of electromobility of the formed complexes (Figure 2a and SI Figure S10), revealing a poor 5mC/5hmC selectivity.

To increase the stringency of selective complex formation and to provide a novel technique for direct epigenetic nucleobase analysis at user-defined sites, we designed an assay for solid phase DNA sequence isolation by immobilized TALEs. On the basis of experiments with different solid phases and immobilization strategies (SI Figures S8−S9), we immobilized TALEs via a C-terminal His6 tag to Ni-NTA magnetic agarose beads. Sequence isolation was then conducted by in-solution TALE-DNA complex formation, immobilization of complexes on beads and stringent washing, followed by elution and quantification of the enriched gDNA sequences by qPCR (Figure 2b). We employed TALE_HEY2b together with TALEs_HEY2a and c, targeting DNA sequences of 18−20 nt in the HEY2 gene (Figure 1c) for sequence isolation. Using 200 ng nonmethylated input gDNA of one zebrafish individual, we recovered 690−1080 copies of target sequences, corresponding to 0.84 ± 0.27%, 0.53 ± 0.03% and 0.61 ± 0.2% of the input amount (Figure 2c). To test the sequence selectivity of these TALEs, we performed assays with off-target gDNA (human, *Homo sapiens*) and analyzed the enrichment efficiency of the promoters of the two cancer biomarker genes BRCA1 and MGMT. We observed only negligible enrichment (Figure 2c). As cross-experiment, we constructed two TALEs each targeting a 26 nt sequence in the tested promoters of BRCA1 and

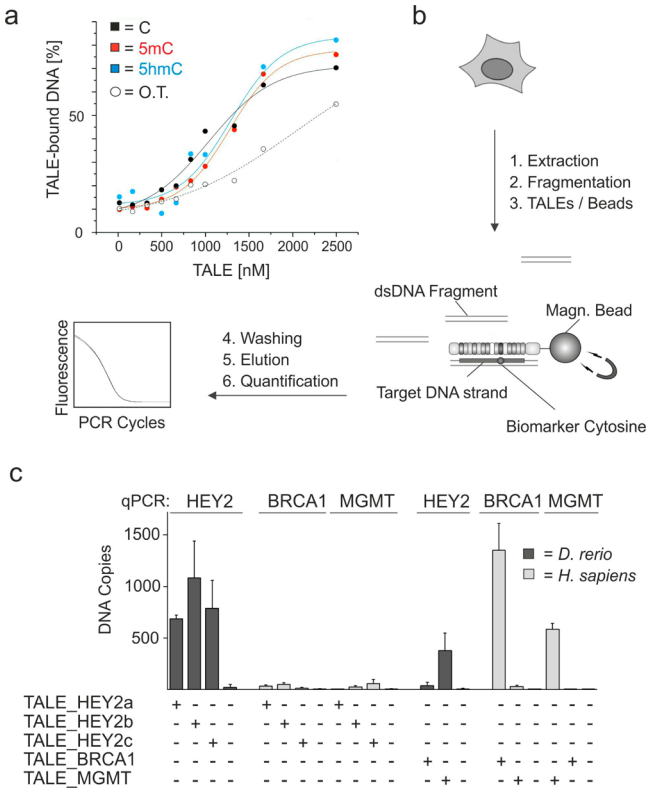


Figure 2. Selective isolation of user-defined genomic DNA sequences with immobilized TALEs. (a) EMSA assay reveals poor selectivity for C, 5mC and 5hmC in in-solution complex formation (for raw data, see SI Figure S10). O. T. = off target DNA sequence HEY2c (Figure 1c). (b) Workflow of TALE-based isolation of user-defined DNA sequences from genomes on the solid phase. (c) Cross-genome and cross-gene enrichment experiments for analysis of TALE-based sequence isolation efficiency and selectivity. Diagram shows target DNA copies obtained from enrichment experiments conducted as shown in (b).

MGMT (TALE_BRCA1 and TALE_MGMT), employed them in assays with zebrafish gDNA and analyzed enrichment of the HEY2 gene. Whereas only negligible enrichment was observed for TALE_BRCA1, some enrichment was visible for TALE_MGMT (Figure 2c), potentially due to a MGMT-related sequence in the HEY2 gene (SI Figure S4). Finally, we analyzed off-target enrichment of these two TALEs within human gDNA in a cross-experiment. We recovered 2.38 ± 0.46% and 1.00 ± 0.10% of the respective input target sequences for TALE_BRCA1 and TALE_MGMT, but observed only negligible enrichment of the respective off-target sequence (Figure 2c). These data demonstrate a high efficiency and selectivity of canonical sequence isolation for most employed TALEs, without the application of design rules to avoid sequence-dependent off-target binding.

3.2. Isolation of Human Genomic DNA Sequences with Selectivity for 5mC. We next aimed to exploit the 5-methylation-sensitive interaction between TALE RVD HD and C to differentiate between C and 5mC (Figure 3a, b).[39−41,46,47] We conducted assays with TALEs_HEY2a−c using non-methylated or methylated zebrafish gDNA as above (SI Figures S5−S6). The target sequences contained a single CpG dinucleotide in varying sequence contexts and positions within the TALE-DNA complex (Figure 1c). We observed discrimination factors of 8.2-fold, 5.2-fold and 4.6-fold (Figure 3c; for

DOI: 10.1021/jacs.6b04807
J. Am. Chem. Soc. 2016, 138, 9910−9918

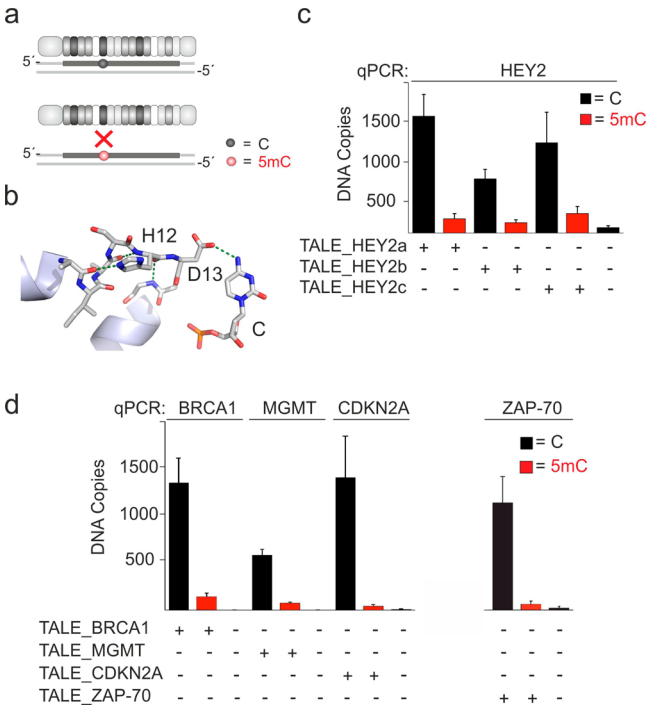


Figure 3. The expanded nucleobase selectivity of TALEs enables the robust analysis of single 5mC nucleobases at user-defined genomic positions by direct sequence isolation. (a) Cartoon showing the principle of programmable 5mC-sensitivity of TALEs based on RVD HD. Black sphere = C, red sphere = 5mC. (b) Interaction of RVD HD (amino acids 12 and 13 of TALE repeat) with cytosine (C) in a crystal structure of a TALE-DNA complex (pdb entry 3V6T).[32] Hydrogen bonds are shown as dotted green lines. (c) Detection of single 5mC in user-defined sequences in the zebrafish genome by solid-phase isolation with TALEs as shown in Figure 2b. (d) Detection of single 5mC in user-defined cancer biomarker sequences in the human genome by solid-phase isolation with TALEs as shown in Figure 2b.

reference data in oligonucleotides, see SI Figure S7). We next conducted analogous assays with gDNA of a human individual using TALE_BRCA1, TALE_MGMT and the additional TALE_CDKN2A, all targeting promoters of biomarker tumor-suppressor genes with known aberrant methylation in cancer tissue.[48] Target sequences were 26 nt and contained a single CpG (Figure 1c; TALE_BRCA1 contained two CpG). Despite the (compared to TALEs_HEY2a−c) increased number of repeats and the resulting increased ratio of matched vs C/5mC-mismatched repeat-nucleobase interactions, we observed high discrimination factors of 9.5-fold, 8.0-fold and 34.4-fold (Figure 3d; for reference data in oligonucleotides, see SI Figure S7). These data show that direct sequence isolation in combination with the 5mC selectivity of TALE RVD HD enables the simple and robust analysis of single 5mC nucleobases at user-defined positions in large genomes.

We next tested our approach for the detection of a validated clinical biomarker used for prognostics in chronic lymphocytic leukemia (CLL), the most prevalent leukemia in adults.[49] Expression of the oncogene zeta-chain-associated protein kinase (ZAP) 70 is a key factor in CLL pathology[50] and loss of 5mC at a single CpG in the sense strand close to the transcription start site (CpG + 223) is predictive for a poor outcome of CLL.[51]

However, a direct and selective isolation of this biomarker from genomes is currently prevented by the low resolution of available affinity probes. We designed TALE_ZAP-70, binding a 26 nt region that contains CpG + 223 (Figure 1c) and conducted our assay as above. We observed a discrimination factor of 18.2-fold, demonstrating the direct and selective isolation/analysis of this prognostic biomarker (Figure 3d).

Quantification of the 5mC level at single nucleotide positions is an important analytical goal, since DNA samples typically exhibit heterogeneous 5mC patterns. To evaluate our approach for this ability, we performed enrichments with TALE_-CDKN2A and either nonmethylated, methylated or mixed gDNA. We observed a linear dependence between the 5mC level at the target position and the amount of isolated gDNA, indicating the ability to quantify genomic 5mC at single positions (Figure 4a).

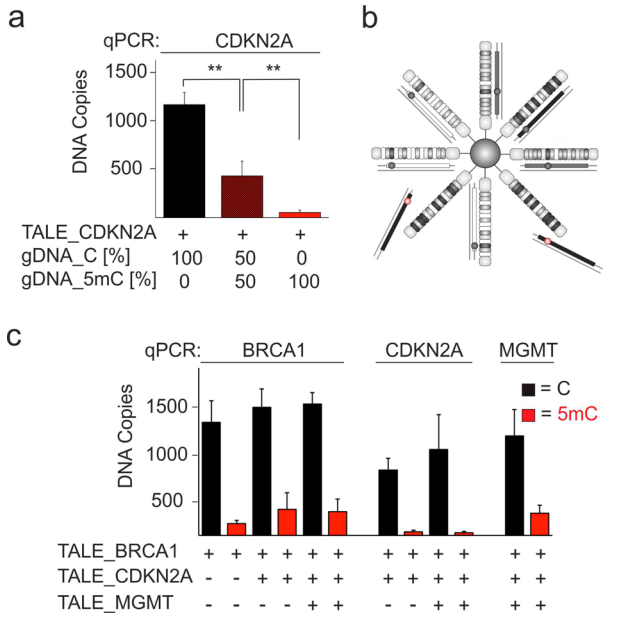


Figure 4. TALE-based isolation of human genomic sequences for quantitative and multiplexed analysis of single cancer biomarker 5mC. (a) Quantification of 5mC at a single position in the human CDKN2A gene. Sequence isolation with TALEs was conducted with non-methylated, mixed or methylated human gDNA. ** = $p < 0.01$ from Student t test. (b) Cartoon showing principle of multiplexed sequence enrichment with immobilized custom TALE mixes. Black sphere = C, red sphere =5mC, bars = target DNA strand with biomarker sequence. (c) Multiplexed detection of 5mC at single sites in the human BRCA1, CDKN2A and MGMT genes. Sequence isolation was conducted with single or up to three mixed TALEs as indicated.

Our assay exclusively involves standard liquid handling and magnetic bead separation steps and should thus be scalable and automatable. Moreover, profiling of multiple CpG in multiple target sequences within a single enrichment may be feasible by using custom TALE mixes (Figure 4b). As proof of principle, we performed experiments with 1−3 mixed TALEs in single enrichments using 200 ng input gDNA (Figure 4c). We observed robust discrimination in all cases, independent of the multiplexing grade. This confirms minimal cross-talk between TALEs (Figure 2c), enabling multiplexing without the need for increased gDNA input amounts.

3.3. Comparison of Anti-5mC Antibodies and TALEs As Affinity Probes for DNA Enrichment. CpG dinucleotides can exist in fully, non-, or hemimethylated states as a result of the absence or presence of de novo and maintenance

DOI: 10.1021/jacs.6b04807
J. Am. Chem. Soc. 2016, 138, 9910−9918

methylation. The ability to study the modification symmetry of CpG is thus of biological interest, but requires 5mC detection in a strand-specific manner. To evaluate our approach for this ability, we designed TALE_BRCA1_a that targets the same locus as TALE_BRCA1, but the opposite DNA strand (Figure 1c). In assays using PCR products with defined 5mC in either strand in a full gDNA background, TALE_BRCA1_a did not discriminate between C and 5mC in the opposite DNA strand that was discriminated by TALE_BRCA1 before, but exhibited a discrimination factor of 15.5-fold for its own target DNA strand (Figure 5a and Figure 3d). This shows that TALEs can provide strand-specific detection of 5mC at single sites.

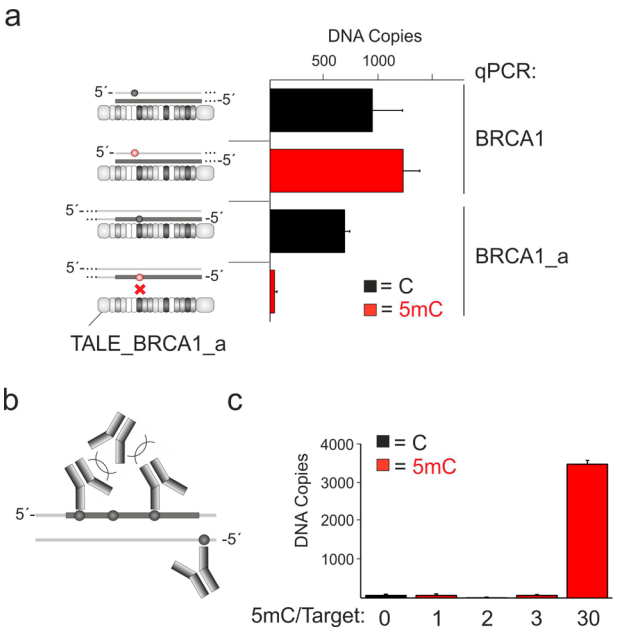


Figure 5. TALE-based analysis of 5mC with strand-specificity and with higher resolution and sensitivity than offered by MeDIP. (a) Strand-specific detection of 5mC in the BRCA1 gene by TALEs. Black sphere: C, red sphere: 5mC. Both sense and antisense strand of target DNA are shown as light gray line, with TALE target sequence as dark gray bar and TALE as cartoon. (b) Cartoon illustrating the absence of positional and strand resolution, as well as the bias of 5mC level quantification with MeDIP. (c) Determination of 5mC sensitivity of anti-5mC-antibodies in the MeDIP protocol in the context of the HEY2 gene sequence using target DNAs with 0, 1, 2, 3, or 30 5mC positions.

We next benchmarked our approach against the widely used, anti-5mC-antibody-based methylated DNA immunoprecipitation (MeDIP) protocol.[52] Owing to the lack of programmable sequence selectivity of antibodies, the resolution of MeDIP is defined by the DNA fragment size of the sample (typically hundreds of nt), it is not strand-specific, prone to density bias due to the large size of antibodies, and thus offers only qualitative data (Figure 5b).[3] Though our previous data show that genomic sequence isolation with TALEs offers superior resolution and quantitativeness than MeDIP, it is not clear which approach provides higher sensitivity. We performed MeDIP assays with spike-in PCR products in the full gDNA background that covered a locus of the zebrafish HEY2 gene and contained 0−3 5mC sites in one strand of the target sequence of TALE_HEY2b. In contrast to TALEs (including TALE_HEY2b) that exhibited single 5mC sensitivity in all previously tested cases (Figure 3c, d), MeDIP did not exhibit 5mC-dependent target sequence enrichment in any of the cases (Figure 5c). We repeated the experiment with the same spike-in sequence, but methylated at all CpG, resulting in 30 5mC sites (i.e., 15 fully methylated CpG). Only in this extreme case enrichment was observed, with recovery of 7% of the input target gDNA sequences (Figure 5c). These data reveal a superior sensitivity of TALEs compared to antibodies in MeDIP. Moreover, MeDIP exhibited a protocol time of >24 h compared to 7.5 h for TALEs, and required significantly more handling steps.

3.4. Size-Reduced TALE Repeats for Isolation of Human Genomic DNA Sequences with Selectivity for 5hmC and 5fC. The observed selectivity of TALEs for genomic C/5mC raise the question, if TALE repeats could be identified that allow for the detection of oxidized 5mC derivatives at single genomic positions by sequence isolation. Crystal structures of the Dickerson−Drew dodecamer containing single 5mC, 5hmC, 5fC or 5caC revealed regular duplex geometries, suggesting the ability of regular complex formation with TALEs. However, they also revealed the presentation of unique chemical information by each of the nucleobases in the major groove close to the RVD (Figure 6a).[34,53,54] Besides the inherent differences of the 5-substituents in the capability to accept or donate hydrogen bonds, there are differences in steric demand. In fact, even 5fC and 5hmC, though similar in size, exhibit marked differences in terms of conformational flexibility. 5hmC exhibits two different conformations in crystal structures,[34] whereas the formyl group of 5fC is fixed in plane with the cytosine ring by hydrogen bonding to N4. This type of interaction is also observed for 5caC (Figure 6a).[55,56]

We have previously shown in oligonucleotides by a DNA polymerase accessibility assay that 5hmC can selectively influence the local stability of TALE-DNA complexes when positioned opposite a TALE repeat N* (* = amino acid deletion at position 13). Similar to RVD NG that is missing a side chain at this position and binds 5mC, but not C (Figure 6a),[42] RVD N* is capable of accommodating a cytosine 5-methyl group, but binds C with similar affinity as 5mC, and 5hmC with only low affinity.[40] However, the interactions of these and other TALE repeats with 5fC and 5caC have not been studied, and the structural requirements for selective sensing of epigenetic cytosine nucleobases by TALEs are not known. To gain insights into these aspects, we engineered a set of additional size-reduced TALE repeats for structure−function relationships. We engineered these repeats into position 6 of TALE_HEY2b and established selectivity profiles including all five human cytosine nucleobases by the DNA polymerase accessibility assay described previously (Figure 6b, SI Figures S11−S12).[39−41] TALE_HEY2b_N* exhibited strong binding of C and 5mC and the expected selective discrimination against 5hmC. Interestingly, 5fC was bound with equal affinity as 5mC, whereas 5caC reduced binding to a similar extent as 5hmC (Figure 6b, note that the graphs are based on single TALE concentrations and underemphasize the true repeat selectivities on the basis of K_i; for N*, these are ~25-fold for C/5hmC and ~30-fold for 5mC/5hmC, whereas C and 5mC differ by only 1.2-fold).[40] These data correspond to the structural similarity of 5mC and 5fC at the face that is presumably oriented toward the TALE repeat (Figure 6a). When taking RVD NG opposite 5mC as model, the steric demand of 5mC and 5fC is likely similar and a methyl or methylene moiety is forming the nucleobase surface at this face. In contrast, the hydroxyl or

DOI: 10.1021/jacs.6b04807
J. Am. Chem. Soc. 2016, 138, 9910−9918

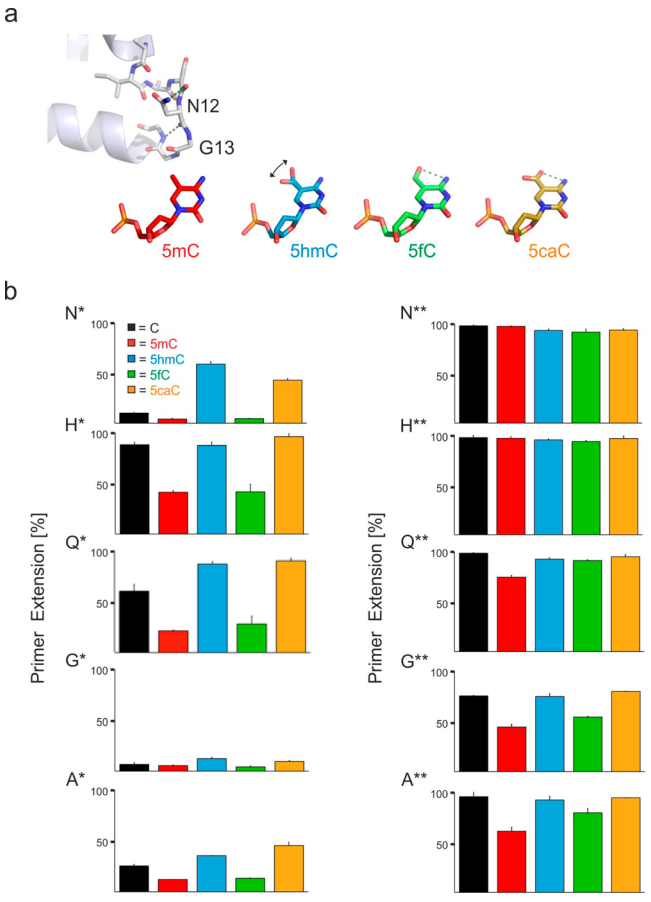


Figure 6. A mutational analysis reveals structural requirements of size-reduced TALE repeats for the selective recognition of epigenetic cytosine nucleobases. (a) Interaction of RVD NG (amino acids 12 and 13 of TALE repeat) with 5mC and structures of 5hmC, 5fC, and 5caC as found in crystal structures of DNA duplexes (pdb entries 4GJP[42] as well as 4I9V, 4QC7, and 4PWM).[34,53,54] Hydrogen bonds are shown as dotted green lines. For 5hmC, the two observed conformations of the hydroxmethyl substituent are shown, with an arrow indicating conformational flexibility. (b) Selectivity profiles for ten size-reduced TALE repeats including all five human cytosine nucleobases. Repeats are defined by the amino acid at position 12 and the presence or absence of deletions (*) at positions 13 and 14, with otherwise identical repeat sequences (panel a shows positions). TALE-binding to oligonucleotide primer template duplexes containing a single variable cytosine in the template strand was measured by inhibition of a DNA polymerase-catalyzed primer extension (SI Figure S12). Primer extension product formation of a reaction w/o TALE was set to 100% and inhibition of primer extension was measured to reveal affinity of TALE-binding.

carboxylate group of 5hmC and 5caC lead to an enlarged, polar surface at this face that may account for reduced TALE binding.

We next tested the influence of amino acid N12 that in canonical TALE repeats is engaged in an intraloop hydrogen bond that contributes to preorganization (Figure 6a). N12 is replaced by an H residue in the canonical TALE repeat HD where it can donate an alternative intraloop hydrogen bond that leads to a slightly altered loop geometry (Figure 3b).[32,33]

The resulting TALE_HEY2b_H* exhibited reduced binding affinity and relaxed selectivity with an overall similar profile. When we installed a Q side chain with similar hydrogen bonding capability as N, but with increased length and flexibility TALE_HEY2b-Q*, we observed a slightly improved overall affinity and selectivity with an overall similar profile. We next replaced N12 by G12 to remove hydrogen bonding and potentially increase the loop flexibility. This led to strongly increased affinity, combined with a relaxed selectivity. To test, if this effect was a result of the lost hydrogen bonding capability or the increased backbone flexibility, we evaluated TALE_HEY2b_A* with an expected decreased flexibility (and slightly increased steric demand). This reduced the affinity but partially restored the selectivity. Finally, we tested combinations of the five amino acids with an additional deletion to remove G14 and potentially create more space for the bulky 5-substituents of 5hmC and 5caC (repeats N**, H**, Q**, G** and A**). This led to significantly reduced binding in all cases, with some residual selectivity in the three latter ones, again with similar selectivity profile (Figure 6b). These combined experiments reveal a dominating role of the amino acid 13 deletion for the overall selectivity profile of TALE repeats. The side chain at position 12 can further modulate both the affinity and selectivity of repeats. Moreover, the presence of G14 is critical for overall affinity, though some selective binding of 5mC and 5fC is retained in the case of the small, nonpolar amino acids G and A at position 12.

We next aimed to exploit the observed selectivity of RVD N* for the detection of single genomic 5hmC by sequence isolation. The selectivity requirements for nucleobase detection in genomic DNA samples are defined by the levels at which the individual nucleobases exist in the cell type under study. All four epigenetic nucleobases have been mapped in a variety of cell types, revealing marked differences. Besides 5mC, only 5hmC has been shown to be stable and to exist at high levels throughout genomes, with high abundances in, e.g., brain and embryonic stem cells.[10] 5fC, though >1−2 orders of magnitude less abundant than 5hmC, can exist at high levels at certain sites.[18] However, previously observed levels of 5caC are so low[57] that its influence within the signal/noise of TALE-based nucleobase analysis is negligible. The relevant cytosine nucleobase range for TALE-based genomic detection thus includes C, 5mC, 5hmC, and 5fC. We performed assays with TALE_HEY2b_N* using PCR spike-ins with C, 5mC, 5hmC, 5fC or 5caC in the full gDNA background. The obtained selectivity of sequence isolation corresponded well to the data obtained by the DNA polymerase accessibility assay (Figure 6b and 7a). C, 5mC and 5fC were enriched with similar efficiencies, whereas the presence of a single 5hmC lead to markedly reduced enrichment (Figure 7a, 5caC also reduced enrichment). This indicates that RVD N* is fully selective for 5hmC in the relevant nucleobase range and that its selectivity can be sufficient for the detection of a single 5hmC nucleobase in genomic DNA samples. To enable detection also of 5fC, we combined the 5fC/5hmC selectivity of RVD N* with sodium borohydride treatment of DNA prior enrichment. This selectively and quantitatively reduces genomic 5fC to 5hmC under mild conditions (Figure 7b).[18] Indeed, the enrichment efficiency of 5fC-containing DNA after reduction was lowered to the one of 5hmC-containing DNA, whereas the efficiency for 5hmC-containing DNA was unaffected by reduction (Figure 7c). This shows that TALE-based sequence isolation in combination with chemical reduction enables the selective detection of a single 5fC position in genomic DNA samples.

4. CONCLUSION

In summary, TALEs enable the first approach for the direct isolation of user-defined DNA sequences from large genomes

DOI: 10.1021/jacs.6b04807
J. Am. Chem. Soc. 2016, 138, 9910−9918

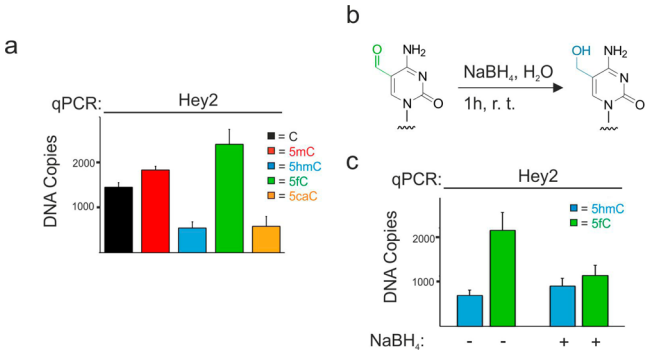


Figure 7. A size-reduced TALE repeat enables the selective detection of 5hmC and 5fC in genomic DNA. (a) Isolation of the HEY2b sequence from zebrafish genomic DNA samples with TALE_-HEY2b_N* in dependence of a single C, 5mC, 5hmC, 5fC, or 5caC opposite RVD N*. (b) Selective reduction of 5fC to 5hmC by sodium borohydride. (c) Isolation of the HEY2b sequence with TALE_-HEY2b_N* as in Figure 6c with zebrafish genomic DNA samples before and after reduction of 5fC to 5hmC with sodium borohydride.

with selectivity for both canonical and epigenetic nucleobases. Compared to previous, nonprogrammable affinity probes such as antibodies, TALEs offer a higher sensitivity, technical ease, and analytical resolution, i.e., the quantification of 5mC at single genomic positions in a strand-specific manner. This enabled the direct analysis of ZAP-70 CpG + 223, a validated clinical biomarker for CLL that cannot be selectively detected with any previous affinity probe. Moreover, we provide insights into the structural requirements of size-reduced TALE repeats for selective recognition of the full range of human epigenetic nucleobases and show first proofs of principle for the direct detection of single genomic 5hmC and 5fC nucleobases. The automatable workflow, ability of multiplexing and high throughput of TALE assembly[58] support a high scalability of the assay. Compared to highly optimized bisulfite-based protocols, TALE-based analysis by sequence enrichment currently exhibits reduced signal/noise ratios and quantitativeness. Multiple aspects of TALE design offer potential for improvements. Besides general adjustments of enrichment conditions, these include the directed evolution of TALE repeats with further improved selectivities as well as balancing the binding forces of the single discriminatory and all other repeats by taking into account the individual affinities of RVD types, the overall repeat number and the polarity of TALE binding.[37,38,59] Similarly, removing electrostatic interactions in the N-terminus and/or in individual repeats provides means to significantly increase selectivity.[60] Future efforts deal with the construction of larger TALE libraries and multiplexed sequence isolation for epigenetic profiling by high throughput sequencing and the exploration of in vivo monitoring of dynamic epigenetic modification at user-defined sites. These efforts will further expand the applicability of TALEs in complementary technologies for epigenetics research and diagnostics/prognostics.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b04807.

Oligonucleotide and protein sequences, data of protein expressions/purifications and of biochemical assays. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*daniel.summerer@uni-konstanz.de

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge support by the TU Dortmund, the Zukunftskolleg of the University of Konstanz and the Konstanz Research School Chemical Biology. We thank Christoph Plass for helpful discussions. We thank A. J. Bogdanove and D. F. Voytas for TALE assembly plasmids obtained via Addgene. This work was supported by grants from the Deutsche Forschungsgemeinschaft (Su 726/5−1 in SPP1784 and Su 726/6−1 in SPP1623).

■ REFERENCES

(1) Law, J. A.; Jacobsen, S. E. *Nat. Rev. Genet.* **2010**, *11*, 204.
(2) Wu, H.; Zhang, Y. *Cell* **2014**, *156*, 45.
(3) Booth, M. J.; Raiber, E. A.; Balasubramanian, S. *Chem. Rev.* **2015**, *115*, 2240.
(4) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930.
(5) Kriaucionis, S.; Heintz, N. *Science* **2009**, *324*, 929.
(6) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1300.
(7) He, Y. F.; Li, B. Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C. X.; Zhang, K.; He, C.; Xu, G. L. *Science* **2011**, *333*, 1303.
(8) Pfaffeneder, T.; Hackner, B.; Truss, M.; Munzel, M.; Muller, M.; Deiml, C. A.; Hagemeier, C.; Carell, T. *Angew. Chem., Int. Ed.* **2011**, *50*, 7008.
(9) Shen, L.; Song, C. X.; He, C.; Zhang, Y. *Annu. Rev. Biochem.* **2014**, *83*, 585.
(10) Bachman, M.; Uribe-Lewis, S.; Yang, X. P.; Williams, M.; Murrell, A.; Balasubramanian, S. *Nat. Chem.* **2014**, *6*, 1049.
(11) Du, Q.; Luu, P. L.; Stirzaker, C.; Clark, S. J. *Epigenomics* **2015**, *7*, 1051.
(12) Wang, L.; Zhou, Y.; Xu, L.; Xiao, R.; Lu, X.; Chen, L.; Chong, J.; Li, H.; He, C.; Fu, X. D.; Wang, D. *Nature* **2015**, *523*, 621.
(13) Kellinger, M. W.; Song, C. X.; Chong, J.; Lu, X. Y.; He, C.; Wang, D. *Nat. Struct. Mol. Biol.* **2012**, *19*, 831.
(14) Ngo, T. T.; Yoo, J.; Dai, Q.; Zhang, Q.; He, C.; Aksimentiev, A.; Ha, T. *Nat. Commun.* **2016**, *7*, 10813.
(15) Spruijt, C. G.; Gnerlich, F.; Smits, A. H.; Pfaffeneder, T.; Jansen, P. W.; Bauer, C.; Munzel, M.; Wagner, M.; Muller, M.; Khan, F.; Eberl, H. C.; Mensinga, A.; Brinkman, A. B.; Lephikov, K.; Muller, U.; Walter, J.; Boelens, R.; van Ingen, H.; Leonhardt, H.; Carell, T.; Vermeulen, M. *Cell* **2013**, *152*, 1146.
(16) Iurlaro, M.; Ficz, G.; Oxley, D.; Raiber, E. A.; Bachman, M.; Booth, M. J.; Andrews, S.; Balasubramanian, S.; Reik, W. *Genome Biol.* **2013**, *14*, R119.
(17) Frommer, M.; Mcdonald, L. E.; Millar, D. S.; Collis, C. M.; Watt, F.; Grigg, G. W.; Molloy, P. L.; Paul, C. L. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 1827.
(18) Booth, M. J.; Marsico, G.; Bachman, M.; Beraldi, D.; Balasubramanian, S. *Nat. Chem.* **2014**, *6*, 435.
(19) Booth, M. J.; Branco, M. R.; Ficz, G.; Oxley, D.; Krueger, F.; Reik, W.; Balasubramanian, S. *Science* **2012**, *336*, 934.
(20) Yu, M.; Hon, G. C.; Szulwach, K. E.; Song, C. X.; Zhang, L.; Kim, A.; Li, X.; Dai, Q.; Shen, Y.; Park, B.; Min, J. H.; Jin, P.; Ren, B.; He, C. *Cell* **2012**, *149*, 1368.
(21) Song, C. X.; Szulwach, K. E.; Dai, Q.; Fu, Y.; Mao, S. Q.; Lin, L.; Street, C.; Li, Y.; Poidevin, M.; Wu, H.; Gao, J.; Liu, P.; Li, L.; Xu, G. L.; Jin, P.; He, C. *Cell* **2013**, *153*, 678.

DOI: 10.1021/jacs.6b04807
J. Am. Chem. Soc. 2016, 138, 9910−9918

(22) Lu, X.; Song, C. X.; Szulwach, K.; Wang, Z.; Weidenbacher, P.; Jin, P.; He, C. *J. Am. Chem. Soc.* **2013**, *135*, 9315.
(23) Stoddart, D.; Heron, A. J.; Mikhailova, E.; Maglia, G.; Bayley, H. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 7702.
(24) Lieberman, K. R.; Cherf, G. M.; Doody, M. J.; Olasagasti, F.; Kolodji, Y.; Akeson, M. *J. Am. Chem. Soc.* **2010**, *132*, 17961.
(25) Wescoe, Z. L.; Schreiber, J.; Akeson, M. *J. Am. Chem. Soc.* **2014**, *136*, 16582.
(26) Flusberg, B. A.; Webster, D. R.; Lee, J. H.; Travers, K. J.; Olivares, E. C.; Clark, T. A.; Korlach, J.; Turner, S. W. *Nat. Methods* **2010**, *7*, 461.
(27) Summerer, D. *ChemBioChem* **2010**, *11*, 2499.
(28) Kubik, G.; Summerer, D. *ACS Chem. Biol.* **2015**, *10*, 1580.
(29) Kubik, G.; Summerer, D. *ChemBioChem* **2016**, *17*, 975.
(30) Boch, J.; Bonas, U. *Annu. Rev. Phytopathol.* **2010**, *48*, 419.
(31) Bogdanove, A. J.; Voytas, D. F. *Science* **2011**, *333*, 1843.
(32) Deng, D.; Yan, C.; Pan, X.; Mahfouz, M.; Wang, J.; Zhu, J. K.; Shi, Y.; Yan, N. *Science* **2012**, *335*, 720.
(33) Mak, A. N. S.; Bradley, P.; Cernadas, R. A.; Bogdanove, A. J.; Stoddard, B. L. *Science* **2012**, *335*, 716.
(34) Szulik, M. W.; Pallan, P. S.; Nocek, B.; Voehler, M.; Banerjee, S.; Brooks, S.; Joachimiak, A.; Egli, M.; Eichman, B. F.; Stone, M. P. *Biochemistry* **2015**, *54*, 1294.
(35) Moscou, M. J.; Bogdanove, A. J. *Science* **2009**, *326*, 1501.
(36) Boch, J.; Scholze, H.; Schornack, S.; Landgraf, A.; Hahn, S.; Kay, S.; Lahaye, T.; Nickstadt, A.; Bonas, U. *Science* **2009**, *326*, 1509.
(37) Miller, J. C.; Zhang, L.; Xia, D. F.; Campo, J. J.; Ankoudinova, I. V.; Guschin, D. Y.; Babiarz, J. E.; Meng, X.; Hinkley, S. J.; Lam, S. C.; Paschon, D. E.; Vincent, A. I.; Dulay, G. P.; Barlow, K. A.; Shivak, D. A.; Leung, E.; Kim, J. D.; Amora, R.; Urnov, F. D.; Gregory, P. D.; Rebar, E. J. *Nat. Methods* **2015**, *5*, 465.
(38) Yang, J.; Zhang, Y.; Yuan, P.; Zhou, Y.; Cai, C.; Ren, Q.; Wen, D.; Chu, C.; Qi, H.; Wei, W. *Cell Res.* **2014**, *24*, 628.
(39) Kubik, G.; Summerer, D. *ChemBioChem* **2015**, *16*, 228.
(40) Kubik, G.; Batke, S.; Summerer, D. *J. Am. Chem. Soc.* **2015**, *137*, 2.
(41) Kubik, G.; Schmidt, M. J.; Penner, J. E.; Summerer, D. *Angew. Chem., Int. Ed.* **2014**, *53*, 6002.
(42) Deng, D.; Yin, P.; Yan, C.; Pan, X.; Gong, X.; Qi, S.; Xie, T.; Mahfouz, M.; Zhu, J. K.; Yan, N.; Shi, Y. *Cell Res.* **2012**, *22*, 1502.
(43) Cermak, T.; Doyle, E. L.; Christian, M.; Wang, L.; Zhang, Y.; Schmidt, C.; Baller, J. A.; Somia, N. V.; Bogdanove, A. J.; Voytas, D. F. *Nucleic Acids Res.* **2011**, *39*, 7879.
(44) Cuculis, L.; Abil, Z.; Zhao, H.; Schroeder, C. M. *Nat. Commun.* **2015**, *6*, 7277.
(45) Miller, J. C.; Tan, S. Y.; Qiao, G. J.; Barlow, K. A.; Wang, J. B.; Xia, D. F.; Meng, X. D.; Paschon, D. E.; Leung, E.; Hinkley, S. J.; Dulay, G. P.; Hua, K. L.; Ankoudinova, I.; Cost, G. J.; Urnov, F. D.; Zhang, H. S.; Holmes, M. C.; Zhang, L.; Gregory, P. D.; Rebar, E. J. *Nat. Biotechnol.* **2011**, *29*, 143.
(46) Bultmann, S.; Morbitzer, R.; Schmidt, C. S.; Thanisch, K.; Spada, F.; Elsaesser, J.; Lahaye, T.; Leonhardt, H. *Nucleic Acids Res.* **2012**, *40*, 5368.
(47) Valton, J.; Dupuy, A.; Daboussi, F.; Thomas, S.; Marechal, A.; Macmaster, R.; Melliand, K.; Juillerat, A.; Duchateau, P. *J. Biol. Chem.* **2012**, *287*, 38427.
(48) Rodriguez-Paredes, M.; Esteller, M. *Nat. Med.* **2011**, *17*, 330.
(49) Hallek, M.; Cheson, B. D.; Catovsky, D.; Caligaris-Cappio, F.; Dighiero, G.; Dohner, H.; Hillmen, P.; Keating, M. J.; Montserrat, E.; Rai, K. R.; Kipps, T. J.; International Workshop on Chronic Lymphocytic, L.. *Blood* **2008**, *111*, 5446.
(50) Del Poeta, G.; Del Principe, M. I.; Consalvo, M. A. I.; Maurillo, L.; Buccisano, F.; Venditti, A.; Mazzone, C.; Bruno, A.; Gianni, L.; Capelli, G.; Lo Coco, F.; Cantonetti, M.; Gattei, V.; Amadori, S. *Cancer* **2005**, *104*, 2743.
(51) Claus, R.; Lucas, D. M.; Ruppert, A. S.; Williams, K. E.; Weng, D.; Patterson, K.; Zucknick, M.; Oakes, C. C.; Rassenti, L. Z.; Greaves, A. W.; Geyer, S.; Wierda, W. G.; Brown, J. R.; Gribben, J. G.; Barrientos, J. C.; Rai, K. R.; Kay, N. E.; Kipps, T. J.; Shields, P.; Zhao, W.; Grever, M. R.; Plass, C.; Byrd, J. C. *Blood* **2014**, *124*, 42.
(52) Weber, M.; Davies, J. J.; Wittig, D.; Oakeley, E. J.; Haase, M.; Lam, W. L.; Schubeler, D. *Nat. Genet.* **2005**, *37*, 853.
(53) Lercher, L.; McDonough, M. A.; El-Sagheer, A. H.; Thalhammer, A.; Kriaucionis, S.; Brown, T.; Schofield, C. J. *Chem. Commun. (Cambridge, U. K.)* **2014**, *50*, 1794.
(54) Renciuk, D.; Blacque, O.; Vorlickova, M.; Spingler, B. *Nucleic Acids Res.* **2013**, *41*, 9891.
(55) La Francois, C. J.; Jang, Y. H.; Cagin, T.; Goddard, W. A., 3rd; Sowers, L. C. *Chem. Res. Toxicol.* **2000**, *13*, 462.
(56) Sumino, M.; Ohkubo, A.; Taguchi, H.; Seio, K.; Sekine, M. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 274.
(57) Lu, X.; Han, D.; Zhao, B. S.; Song, C. X.; Zhang, L. S.; Dore, L. C.; He, C. *Cell Res.* **2015**, *25*, 386.
(58) Kim, Y.; Kweon, J.; Kim, A.; Chon, J. K.; Yoo, J. Y.; Kim, H. J.; Kim, S.; Lee, C.; Jeong, E.; Chung, E.; Kim, D.; Lee, M. S.; Go, E. M.; Song, H. J.; Kim, H.; Cho, N.; Bang, D.; Kim, S.; Kim, J. S. *Nat. Biotechnol.* **2013**, *31*, 251.
(59) Meckler, J. F.; Bhakta, M. S.; Kim, M. S.; Ovadia, R.; Habrian, C. H.; Zykovich, A.; Yu, A.; Lockwood, S. H.; Morbitzer, R.; Elsaesser, J.; Lahaye, T.; Segal, D. J.; Baldwin, E. P. *Nucleic Acids Res.* **2013**, *41*, 4118.
(60) Guilinger, J. P.; Pattanayak, V.; Reyon, D.; Tsai, S. Q.; Sander, J. D.; Joung, J. K.; Liu, D. R. *Nat. Methods* **2014**, *11*, 429.

DOI: 10.1021/jacs.6b04807
J. Am. Chem. Soc. 2016, 138, 9910−9918