**UdgX-Mediated Uracil Sequencing at Single-Nucleotide Resolution**

Liudan Jiang,▲ Jiayong Yin,▲ Maoxiang Qian,▲ Shaoqin Rong, Shengqi Zhang, Kejing Chen, Chengchen Zhao, Yuanqing Tan, Jiaxin Guo, Hao Chen, Siyun Gao, Tingting Liu, Yi Liu, Bin Shen, Jian Yang, Yong Zhang, Fei-Long Meng, Jinchuan Hu,* Honghui Ma,* and Yi-Han Chen*

**ABSTRACT:** As an aberrant base in DNA, uracil is generated by either deoxyuridine (dU) misincorporation or cytosine deamination, and involved in multiple physiological and pathological processes. Genome-wide profiles of uracil are important for study of these processes. Current methods for whole-genome mapping of uracil all rely on uracil-DNA N-glycosylase (UNG) and are limited in resolution, specificity, and/or sensitivity. Here, we developed a UdgX cross-linking and polymerase stalling sequencing ("Ucaps-seq") method to detect dU at single-nucleotide resolution. First, the specificity of Ucaps-seq was confirmed on synthetic DNA. Then the effectiveness of the approach was verified on two genomes from different sources. Ucaps-seq not only identified the enrichment of dU at dT sites in pemetrexed-treated cancer cells with globally elevated uracil but also detected dU at dC sites within the "WRC" motif in activated B cells which have increased dU in specific regions. Finally, Ucaps-seq was utilized to detect dU introduced by the cytosine base editor (nCas9-APOBEC) and identified a novel off-target site in cellular context. In conclusion, Ucaps-seq is a powerful tool with many potential applications, especially in evaluation of base editing fidelity.

**INTRODUCTION**

Uracil is a pyrimidine base that possesses similar chemical structure to thymine and forms base-pairing with adenine as well. Since most DNA polymerases are unable to distinguish thymine and uracil, they occasionally incorporate dU instead of deoxythymidine (dT) into DNA, especially when the synthesis of thymine is disturbed. The other hand, uracil can also be generated through cytosine deamination catalyzed by the AID/APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) family proteins or by spontaneous hydrolysis, leading to C to T conversion. Uracil can be efficiently excised from deoxyribose by UNG, generating an apurinic site (AP site) which can be further repaired by the base excision repair (BER) pathway. Although the steady-state frequency of uracil in mammalian genomes is very low (∼10⁻⁷ per nucleotide), it plays critical roles in diverse biological processes. If uracil is continually incorporated into DNA, hyperactive BER may lead to DNA breaks and even cell death. This so-called "thymine-less cell death" has been exploited in several chemotherapeutic agents. In addition, two essential processes during B cell maturation, somatic hypermutation (SHM) and class switch recombination (CSR), are initiated by AID-catalyzed cytosine deamination in immunoglobulin genes. Furthermore, dysregulation of APOBEC can accumulate undesired C-to-T mutations in genome as well, which might facilitate the progression of specific cancer subtypes. Meanwhile, cytosine deamination-mediated C-to-T conversion has been applied in the genome-editing tool CBE (cytosine base editor) that holds great potential for the treatment of genetic diseases. Therefore, there is a growing demand for mapping uracil in the whole genome.

Recently, several UNG-based uracil-sequencing methods have been developed (reviewed in ref 16). They generally utilized UNG to convert uridines to AP sites, followed by chemical labeling or incision by an AP endonuclease to generate free 3' OH for further capture. Alternatively, an excision-defective UNG-deriv was applied to pull down uracil-containing DNA fragments. These methods may generate false positives caused by pre-existing DNA strand breaks, AP sites, or other modified nucleotides with aldehyde groups including 5-formyldeoxyuridine (5fU) and 5-formyl-deoxycytosine (5fC), although appropriate pretreatment might reduce some interferences. In addition, among these methods only Excision-seq™ and AI-seq™ can detect dU at base resolution. However, Excision-seq required uracils on both strands which are close to each other to generate proper double-stranded DNA fragments by UNG/Endo IV treatment.

**Received:** October 25, 2021

**Cite This:** https://doi.org/10.1021/jacs.1c11269

© XXXX The Authors. Published by American Chemical Society.
Thus, it is limited to genomes with a high density of uracil, e.g., repair-defective *S. cerevisiae* and *E. coli* (∼3–8 dUs per 10^3 nucleotides^23,24). Al-seq chemically converted UNG-generated AP sites to azide-cytosine and calculated the locations of these modified bases by comparing them to input sequences, making it sensitive to detect dT-derived dUs. In contrast, detecting uracil produced from cytosine deamination requires a substantial ratio of U:G mismatch at an initial C site (≥20%), which needs more than 20× coverage of sequencing depth for input samples. However, if considerable C to U (T) conversion can be identified in input data, there is no need to sequence dU in the scenario of finding off-targeting sites. Thus, we believe that a novel approach for directly labeling dU without producing reactive AP intermediates and DNA strand breaks would be superior compared to UNG-based methods.

A novel member of the UDG superfamily, named UdgX, was recently identified from *Mycobacterium smegmatis*. It not only specifically recognizes and excises uracil from DNA but also forms an irreversible covalent bond with the deoxyribose, which distinguishes it from other UDG family members. This UdgX-DNA complex is extremely stable even under harsh conditions such as SDS, NaOH, and heat,^25−27 which enables the detection of uracil by UdgX.

Here we describe a method termed “Ucaps-seq (UdgX crosslinking and polymerase stalling sequencing)” by combining the unique property of UdgX and the strategy of high-fidelity DNA polymerase stalling^28 to pinpoint uracil at single nucleotide resolution. After verifying the specificity of this method on synthetic DNA, Ucaps-seq was successfully applied to detect dUs derived from misincorporation and cytosine deamination. Moreover, it was utilized to trace the editing events of cytosine base editors and identified a novel off-target site in vivo, suggesting that Ucaps-seq is a potential tool for assessing base editor fidelity.

**RESULTS**

For efficient and specific capture of the UdgX-DNA complex, we plan to add a biotin-tag to the protein. To this end, UdgX was fused with N-terminal Avi-tag and coexpressed with BirA ligase which can add a biotin to the Avi-tag (Figure S1).^29,30 To examine the specificity of UdgX to uracil, the protein was incubated with synthetic single-stranded DNA (ssDNA) harboring a dU, dT, AP site or other dT-derivates such as 5-hydroxymethyluracil (5hmU) and 5fU. After reaction, the gel shift results showed that UdgX quantitatively reacted with uracil only (Figure 1a), in line with previous findings.^^25

---

**Figure 1.** The strategy of Ucaps-seq. (a) Reactivity of UdgX on ssDNA harboring various modifications. Reactions were stopped by alkali-heat-treatment which could also cleave AP sites (lanes 10 and 11). (b) Activity of UdgX on ssDNA and dsDNA substrates. (c) Blockage of primer extension by UdgX-dU cross-link. M53, M54, and M80 were markers indicating extension products stopping at one nucleotide before the uracil, at the uracil site, and bypassing the uracil, respectively. Comp, DNA-UdgX complex. Free, free DNA. (d) Workflow of Ucaps-seq.
Although there are arguments on the reactivity of UdgX with uracil in double-stranded DNA (dsDNA) or ssDNA, our results showed that UdgX could only efficiently cross-link with uracil on ssDNA but not those on dsDNA (Figure 1b), which is important for us to design the proper workflow for sequencing. In order to examine whether and where polymerase may stall in the DNA polymerization reaction, we performed a primer extension reaction on a dU containing ssDNA which was cross-linked with UdgX. As shown in Figure 1c, the polymerase stopped precisely at the nucleotide before the uracil site, suggesting that the exact position of uracil can be detected accurately by UdgX cross-linking and primer extension.

Encouraged by the above results, we developed Ucaps-seq for dU profiling in genomic DNA (Figure 1d). Genomic DNA was sheared, end-repaired, and ligated to Adaptor 1. Since UdgX only reacts with ssDNA, the ligated products were denatured and subsequently reacted with UdgX. We further optimized reaction conditions to achieve maximum labeling efficiencies of uracils in genomic DNA by incubation with UdgX (0.5 μg enzyme per 1 μg DNA) for 3 h on ice (Figure S2a, S2b). Then the mixture was purified with SPRI (Solid Phase Reversible Immobilization) beads and phenol extraction to remove excess free UdgX, which would also occupy the streptavidin beads and compete with the UdgX-DNA complex. Afterward, the DNA-UdgX complexes were captured by streptavidin beads. Then a primer was attached to the 3′ end of Adaptor 1 and extended by KAPA HiFi DNA polymerase. The extension products were released from the beads by alkaline treatment and ligated to Adaptor 2, followed by PCR-amplification to generate sequencing libraries. The purification of streptavidin beads and primer extension were highly efficient as shown in Figure S2c.

We first evaluated the Ucaps-seq workflow with synthetic DNA. Because UdgX could only react with uracil in ssDNA, the surrounding sequences of dU or substitute dT in corresponding synthesized oligonucleotides (UF or TF) were comprised of random bases (14 bases on each side) (Figure 2a, Table S1) to prevent reannealing during reaction. Equal amounts of unmodified and dU-containing dsDNA were mixed and subjected to input and Ucaps-seq libraries (Figure 2a). Sequencing results showed that, compared with input, the dU-containing strands (UF) were highly enriched in Ucaps-seq (98.74%, Figure 2b). More importantly, the polymerase mainly stopped right before the known uracil site (∼85.2%) (Figure 2c). As for the remaining 14.8%, there might be two reasons. First, our input data indicated that the quality of long synthetic oligonucleotide with random sequence and modification is somehow not satisfactory. It was reported that single-base deletions and depurination are primary synthesis-related errors especially for long polynucleotides, which could lead to early termination of extension. Second, there might be occasional resection at the 3′ ends of extension strands by the strong proofreading activity of KAPA HiFi DNA polymerase. In conclusion, these results suggested that Ucaps-seq was a specific and precise method for uracil sequencing.

We next applied Ucaps-seq to map uracil in mammalian genome. Given the very low abundance of dU in unchallenged cells, we decided to treat cells with pemetrexed (PMX) and use untreated cells as a control. PMX is an antifolate chemotherapy agent which inhibits the thymidylate synthase, leading to accumulation of cellular dUMP and misincorporation of uracil into DNA (Figure S3a). In order to prevent incorporated uracil from repairing, we knocked out UNG in HeLa cells and confirmed it by Sanger sequencing, Western blot, and PMX sensitivity assays (Figure S3b, S3c, S3d). The elevated uracil content in PMX-treated UNG−/− HeLa cells (∼1 per 1.7 × 10^4 nucleotides) was verified by a modified UdgX-based dot-blot method (Figure S3e, S3f). Quality check of Ucaps-seq libraries revealed a higher yield for the PMX-treated sample (Figure...
S3g), consistent with higher uracil content in this sample. As described in Figure 1d, the DNA polymerase stopped right before the dU site; thus, the uracil should be at the nucleotide adjacent to the 5′ end of the reads, which ought to be thymine in the reference genome for PMX-treated cells. Thymine was indeed highly enriched at expected positions in PMX-treated cells (Figures 3a, S3h). In contrast, as the endogenous dU in UNG−/− cells is very low (∼1 per 3 × 10⁶ nucleotides as reported2) and can be generated from both misincorporation (at dT sites) and deamination (at dC sites), only a slight enrichment of cytosine but not thymine was observed in untreated cells (Figures 3a, S3h). Then we analyzed the flanking nucleobases of detected modification sites in PMX-treated cells, and identified a preference for T and C at the 5′ upstream site of dU (dT) (Figures 3b, S3i, left). This pattern was not observed around other nonspecific bases (A, G, and C) (Figures 3b, S3i, right), excluding the possibility of artifacts introduced during the library construction. Furthermore, the neighboring bases of dT were random in the reference genome (Figure S3j), suggesting that the preference for surrounding bases of dU was due to selective incorporation rather than inherent sequence context.

As uracil was incorporated during DNA replication upon PMX-treatment, we profiled the uracil distribution in different replicating timing domains and found that uracil content was lowest in early replicating domains (ERDs) (Figure 3c, Figure S3k). To investigate whether it was due to the variance of local sequence contexts, we calculated the prevalence of nucleotides surrounding dTs in the reference genome in different replicating domains. As shown in Figure S3l, the patterns of neighbor nucleotides were similar among replication domains, excluding the effect of sequence context. One possible explanation was that ERDs were strongly correlated with open chromatin35 and more accessible for other uracil glycosylases (e.g., SMUG136) in UNG knockout cells, and thus were repaired faster and had less uracil.

Then, Ucaps-seq was used to measure uracil originated from cytosine deamination. Uracil generated by AID can initiate CSR which determines the diversity of antibody. AID preferentially deaminates cytidines within the 5′ Sμ region...
and a downstream acceptor S region at the WRC (W = A, T; R = A, G) motif, and mostly the AGCT motif. Ucaps-seq uncovered that uracil was preferentially located at WRC, especially the AGCT motif within the strongest AID-hotspot region in CSR-activated UNG−/− B lymphocytes but not AID-deficient cells (Figure S4), suggesting that our method could be applied to detect region-specific dU as well.

Uracil is a crucial intermediate in cytosine base editors (CBEs) mediated C-to-T transition (Figure 4a), so Ucaps-seq can be used to search off-target sites by capturing the dU intermediate. In order to increase the incidence of off-target editing, two well-characterized gRNAs, i.e., "HEK_293_site3" (site3) and "RNF2", were simultaneously transfected into HEK293T cells with AncBE4max. Editing efficiencies at targeted loci were confirmed by Sanger Sequencing (Figure 4b, 4c), and the occupations of dU within editing windows were proved by a modified ligation-mediated PCR assay (Figure S5a, S5b). As expected, screenshots of Ucaps-seq at the on-target site of RNF2. Edited DNA were diluted with unedited DNA to evaluate the sensitivity of Ucaps-seq (e) Screenshot of Ucaps-seq at the on-target site of site3. (d, e) Edited windows were indicated by pink shadows.
target sites which have much lower efficiencies than on-target sites. For this purpose, we developed a bioinformatic pipeline (Figure S5d, see Methods for details) and identified six candidate sites which possessed approximately 0.2–3.8 unique dU reads per million filtered reads (Figure 5a). The top five sites were targeted loci of RNF2 and site3, and the sixth site (OT1) showed high sequence similarity to the RNF2 locus (Figure 5b). The OT1 site was unexpected since RNF2 was reported to have no off-target in vivo.21,38,39 Targeted amplicon sequencing verified that OT1 exhibited an approximately 2.6% editing ratio (Figure 5c). To validate that OT1 was a bona fide off-target site of RNF2, the site3 or RNF2 gRNA was separately coexpressed with AncBE4max in HEK293T cells. Targeted amplicon sequencing confirmed that OT1 was edited only in cells treated with RNF2 gRNA (Figure S5e).

■ DISCUSSION

We developed Ucaps-seq in this work by employing a novel protein called “UdgX” to map uracil genome-wide at single nucleotide resolution. Ucaps-seq provides several advantages over prior UNG-based dU-sequencing methods. Unlike UNG, which can excise both dU and 5fU to produce an AP site,2 UdgX only reacts with dU to form a covalent complex, avoiding interference from pre-existing DNA damages such as AP sites and single strand breaks. Furthermore, the blockage of DNA polymerase not only assures single-nucleotide resolution but also performs an extra dU verification step. If there was no UdgX-dU cross-linking, the polymerase could extend to the 5′ end of the templates and generate a full-length product containing the complementary sequence of 5′ end of the first adaptor, which can be picked out during the following steps (Figure 1d and Methods). When compared with previous UNG-based single-nucleotide dU sequencing methods (i.e., excision-seq and AI-seq), our Ucaps-seq exhibited much higher sensitivity for all dUs than excision-seq and for dC-derived dUs than AI-seq, allowing Ucaps-seq to be used to assess cytosine base editing.

The specificity and accuracy of Ucaps-seq were first validated by synthetic DNA. Besides, the enrichment of dT at the expected site in the PMX-treated cells provides another way for assessment of the specificity of Ucaps-seq. In the human genome, the ratio of T is approximately 30%, and it means that the T/(A+G+C) ratio is about 30%/ (1–30%) = 0.43. Since the distribution of nonspecific reads is random, the nonspecific ratio of T could be calculated from the ratio of A +G+C. Then the ratio of specific T (dT) in all Ts could be obtained. As shown in Figure 3a, the T ratio of PMX-treated cells is 75% and the sum of A+G+C is 25%, so the nonspecific ratio of T is 25% × 0.43 = 10.75%. Therefore, the specific Ts (Us) should occupy about (1 − 10.75%/75%) × 100% ≈ 86% of all Ts at predicted sites, indicating that Ucaps-seq is able to detect the precise positions of dUs across the genome with a frequency of ~1 per 1.7 × 104 nucleotides.

The off-target activities of genome editing tools have long been a concern, and many methods21,38–40 have been developed to characterize them, the majority of which capture DNA double strand breaks (DSB) induced by Cas9 nucleases. However, CBEs can edit the genome without generating DSBs. During the preparation of our manuscript, Lei et al. reported a method named Detect-seq to evaluate the specificity of CBE through dU sequencing.23 However, Detect-seq still relied on UNG and did not achieve base resolution. More importantly, it found no off-target sites in HEK293T cells transfected with BE4max and RNF2 gRNA. When we employed the AncBE4max variant with the same gRNA, an off-target site (OT1) was detected by Ucaps-seq and verified by targeted amplicon sequencing. Furthermore, our findings revealed that CBE and classical Cas9 nuclease had different off-target effects, implying that it might be an alternative technique to evaluate CBE fidelity.
It should be noted that several on-target dUs (e.g., C3 at RNF2 locus and C5 at site3 locus) were poorly detected by Ucaps-seq. According to the procedure of Ucaps-seq, when there were two dUs within the editing window (but not next to each other), the polymerase would be blocked by the downstream dU (Figure S5c, left). This may explain the decreased Ucaps-seq signal of C3 at RNF2 locus, despite the fact that it was edited efficiently. Moreover, although C4 and C5 at site3 locus were both edited, Ucaps-seq only identified C4 but not the downstream C5. One probable reason is that when two neighboring dUs were present, their reactions with UdgX would interfere with each other, especially the reaction at 3’ dU which would be inhibited (Figure S5c, right). Nevertheless, the sensitivity for detecting off-target events would not be affected for two reasons. First, dUs at off-target sites should only have minor effects on identification of surrounding edited sites due to the low occurrences. Moreover, even if there were multiple edited Cs in the same editing window, Ucaps-seq could identify at least one dU, and then targeted amplicon sequencing (see methods) was performed to detect surrounding editing events.

In summary, Ucaps-seq, which takes advantage of UdgX protein and the high-fidelity DNA polymerase, is a valuable technique for exploring uracil-related topics such as the hotspots of dysregulated APOBECs and the specificity of CBEs. Furthermore, since many other proteins can covalently cross-link with specific base modifications (e.g., HMCS and AP sites\(^ {17}\)), Ucaps-seq also provides an innovative strategy for accurate mapping of these modifications.

**ASSOCIATED CONTENT**

**Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.1c11269.

Experimental details, Supplementary Figures 1–5, and Supplementary Table 1 (PDF)

**Accession Codes**

All raw sequencing data are available at NCBI Sequence Read Archive with BioProject ID PRJNA728500. The source codes used in this paper are available at https://github.com/Jyyin333/Ucaps-seq.

**AUTHOR INFORMATION**

**Corresponding Authors**

Jinchuan Hu — Shanghai Fifth People’s Hospital, Fudan University, and Shanghai Key Laboratory of Medical Epigenetics, International Co-laboratory of Medical Epigenetics and Metabolism (Ministry of Science and Technology), Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China; Email: hujinchuan@fudan.edu.cn

Honghui Ma — Department of Cardiology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China; Key Laboratory of Arrhythmias of the Ministry of Education of China, Tongji University School of Medicine, Shanghai 200092, China; Research Units of Origin and Regulation of Heart Rhythm, Chinese Academy of Medical Sciences, Shanghai 200092, China; Email: honghui2ma@tongji.edu.cn

Yi-Han Chen — Department of Cardiology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China; Key Laboratory of Arrhythmias of the Ministry of Education of China, Tongji University School of Medicine, Shanghai 200092, China; Email: yihanchen@tongji.edu.cn

**Authors**

Luidan Jiang — Department of Cardiology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China; Shanghai Fifth People’s Hospital, Fudan University, and Shanghai Key Laboratory of Medical Epigenetics, International Co-laboratory of Medical Epigenetics and Metabolism (Ministry of Science and Technology), Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China; Key Laboratory of Arrhythmias of the Ministry of Education of China, Tongji University School of Medicine, Shanghai 200092, China

Jiayong Yin — Institute of Pediatrics and Department of Hematology and Oncology, Children’s Hospital of Fudan University, and Shanghai Key Laboratory of Medical Epigenetics, International Co-laboratory of Medical Epigenetics and Metabolism (Ministry of Science and Technology), Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China

Moxiang Qian — Institute of Pediatrics and Department of Hematology and Oncology, Children’s Hospital of Fudan University, and Shanghai Key Laboratory of Medical Epigenetics, International Co-laboratory of Medical Epigenetics and Metabolism (Ministry of Science and Technology), Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China

Shaoqin Rong — Shanghai Key Laboratory of Medical Epigenetics, International Co-laboratory of Medical Epigenetics and Metabolism (Ministry of Science and Technology), Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China

Shengqi Zhang — Department of Cardiology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China; Key Laboratory of Arrhythmias of the Ministry of Education of China, Tongji University School of Medicine, Shanghai 200092, China

Kejing Chen — Department of Cardiology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China; Key Laboratory of Arrhythmias of the Ministry of Education of China, Tongji University School of Medicine, Shanghai 200092, China

Chengchen Zhao — School of Life Science and Technology, Tongji University, Shanghai 200092, China

Yuanning Tan — Shanghai Key Laboratory of Medical Epigenetics, International Co-laboratory of Medical Epigenetics and Metabolism (Ministry of Science and Technology), Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China

Jiayin Guo — State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing 211166, China

Hao Chen — School of Medicine, Southern University of Science and Technology, Shenzhen 518055, China

Siyun Gao — Department of Cardiology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China; Key Laboratory of Arrhythmias of the Ministry of Education of China, Tongji University School of Medicine, Shanghai 200092, China

Tingting Liu — State Key Laboratory of Molecular Biology, Shanghai Institute of Biochemistry and Cell Biology, Shanghai 200031, China

Ministry of Education of China, Tongji University School of Medicine, Shanghai 200092, China; orcid.org/0000-0003-1000-1835; Email: yihanchen@tongji.edu.cn
Yi Liu — Department of Cardiology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China; Key Laboratory of Arrhythmias of the Ministry of Education of China, Tongji University School of Medicine, Shanghai 200092, China

Bin Shen — State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing 211166, China

Jian Yang — Department of Cardiology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China; Key Laboratory of Arrhythmias of the Ministry of Education of China, Tongji University School of Medicine, Shanghai 200092, China

Yong Zhang — School of Life Science and Technology, Tongji University, Shanghai 200092, China

Fei-Long Meng — State Key Laboratory of Molecular Biology, Shanghai Institute of Biochemistry and Cell Biology, Shanghai 200031, China

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.1c11269

Author Contributions

L.J., J.Y., and M.Q. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (82088101, 81930013, and 81770397 to Y.-H.C.; 31971330 and 32170586 to H.M.; 21807013 and 81973997 to M.Q.), the National Key Research and Development Plan (2019YFA0801501, to Y.-H.C.), Key Disciplines Group Construction Project of Pudong Health Bureau of Shanghai (PWZxx2017-05), Top-level Clinical Discipline Project of Shanghai Pudong District (PWYg2018-02), Shanghai Key clinical specialty Project (shslczdzk06202), Research Unit of Origin and Regulation of Heart Rhythm, Chinese Academy of Medical Sciences (2019RU045), Innovative research team of high-level local universities in Shanghai, and a key laboratory program of the Education Commission of Shanghai Municipality (ZDYSYS14005). J.H. and M.Q. are supported by the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning [2019] and the innovative research team of high-level local university in Shanghai. J.H. is supported by the Shanghai Outstanding Young Talent Program. Y.-H.C. is a fellow at the Collaborative Innovation Center for Cardiovascular Disease Translational Medicine, Nanjing Medical University. The authors would like to acknowledge the technical support from Hua Li and Lin Lin at SUSTech CRFT.

REFERENCES

(1) Mosbaugh, D. W. Purification and characterization of porcine liver DNA polymerase gamma: utilization of dUTP and dTTP during in vitro DNA synthesis. Nucleic Acids Res. 1988, 16 (12), 5645–59.

(2) Krokan, H. E.; Drabløs, F.; Slupphaug, G. Uracil in DNA — occurrence, consequences and repair. Oncogene 2002, 21 (58), 8935–8948.

(3) Wardle, J.; Burgers, P. M.; Cann, I. K.; Darley, K.; Heslop, P.; Johansson, E.; Lin, L. J.; McGlynn, P.; Sanvoisin, J.; Stith, C. M.; Connolly, B. A. Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya. Nucleic Acids Res. 2008, 36 (3), 705–11.

(4) Olinski, R.; Jurkowski, M.; Zaremba, T. Uracil in DNA—its biological significance. Mutat. Res. 2010, 705 (3), 239–45.

(5) Wallace, S. S. Base excision repair: a critical player in many games. DNA Repair (Amst) 2014, 19, 14–26.

(6) Krokan, H. E.; Bjørås, M. Base excision repair. Cold Spring Harb Perspect Biol. 2013, 5 (4), a012583.

(7) Galashevskaya, A.; Sarno, A.; Vagbo, C. B.; Aas, P. A.; Hagen, L.; Slupphaug, G.; Krokan, H. E. A robust, sensitive assay for genomic uracil determination by LC/MS/MS reveals lower levels than previously reported. DNA Repair (Amst) 2013, 12 (9), 699–706.

(8) Goulia, M.; Bleile, B. M.; Dickey, L. M.; Graffstrom, R. H.; Ingraham, H. A.; Neynaber, S. A.; Peterson, M. S.; Tseng, B. Y. Mechanism of thymineless death. Adv. Exp. Med. Biol. 1986, 195B, 89–95.

(9) Li, L.; Connor, E. E.; Berger, S. H.; Wyatt, M. D. Determination of apoptosis, uracil incorporation, DNA strand breaks, and sister chromatid exchanges under conditions of thymidylate deprivation in a model of BER deficiency. Biochem. Pharmacol. 2005, 70 (10), 1458–1468.

(10) Petersen-Mahrt, S. K.; Harris, R. S.; Neuberger, M. S. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature 2002, 418 (6893), 99–104.

(11) Di Noia, J. M.; Neuberger, M. S. Molecular Mechanisms of Antibody Somatic Hypermutation. Annu. Rev. Biochem. 2007, 76 (1), 1–22.

(12) Stavnezer, J.; Guikema, J. E.; Schrader, C. E. Mechanism and regulation of class switch recombination. Annu. Rev. Immunol. 2008, 26, 261–92.

(13) Tubbs, A.; Nussenzeig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. Cell 2017, 168 (4), 644–656.

(14) Komor, A. C.; Kim, Y. B.; Packer, M. S.; Zuris, J. A.; Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 2016, 533 (7603), 420–4.

(15) Rees, H. A.; Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. Nat. Rev. Genet. 2018, 19 (12), 770–788.

(16) Mingard, C.; Wu, J.; Meekague, M.; Sturla, S. J. Next-generation DNA damage sequencing. Chem. Soc. Rev. 2020, 49 (20), 7354–7377.

(17) Sakhtemani, R.; Senevirathne, V.; Stewart, J.; Perera, M. L. W.; Pique-Regi, R.; Lawrence, M. S.; Bhagwat, A. S. Genome-wide mapping of regions preferentially targeted by the human DNA-cytosine deaminase APOBEC3A using uracil-DNA pulldown and sequencing. J. Biol. Chem. 2019, 294 (41), 15037–15051.

(18) Wang, Y.; Zhang, X.; Han, S.; Yang, W.; Chen, Z.; Wu, F.; Liu, J.; Weng, X.; Zhou, X. Base-Resolution Analysis of Deoxyuridine at the Genome Scale Based on the Artificial Incorporation Modified Nucleoside. ACS Central Science 2021, 7 (6), 973–979.

(19) Bryan, D. S.; Ransom, M.; Adane, B.; York, K.; Hesselberth, J. R. High resolution mapping of modified DNA nucleobases using excision repair enzymes. Genome Res. 2014, 24 (9), 1534–42.

(20) Shu, X.; Liu, M.; Lu, Z.; Zhu, C.; Meng, H.; Huang, S.; Zhang, X.; Yi, C. Genome-wide mapping reveals that deoxyuridine is enriched in the human centromeric DNA. Nat. Chem. Biol. 2018, 14 (7), 680–687.

(21) Lei, Z.; Meng, H.; Lv, Z.; Liu, M.; Zhao, H.; Wu, H.; Zhang, X.; Liu, L.; Zhuang, Y.; Yin, K.; Yan, Y.; Yi, C. Detect-seq reveals out-of-protospecies editing and target-strand editing by cytosine base editors. Nat. Methods 2021, 18 (6), 643–651.

(22) Palinkas, H. L.; Bekesi, A.; Rona, G.; Pongor, L.; Papp, G.; Tihanyi, G.; Holub, E.; Poti, A.; Gemma, C.; Ali, S.; Morten, M. J.; Rothenberg, E.; Pagano, M.; Sutz, D.; Gyorffy, B.; Vertessy, B. G. Genome-wide alterations of uracil distribution patterns in human DNA upon chemotherapeutic treatments. Elife 2020, 9, No. e60498.

(23) Horvath, A.; Vertessy, B. G. A one-step method for quantitative determination of uracil in DNA by real-time PCR. Nucleic Acids Res. 2010, 38 (21), e196.
(24) Rona, G.; Scheer, I.; Nagy, K.; Palinkas, H. L.; Tibanyi, G.; Borsos, M.; Bekesi, A.; Vertessy, B. G. Detection of uracil within DNA using a sensitive labeling method for in vitro and cellular applications. Nucleic Acids Res. 2016, 44 (3), e28.

(25) Sang, P. B.; Srinath, T.; Patil, A. G.; Woo, E. J.; Varshney, U. A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily. Nucleic Acids Res. 2015, 43 (17), 8452–63.

(26) Ahn, W. C.; Aroli, S.; Kim, J. H.; Moon, J. H.; Lee, G. S.; Lee, M. H.; Sang, P. B.; Oh, B. H.; Varshney, U.; Woo, E. J. Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision. Nat. Chem. Biol. 2019, 15 (6), 607–614.

(27) Yu, J.; Chen, R.; Yang, Y.; Cao, W.; Xie, W. Suicide inactivation of the uracil DNA glycosylase UdgX by covalent complex formation. Nat. Chem. Biol. 2019, 15 (6), 615–622.

(28) Hu, J.; Adebali, O.; Adar, S.; Sancar, A. Dynamic maps of UV damage formation and repair for the human genome. Proc. Natl. Acad. Sci. U. S. A. 2017, 114 (26), 6758–6763.

(29) Li, Y.; Sousa, R. Expression and purification of E. coli BirA biotin ligase for in vitro biotinylation. Protein Expr Purif 2012, 82 (1), 162–7.

(30) Li, Y.; Sousa, R. Novel system for in vivo biotinylation and its application to crab antimicrobial protein scygonadin. Biotechnol. Lett. 2012, 34 (9), 1629–35.

(31) Datta, M.; Aroli, S.; Karmakar, K.; Dutta, S.; Chakravortty, D.; Varshney, U. Development of mCherry tagged UdgX as a highly sensitive molecular probe for specific detection of uracils in DNA. Biochem. Biophys. Res. Commun. 2019, 518 (1), 38–43.

(32) Stewart, J. A.; Schauer, G.; Bhagwat, A. S. Visualization of uracils created by APOBEC3A using UdgX shows colocalization with RPA at stalled replication forks. Nucleic Acids Res. 2020, 48 (20), e118.

(33) Hughes, R. A.; Ellington, A. D. Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. Cold Spring Harb Perspect Biol. 2017, 9 (1), a023812.

(34) Bulgar, A. D.; Weeks, L. D.; Miao, Y.; Yang, S.; Xu, Y.; Guo, C.; Markowitz, S.; Oleinick, N.; Gerson, S. L.; Liu, L. Removal of uracil by uracil DNA glycosylase limits pemetrexed cytotoxicity: overriding the limit with methoxyamine to inhibit base excision repair. Cell Death Dis 2012, 3, e352.

(35) Marchal, C.; Sima, J.; Gilbert, D. M. Control of DNA replication timing in the 3D genome. Nat. Rev. Mol. Cell Biol. 2019, 20 (12), 721–737.

(36) Olimski, R.; Slupphaug, G.; Foksinski, M.; Krokan, H. E. Genomic Uracil and Aberrant Profile of Demethylation Intermediates in Epigenetics and Hematologic Malignancies. Int. J. Mol. Sci. 2021, 22 (8), 4212.

(37) Honjo, T.; Kinoshita, K.; Muramatsu, M. Molecular mechanism of class switch recombination: linkage with somatic hypermutation. Annu. Rev. Immunol. 2002, 20, 165–96.

(38) Tsai, S. Q.; Zheng, Z.; Nguyen, N. T.; Liebers, M.; Topkar, V. V.; Thapar, V.; Wyvackes, N.; Khayer, C.; Ialstrate, A. J.; Le, L. P.; Ayyee, M. J.; Joung, J. K. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. 2015, 33 (2), 187–197.

(39) Wienert, B.; Wyman, S. K.; Richardson, C. D.; Yeh, C. D.; Ackakaya, P.; Porritt, M. J.; Morlock, M.; Vu, J. T.; Kazane, K. R.; Watry, H. L.; Judge, L. M.; Conklin, B. R.; Mareca, M.; Corn, J. E. Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. Science 2019, 364 (6443), 286–289.

(40) Yan, W. X.; Mirzazadeh, R.; Garnerone, S.; Scott, D.; Schneider, M. W.; Kallas, T.; Custodio, J.; Wernersson, E.; Li, Y.; Gao, L.; Fedorova, Y.; Zetsche, B.; Zhang, F.; Bienko, M.; Crosetto, N. BISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. Nat. Commun. 2017, 8, 15058.

(41) Mohini, K. N.; Wessel, S. R.; Zhao, B.; Wojciechowski, A. C.; Luzwick, J. W.; Layden, H.; Eichman, B. F.; Thompson, P. S.; Mehta, K. P. M.; Cortez, D. HMCS Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. Cell 2019, 176 (1–2), 144–153 e113.