



Review

Detection of Genomic Uracil Patterns

Angéla Békési^{1,2,*}, Eszter Holub^{1,2}, Hajnalka Laura Pálinkás^{1,2} and Beáta G. Vértessy^{1,2,*}

¹ Department of Applied Biotechnology & Food Sciences, Budapest University of Technology and Economics, H-1111 Budapest, Hungary; holub.eszter@gmail.com (E.H.); palinkas.hajnalka@ttk.hu (H.L.P.)

² Genome Metabolism Research Group, Institute of Enzymology, Research Centre for Natural Sciences, Eötvös Lóránd Research Network, H-1117 Budapest, Hungary

* Correspondence: bekesi.angela@ttk.hu (A.B.); vertessy.beata@ttk.hu (B.G.V.)

Abstract: The appearance of uracil in the deoxyuridine moiety of DNA is among the most frequently occurring genomic modifications. Three different routes can result in genomic uracil, two of which do not require specific enzymes: spontaneous cytosine deamination due to the inherent chemical reactivity of living cells, and thymine-replacing incorporation upon nucleotide pool imbalances. There is also an enzymatic pathway of cytosine deamination with multiple DNA (cytosine) deaminases involved in this process. In order to describe potential roles of genomic uracil, it is of key importance to utilize efficient uracil-DNA detection methods. In this review, we provide a comprehensive and critical assessment of currently available uracil detection methods with special focus on genome-wide mapping solutions. Recent developments in PCR-based and in situ detection as well as the quantitation of genomic uracil are also discussed.

Keywords: uracil-DNA; dot blot; in situ detection; PCR-based U-DNA detection; genome-wide uracil mapping



Citation: Békési, A.; Holub, E.; Pálinkás, H.L.; Vértessy, B.G. Detection of Genomic Uracil Patterns. *Int. J. Mol. Sci.* **2021**, *22*, 3902. <https://doi.org/10.3390/ijms22083902>

Academic Editor: Miroslav Chovanec

Received: 20 February 2021

Accepted: 5 April 2021

Published: 9 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The thymine analogue uracil base has long been considered as an error to be excluded from the DNA; however, key experimental evidence has emerged supporting a more fine-tuned view on the potential impact of genomic uracil (for a detailed review, see [1]). The uracil base (U) can appear in DNA from three sources (Figure 1). First, spontaneous hydrolytic cytosine deamination is one of the most frequent DNA base damage—in a mammalian size genome, several hundreds of such events occur daily, to be compared to the rate of spontaneous depurination (2000–10,000/day/genome) [2,3]. Secondly, DNA cytosine deaminases from the AID/APOBEC family can also catalyze the conversion of cytosine (C) to U in the genome [4–11]. Thirdly, in addition to these potentially mutagenic cytosine deamination events, replicative polymerases can efficiently incorporate 2'-deoxyuridine 5'-monophosphate (deoxyuridylylate, dUMP) moieties instead of the naturally occurring 2'-deoxythymidine 5'-monophosphate (deoxythymidylylate, dTMP), when 2'-deoxyuridine 5'-triphosphate (dUTP) precursor is abundant, e.g., upon perturbed nucleotide metabolism [12–15].

Cellular processing and physiological consequences of the genomic uracils are highly context-dependent (Figure 1). Cytosine deaminations result in uracil:guanine (U:G) mismatches that can be fixed as point mutations upon the next replication in one of the daughter cells unless repaired by base excision repair (BER) or mismatch repair (MMR) mechanisms. Thymine-replacing incorporation is less harmful because the DNA coding potential is not changed, however, the missing methyl label from the uracil base as compared to thymine (T) might result in the perturbation of certain DNA binding regulatory functions [16]. Recently, genomic uracil was suggested as a potential epigenetic marker [1,17].

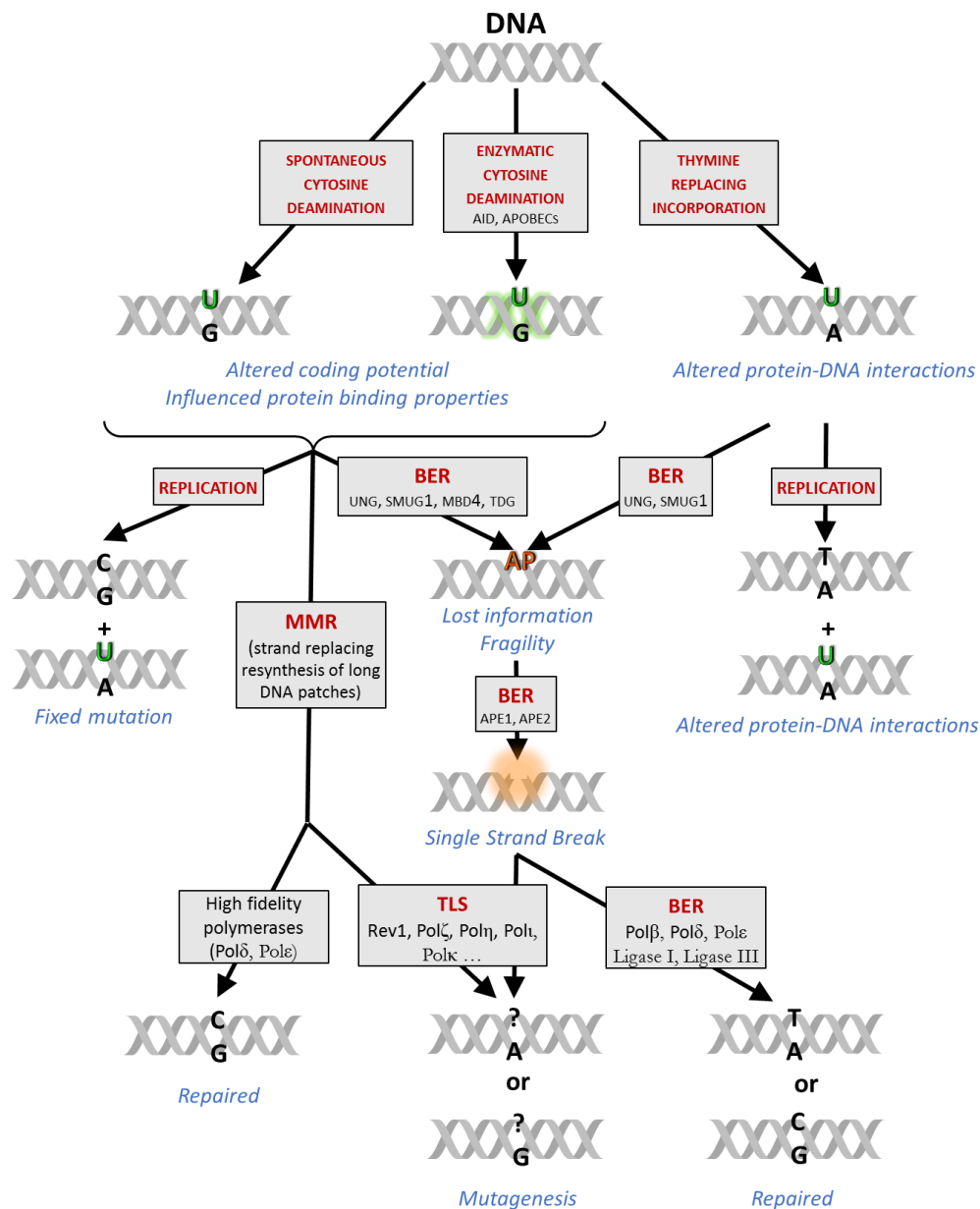


Figure 1. Sources and effects of genomic uracil involving distinct repair mechanisms. Different processes (red letters in grey boxes) are shown with the most important enzymes involved (black letters in grey boxes). Possible consequences of the different outputs are also indicated (blue). Uracil in DNA can arise by spontaneous or enzymatic deamination of cytosines, and by thymine-replacing incorporation. Without repair, U:G mismatches due to cytosine deamination events lead to point mutations right after the next replication, because cytosine and uracil (or thymine) have different base pairing abilities hence different coding potential. Cytosine deamination or the resulted point mutations as well as the “innocent” thymine-replacing uracils (resulted in uracil:adenine (U:A) pairs) can potentially influence interactions between DNA and different DNA binding proteins (such as transcription factors, or polymerases). Uracils can be repaired via base excision repair (BER) initiated by the uracil-DNA-glycosylases: the major enzyme is the uracil-DNA N-glycosylase (UNG); a potential backup is the single-strand selective monofunctional uracil-DNA-glycosylase (SMUG1), while the thymine–DNA–glycosylase (TDG) and the methyl-CpG-binding domain protein 4 (MBD4) can only excise uracils from U:G mismatches. The resulting apyrimidinic/apurinic sites (AP) can be further processed by the AP endonucleases (APE1, APE2). The resulted gap (orange transparent cloud at the single-stranded DNA (ssDNA) break) can be filled correctly by the classic BER polymerases (pol β , δ , and ϵ) or in error-prone manner by translesion synthesis (TLS) polymerases (Rev1, pol ζ , η , ι , κ , ...). The U:G mismatches can be repaired by the mismatch repair (MMR) as well, which involves re-synthesis of long DNA strands usually by high fidelity but sometimes by error-prone polymerases. Both repair processes could result in either mutagenesis or correct repair depending on the recruited polymerases. In a cell, these different steps and ways of repair might occur simultaneously, and their contribution is also influenced by the actual cell cycle phase, the amount of genomic uracils, or the availability of all of these repair factors.

Both the overall amount and the genome-wide distribution of uracils contribute to cellular effects in a strongly context-dependent manner, influenced by numerous factors (Figure 1). For example, mutagenesis due to an increased number of unrepaired deaminated cytosines could be harmful for cell viability [18], but could also potentially lead to malignant transformation. However, limited and strictly regulated C to U deamination events by the activation-induced cytidine deaminase (AID) targeted to the immunoglobulin (Ig) encoding genomic loci [19] are indispensable to trigger essential processes in adaptive immunity such as somatic hypermutation (SHM) and class switch recombination (CSR) [20,21]. Other apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like (APOBEC) enzymes are involved in antiviral defense of the innate immune system [10,22], but overexpression and/or mistargeting of APOBEC deaminases may also lead tumorigenesis, tumor progression, and drug resistance [10,23–26].

Another particular example for uracil appearance in DNA occurs upon drug treatments by targeting thymidylate biosynthesis, where the massive thymine-replacing uracil incorporation and/or its consequently activated repair by futile cycles eventually lead to cell cycle arrest, cellular senescence, or apoptosis [27–29]. The thymineless cell death pathway has been investigated for many years and has been exploited in widely used anti-cancer therapeutic strategies; however, the molecular mechanism and the exact causative factors have not yet been clearly elucidated.

The main repair pathway for uracil (BER) is initiated by uracil-DNA glycosylases (UDGs) that hydrolyze the glycosylic bond between the base and the deoxyribose providing apyrimidinic/apurinic sites (AP sites) for AP endonucleases (APE1 and APE2) [30,31]. The task to recognize and cleave the uracil base from DNA is shared by four different UDGs in mammalian cells depending on the actual cellular and molecular context. The major enzyme termed uracil-DNA-N-glycosylase (UNG) can efficiently process both U:A and U:G as well as U in ssDNA [32,33]. It has both nuclear (UNG2) and mitochondrial (UNG1) isoforms; the latter provides the sole UDG activity controlling uracil content of the independently replicated mitochondrial DNA [34]. UNG can also recognize further uracil-derived non-canonical bases, e.g., 5-fluoro-uracil, hydroxyuracil, isodialuric acid, and alloxan [33]. The single-strand selective monofunctional uracil-DNA-glycosylase (SMUG1) prefers ssDNA in vitro [35], however, it can also constitute an efficient backup for UNG within the cellular context [33]. SMUG1 recognizes and processes formyluracil, hydroxyuracil, and hydroxymethyl-uracil (hmU) bases, as well [36,37].

In contrast to UNG and SMUG1, the uracil excision activities of the other two UDGs are confined to U:G mismatches. Thymine-DNA glycosylase (TDG) can efficiently remove not only U but also T and various 5-substituted uracil derivatives once they are mispaired with G [38–40]. TDG is also implicated in epigenetic and developmental processes via its contribution to active DNA demethylation [41–43]. The methyl-CpG-binding domain protein 4 (MBD4) is also specialized for thymine analogues in mispaired context, and its activity has been implied in epigenetic remodeling processes at CpG islands [38,44]. TDG and MBD4 have a key role in the interplay between BER and MMR to repair U:G mismatches [45].

AP sites represent a potentially mutagenic information-lost-state (e.g., when processed by translesion synthesis (TLS) polymerases). If present at a high level, AP sites might influence essential cellular processes such as replication and transcription. The BER pathway is normally completed by resynthesis of the single missing nucleotide, or a longer piece of DNA replacing the original DNA strand downstream, followed by ligation of the single-strand break and also removing the 5'-deoxyribose phosphate (5'dRP) or the longer flanking DNA track [46,47]. The repair synthesis is mostly mediated by polymerases, such as polymerase β in short patch BER, and either one of polymerases β , δ or ϵ in long-patch BER. However, during SHM, error-prone polymerases (e.g., REV1 and pol η) are also implicated in the concerted action of BER and MMR [45,48,49]. In general, the choice of the polymerase depends on the actual availability and post-translational modifications of different polymerases that is also regulated at multiple level (e.g., by cell cycle, or DNA-

damage response (DDR)) [50]. The numerous different steps and routes of repair occur simultaneously within the cell and need to be correctly considered in experimental design and interpretation of results.

The biosynthetic pathways of pyrimidine nucleotides have evolved to keep an extremely low cellular dUTP/dTTP ratio to minimize the frequency of thymine-replacing uracil incorporation (Figure 2) [1]. Within this network, some enzymes show strict substrate specificity, such as dCMP deaminase (DCTD) [51,52], deoxyuridine triphosphatase (dUTPase) [53], and thymidylate synthase (TS) [54], which are dedicated to efficiently catalyze specific reactions producing dUMP and dTMP, respectively (Figure 2). Other enzymes in these pathways have a wider substrate pool: nucleoside-diphosphate kinase (NDPK), nucleoside-monophosphate kinase (NMPK), and dTMP kinase (TK) act on nucleotides containing different bases. Nucleotide kinases (NK and TK1) are involved in salvage pathways utilizing 2'-deoxynucleosides from external sources to produce the corresponding 2'-deoxynucleotide-monophosphates. The folate cycle (involving dihydrofolate reductase (DHFR) and serine hydroxymethyltransferase (SHMT)) is essential to recycle methylene-tetrahydrofolate (MTHF), the co-substrate of TS that serves as a methyl donor for dTMP synthesis.

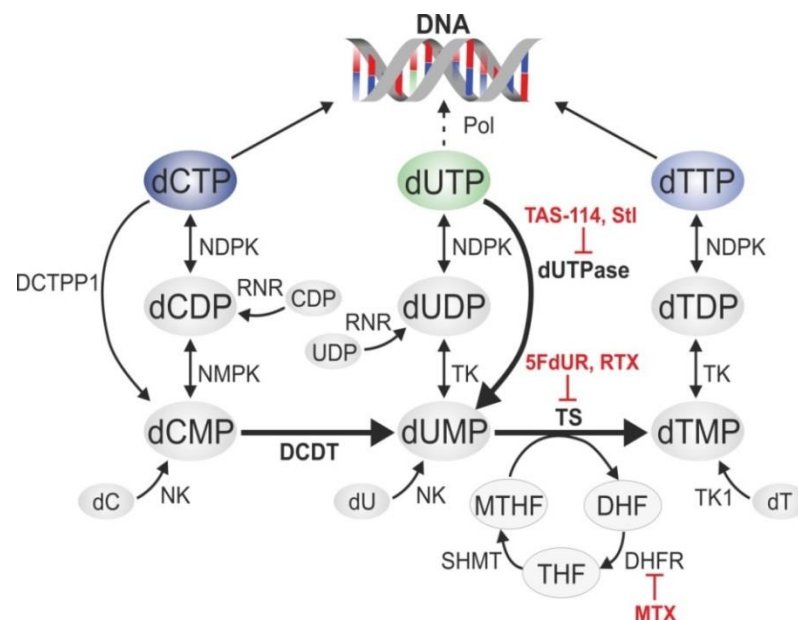


Figure 2. Pathways and potential target enzymes in thymidylate biosynthesis in eukaryotes. Thymidylate synthase (TS) produces deoxythymidylate (dTMP) from deoxyuridylate (dUMP). Inefficient catalysis of this process leads to a decrease in the cellular dTMP level that can result in deoxyribonucleotide triphosphate (dNTP) pool imbalance and perturbed DNA replication. Thymidylate synthase inhibitors raltitrexed (RTX) and 5-fluoro-2'-deoxyuridine (5FdUR) are indicated. For efficient TS activity, the 5,10-methylene tetrahydrofolate (MTHF) co-substrate is indispensable that is recycled from dihydrofolate (DHF) in the folate cycle involving dihydrofolate reductase (DHFR) and serine hydroxymethyltransferase (SHMT) via a tetrahydrofolate (THF) intermediary. Antifolate derivatives (e.g., methotrexate (MTX)) inhibit the folate cycle. Furthermore, production of the precursor dUMP is ensured by two processes: catalyzed by deoxyuridine triphosphatase (dUTPase) and dCMP deaminase (DCTD) together with deoxycytidine triphosphatase (DCTPP1). dUTPase inhibitors TAS-114 and Sfl are also shown. Further abbreviations are as follows: nucleoside-diphosphate kinase (NDPK), ribonucleotide reductase (RNR), nucleoside-monophosphate kinase (NMPK), nucleoside kinase (NK), thymidine kinase (TK1), dTMP kinase (TK), DNA polymerase (Pol).

In thymidylate biosynthesis, TS plays a key role and its inhibition causes massive decrease in the cellular dTMP level, leading to an increased dUTP/dTTP ratio and massive

uracil incorporation into genomic DNA [53]. Notably, TS also has an RNA binding function, regulating the translation of its own mRNA in a negative feedback loop, and also the mRNA of p53 [55–58]. Further effects include imbalances in the overall dNTP pool that eventually lead to programmed cell death. Direct targeting of dUTPase by small molecular drugs (e.g., TAS-114 [59–61]) or the protein inhibitor Stl also provide a promising possibility for anti-cancer therapies [14,62–68]. RNAi-directed suppression of dUTPase has been shown to increase cellular sensitivity towards TS-inhibiting drugs [69,70].

Inhibitors of TS and DHFR are among the most widely used anticancer drugs (Figure 2), and a wealth of experimental and clinical evidences is available about the possible mechanisms and their respective efficiency, as well as various sensitizing and resistance factors [71]. Two major mechanisms are involved: fluoro-pyrimidines and their active metabolites efficiently block and modify the nucleotide binding site of TS, while antifolates, being structurally similar to folates, compete with binding of the MTHF co-substrate.

In summary, uracil in DNA can be considered as undesired DNA damage, which also can be exploited in anti-cancer therapeutic strategies. Moreover, genomic uracils can serve as signals possibly involved in crucial biological processes such as immune diversity, antiviral defense, and epigenetic regulation of transcription. To better understand these possible roles, it is essential to apply state-of-the-art methodology for characterizing genomic uracil patterns. In this review, we provide a comprehensive and critical assessment of the repertoire of available methods measuring uracil-containing DNA (U-DNA) in different contexts with a special focus on next-generation sequencing (NGS)-based genome-wide mapping solutions.

2. Uracil-DNA Detection Methods

Investigation of DNA damage, repair, and epigenetic base modifications became a rapidly developing scientific field, especially in the last decade, fed by numerous new technical solutions such as a new generation of DNA sequencing approaches [72,73]. Li and Sancar provided a comprehensive overview on crucial methods and developments in the field of genome-wide DNA-damage mapping approaches; however, uracil was not fully covered in their work [72]. Another review from Sturla's lab focuses on NGS-based DNA damage sequencing methods, providing a thorough categorization based on the different technical solutions for library preparation [73].

Here, we provide a summary of diverse uracil-DNA detection methods with their advantages and limitations, and discuss their results and conclusions. We detail the global quantitative U-detection methods as well as various emerging solutions for in situ uracil-DNA detection. Regarding the genome-wide mapping methods, relevance, and benefits of single base resolution, as well as the potential pitfalls in data analysis, are also considered.

2.1. Global Quantification of Uracil in DNA

The most straightforward way to quantify the overall uracil content of a DNA sample is a liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) method [74]. It is based on enzymatic digestion of the DNA to 2'-deoxy-ribonucleosides using DNase I and nuclease P1, followed by a preparative HPLC purification coupled MS/MS identification of deoxyuridine (dUrd) and employs an isotope labelled internal standard. With this approach, and systematically addressing possible technical pitfalls, the uracil content of the murine and the human genome was determined to be ~0.15 and ~0.08 uracil/10⁶ bases, respectively, considerably lower than suggested previously by other MS-based methods [75–78]. Ung deficiency led to some increment up to ~1.2 and ~0.35 uracil/10⁶ bases, respectively.

Another approach utilizes alkoxyamine-based aldehyde reactive probes (ARPs) to chemically label the aldehyde group in the deoxyribose moiety at AP sites [79]. Biotinylated ARP reagents were used for the detection of oxidative base damages and AP sites on (ELISA-like) dot blot application [80]. The Ung-ARP assay was developed in Bennett's group, where specific enzymatic removal of the uracil and detection of the resultant

AP sites by biotinylated ARP reagent were combined [81]. Further developments led to two alkoxyamine reagents, AA3 and AA6, associated with increased reactivity and functional groups, appropriate to conjugate with a wide variety of biochemical labels by click chemistry [82,83]. These reagents were used in different applications [84–86].

As an independent approach, a new U-DNA sensor protein was developed for multiple purposes, including quick one-step semi-quantitative dot blot application where uracil is directly recognized, without any further enzymatic or chemical reactions [28]. For this, the inactive D145N/H268N double mutant of human UNG was used as a starting construct from which the N'-terminal 84 residues were deleted to eliminate undesired protein–protein interaction surfaces (Δ UNG sensor) [87,88]. It was demonstrated that such UNG-based sensor equipped with 3xFLAG tag is an appropriate tool in dot blot application to quantify uracil as compared to a standard with known uracil content [28]. Uracil levels were determined in both bacterial (wt, *ung*^{-/-}, and also *ung*^{-/-} *dut*^{-/-} double mutant *E. coli*), and higher eukaryotic (*Drosophila* S2 cells, as well as human colon cancer cell line HCT116) genomes upon treatments with thymidylate biosynthesis inhibitory drugs. The fast and straightforward dot blot applications do not require mass spectrometry infrastructure; however, the mass spectrometric methods provide higher accuracy, especially at low uracil levels.

2.2. In Situ U-DNA Detection Methods

Two early approaches, radio labelling [89] and a modified comet assay [90], were applied for in situ U-DNA detection, allowing quantitation with only low resolution. Recently, the Δ UNG sensor described above was further developed to allow in situ detection as well. First, fluorescently tagged Δ UNG sensors were shown to be appropriate for the in situ detection of uracil-containing exogenous plasmid DNA within the context of eukaryotic cells [28]. Later, this sensor was equipped with SNAP-tags and used in super-resolution fluorescent microscopy (STED and dSTORM) to detect endogenous genomic uracils in human cells [91]. The FLAG-tagged Δ UNG sensor was also used for genome-wide mapping (U-DNA-Seq), and the combination of the two approaches revealed that the RTX and the 5FdUR treatment induced uracil-enriched regions/loci colocalized with the active histone mark, H3K36me3, and the facultative heterochromatin mark, H3K27me3, respectively [91].

The mycobacterial UdgX that forms a covalently trapped complex with uracil-DNA was also employed for U-DNA detection [92]. The UdgX enzyme belongs to a bacterial UDG family harboring an Fe–S cluster, with a sequence motif KIRRH (called R-loop) essential for its function. Three-dimensional (3D) structures of UdgX complexes together with LC-MS/MS analysis revealed a covalent link between His109 of the KIRRRH track and the deoxyribose at the AP site [93–95]. It is straightforward to use this unique enzyme as a uracil-DNA sensor, similarly to the Δ UNG described above. An mCherry tagged UdgX was constructed and characterized in detail as a highly sensitive U-DNA sensor for detecting U-DNA by confocal microscopy in wild type (wt) and *ung*^{-/-} *dut*^{-/-} *E. coli* [96]. A FLAG-tagged UdgX-based sensor was also used in human cells to detect uracils in ssDNA arising upon the induction of APOBEC3A in cisplatin-treated HEK293T cells, revealing uracil colocalization with replication protein A in stalled replication forks [85]. An advantage of this sensor might be that it can be applied by transfection into living cells, and then only the immunodetection steps have to be carried out in fixed samples. UdgX demonstrated a strong (3–5 orders of magnitude) preference towards uracils in ssDNA context over U:A in dsDNA that, on one hand, ensures a selectivity, and on the other hand, somewhat limits its applicability for addressing certain biological questions. It is important to note that the SNAP-tagged Δ UNG construct recognizes uracils within fixed cells in all those contexts that are normally recognized by the wt UNG (ssU, U:A and U:G pairs).

These in situ detection solutions provide potent tools within highly different biological samples for relatively quick and efficient detection of genomic uracils either upon their increasing levels (e.g., drug treatments [91]) or upon spatial clustering into genomic loci (e.g., targeted enzymatic cytosine deamination [85]). Earlier, in situ detection method for

AP sites were available [77], but their application for detection of genomic uracil is not straightforward at all. With the new direct approaches described above, it is also possible to identify (or even screen for) new biological situations and/or conditions, where particular patterns emerge indicating special potentially novel biological roles of genomic uracil. In combination with genome-wide localization data that could provide good candidate protein markers for in situ colocalization studies, these sensor constructs and the coupled detection strategies might especially be powerful. Such in cell studies are more cost-efficient and flexible for the wide screening of treatment-induced changes in various biological samples as compared to the genome-wide NGS-based sequencing methods. However, without the knowledge of genome-wide distribution, just the in situ detection alone could not provide essential new insight to the diverse roles of genomic uracil. Hence, combination of the two approaches is indispensable.

2.3. PCR-Based Methods for Uracil Localization within DNA

To localize uracils within a target DNA sequence, several PCR-based methods are available that can provide either an indication for the presence of U:G pairs, or exact localization (even with single-base resolution), or accurate quantitation (within the target sequence) of the uracils. The first published technique to detect C:G to T(U):A transitions due to cytosine deamination in DNA was the differential DNA denaturation PCR (3D-PCR) [97]. This technique relies on the lower denaturation temperature (T_d) of DNA templates with higher AT content. It applies gradiently lowered T_d in the PCR reactions, amplifying a specific target sequence defined by the two PCR primers. The specific PCR product could be detected already at lower T_d in those cases where some C:G to T:A transitions happened in the template DNA within the amplified region [97]. Later, this 3D-PCR technique was applied in combination with a UNG inhibitor, UGI, to detect uracil-DNA intermediates of APOBEC3A-catalyzed cytosine deamination in a reporter plasmid DNA [22].

Almost at the same time as the publication of 3D-PCR, the Gearhart group applied a combined in vitro reaction of UNG and APE1 to detect uracil in an exogenous plasmid from which AID was expressed in bacteria [98]. They could show AID induction-related increases in the number of nicks by a UNG/APE reaction on alkaline agarose gel, and could locate the uracil moieties on the non-transcribed strand using a denaturing Southern blot. Furthermore, they also applied polymerase β without the addition of dNTPs, just using its 5' dRP hydrolyzing function to introduce nicks to the site of the uracil. Then, they applied primer extension on the nicked template using a specific and biotinylated primer that results in dsDNA end that is appropriate for blunt end adapter ligation. By clonal sequencing of the products of this ligation mediated PCR (LM-PCR), they could localize the original positions of uracils within the non-transcribed strand of the AID-expressing plasmid with single base resolution [98]. Later, they further developed this technique and successfully detected uracils in the immunoglobulin genes of ung^- deficient AID expressing B cells as compared to ung^- deficient $Aicda^{-/-}$ cells (ung^- deficient chicken DT40 either overexpressing the chicken AID or $Aicda^{-/-}$ clones, and B220+GL7+ spleen cells from $ung^{-/-}$ and $ung^{-/-}$, $Aicda^{-/-}$ mice that were either immunized right before the cells were isolated, or the isolated cells were stimulated ex vivo with LPS and IL-4) [99].

Another PCR-based approach simply quantifies the difference between the amounts of intact templates in the samples pre-treated with UNG alone or UNG + APE as compared to the non-treated one. Such quantification could be performed by qPCR, and also by the more convenient digital PCR techniques (such as digital droplet PCR (ddPCR)). In the ex-ddPCR method, samples are treated with UNG, and the fraction of amplicons containing at least one uracil on each strand is determined from positive PCR counts in the treated and un-treated samples [100]. It was shown that the viral gag gene accumulates uracils only in monocyte-derived macrophages (MDM), but not in T cells.

While these approaches above rely on specific enzymatic reactions by UNG and APE, a fully independent PCR-based method utilizes the altered sensitivity of archaeal DNA

polymerase Pfu and its V93Q mutant version for the uracil-containing templates [27]. The structural basis and the functional consequences of binding of the archaeal family B polymerases to uracil bases in the template DNA strand has already been well described [101]. While a single uracil base can eventually stall DNA synthesis by wt Pfu polymerase, V93Q mutant Pfu preserves its activity even on fully uracil substituted templates [102]. Applying wt and V93Q mutant Pfu in parallel PCR reactions using the same template dilution series, from the difference between the corresponding Cq values, the uracil can be quantified within the given template region defined by the two PCR primers [27].

2.4. NGS Based U-DNA Detection for Genome-Wide Mapping

All methods described above are PCR-based; hence, they are limited to determination of a local uracil content that can be valid for the whole DNA sample as much as the genome-wide distribution of the uracil is uniform. Depending on the origin of the uracil-DNA, its genome-wide distribution can be more or less patterned: enzymatic cytosine deamination might result in a strongly targeted localization (e.g., focusing on variable and switch regions of the Ig genes, or kataegic-like clusters [25]), while spontaneous cytosine deamination and thymine-replacing misincorporation due to the insensitivity of the DNA polymerases are more stochastic and random processes. In these latter cases, if any pattern exists, it should be originated from several additional mechanisms, such as altered accessibility of the differential packaged genomic DNA, the unequal distribution of repair processes, and different polymerases with altered sensitivity and specificity. Indeed, in the last decade, numerous new results support the hypothesis that distinct repair proteins [103] and/or different polymerases [104] are loaded to certain genomic loci rather differently. Since 2014, many NGS-based approaches have been published addressing epigenetic marks (ChIP-seq), or DNA methylation [105], or DNA repair loci (e.g., XR-seq, HS-XR-seq, and Damage-seq [50,106,107]), or other base modifications (e.g., OG-seq [108], click-code-seq [109]), or AP sites (e.g., snAP-seq [103]). Similarly, genome-wide uracil mapping solutions have been developed and are becoming crucial to better understand the significance and consequences of uracil appearance in DNA within the different biological contexts. For these “seq” methods, the PCR-based uracil localization or quantifying techniques described above provide essential validation opportunities.

The first published method applied for genome-wide uracil mapping was the Excision-seq applied in *E. coli* and yeast [110]. Excision-seq also operates with the coupled enzymatic reactions of bacterial UNG and the AP endonuclease, ENDO IV, and combines this with massively parallel DNA sequencing (NGS). Two versions had been developed: the pre-digestion (Figure 3a) and the post-digestion Excision-seq (Figure 3b). The pre-digestion version requires high uracil content within the studied DNA sample that allows efficient DNA fragmentation already by the *in vitro* UNG/ENDO IV enzymatic treatment. Then, applying a size selection without additional fragmentation procedure, the sequencing library is prepared. The ligation position of the sequencing adapter at the 5' ends will report on the original sites of uracils with practically a single-base resolution (Figure 3a), similarly to the ligation mediated PCR method [99] described above. As a complementary approach, post-digestion Excision-seq applies UNG/ENDO IV treatment on the prepared DNA fragment library, and the increased read coverage in the sequencing results of the excised samples compared to the non-treated controls indicates genomic regions from which uracils were excluded. However, the sensitivity of such inverse approach highly depends on the sequencing depth and requires a rather uniform genome coverage, which might limit the size and the complexity of the genomes addressed by this technique. The two versions of Excision-seq were reported as adequate methods for the efficient detection of elevated uracil levels upon dUTPase and UDG deficiency in smaller genome sizes, as in *E. coli* and yeast strains [110]. They concluded that uracil is excluded from the very early and very late replication timing genomic segments and assumed that such regulation might involve the alterations of the cellular dNTP pool during the DNA synthesis [110]. Nevertheless, a larger genome size (e.g., mammalian genomes) and the low frequency and/or the nature of the

distribution of uracils might result in some biases or underestimation using Excision-seq method, especially its pre-digestion version. In this aspect, enrichment or pull-down-based methods might be more efficient. Moreover, it has not yet been demonstrated how beneficial the single base resolution capability of pre-digestion Excision-seq is. Indeed, the same group used this method addressing the 10 kb-size HIV genomes from different in vitro infected immune cells showing uniformly distributed uracilation of the proviral genome [100]. Although this method was also extended to the mapping of other DNA base modifications [111] and cited by reviews or other research papers, Excision-seq has not yet become widely used to characterize other biological systems. In one case, pre-digestion Excision-seq was applied as a complementary technique to support the results from dU-seq (which will be discussed later [112]).

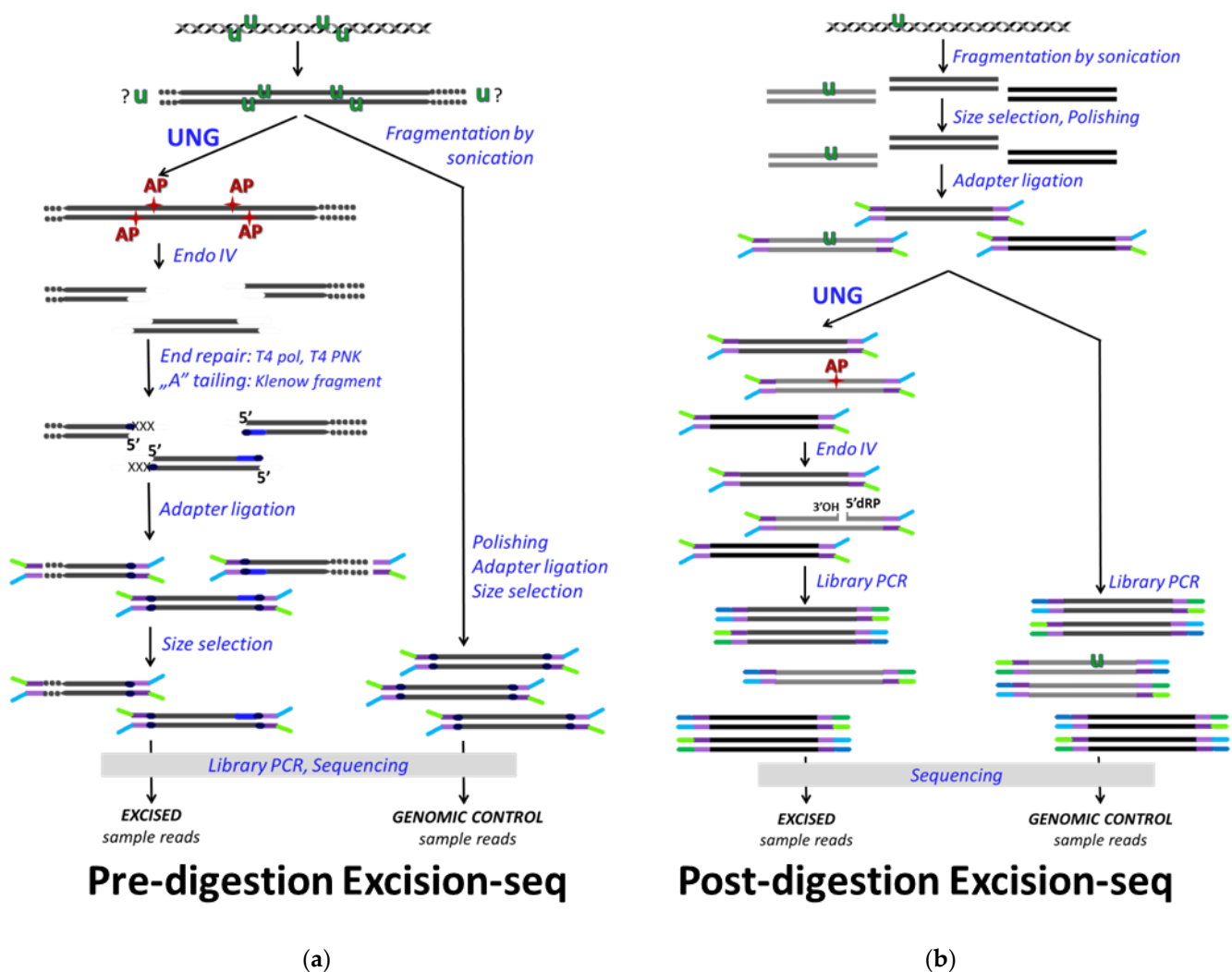


Figure 3. Excision-seq for genome-wide mapping of U-DNA. (a) Pre-digestion Excision-seq. For efficient detection, high density of genomic uracils (green U) is required, because NGS-compatible fragments originate from the UNG/ENDO IV treatment. Longer fragments (indicated with dots at the end) are lost during the size selection. End repair involves T4 polymerase (T4 pol), T4 polynucleotide kinase (T4 PNK), and Klenow fragments. Both ends of the sequencing reads report on uracil sites, but some uracils might escape detection. Non-specific hits are filtered by the comparison to genomic control sequencing. (b) Post-digestion Excision-seq. Genomic fragments are treated with UNG/ENDO IV, and degradation of uracil-containing fragments results in locally decreased read coverage as compared to the sequencing of the non-treated genomic fragments.

Meanwhile, also attempting single base resolution detection of uracil (and other DNA lesions) within the DNA, two other approaches were developed on model DNA and proposed to be used in genomic context too by Burrows' lab [113,114]. Their first method also relies on the UNG/APE enzymatic treatment (or more generally, the other base modification-specific glycosylases and the appropriate AP endonuclease or AP lyase) which is followed by enzymatic labelling of the gapped strand by unnatural nucleotides (dNaM or dMMO2). The bases of these nucleotides are selectively paired with d5SICS unnatural base in PCR reactions forming unnatural base pairs (UBPs). Such UBPs can then be detected either by Sanger sequencing (UBPs stop the seq reactions), or by nanopore sequencing technology, where the position of UBPs can be determined with single nucleotide resolution in the context of single DNA duplexes [113,115]. This method was developed and tested on synthetic DNA models, simulating biologically relevant lesions with their heterogeneous sequence context and also the effect of a large excess of undamaged DNA [113]. It was also suggested that, in combination with an enrichment of DNA lesion-containing DNA-strands, the method can be adapted for complex biological systems. Their second method relies on ligatable gaps that arose upon in vitro treatment by the glycosylase specific for a given modified base; then, sequencing of ligated products by any commonly used NGS techniques and identification of single nucleotide deletions will report on the position of the original DNA lesions [114]. Although this approach seems to be cheaper and more available for the wide scientific community, it has not yet been demonstrated that it could work on large genomes, especially with low uracil content. Basically, the limitations of this method should be similar to those in the pre-digestion Excision-seq. Indeed, the authors suggest that its best application might be in single cell sequencing, where a certain base modification is present at 100% (note: 50% in case of diploid genomes). Furthermore, the relatively high chance of a single nucleotide deletion as a consequence of sequencing error or naturally occurring variation within the sample can also impair the sensitivity.

Two other sequencing methods with single base resolutions, which were developed for the detection of AP sites but can easily be adapted for uracil detection by applying preceding UNG treatment (as is also true for other seq methods designed for AP sites), are also worth being presented here. In Balasubramanian's lab, snAP-seq was developed and used in different size of genomes and for answering different biological questions [103] (Figure 4a). snAP-seq applies a selective chemical labelling of AP sites [116], and enrichment via a biotin-streptavidin system. They demonstrated the selectivity of their method for AP site aldehyde over the formylcytosine aldehyde, via combination of the chemical labelling of the aldehyde groups and the elution from the streptavidin resin using an alkaline condition that hydrolyzes the sugar-phosphate backbone at the AP site. First, chemical labelling is performed on the fragmented DNA, then the P7 sequencing adapter is ligated, followed by the pull-down on streptavidin beads and selective elution by alkaline cleavage. Using the P7 adapter, a primer extension is performed on the eluted ssDNA fraction, then only the AP cleavage-related 5'-phosphates are available for ligation with the other sequencing adapter, P5. Hence, the enrichment of relevant DNA fragments is quite efficient, and a majority (95%) of the sequenced fragment will start exactly one base downstream the original AP sites, as it was measured in a model DNA. This method was applied for single base resolution detection of hmU in *Leishmania major* [103], where hmU is a precursor of the epigenetic marker base J and is supposed to be introduced enzymatically [117–119]. The method was also used in human cell lines to detect AP sites upon the silencing of APE1; however, its single base resolution potential could not really be exploited in this latter case, due to more randomized genomic distribution of the AP sites [103].

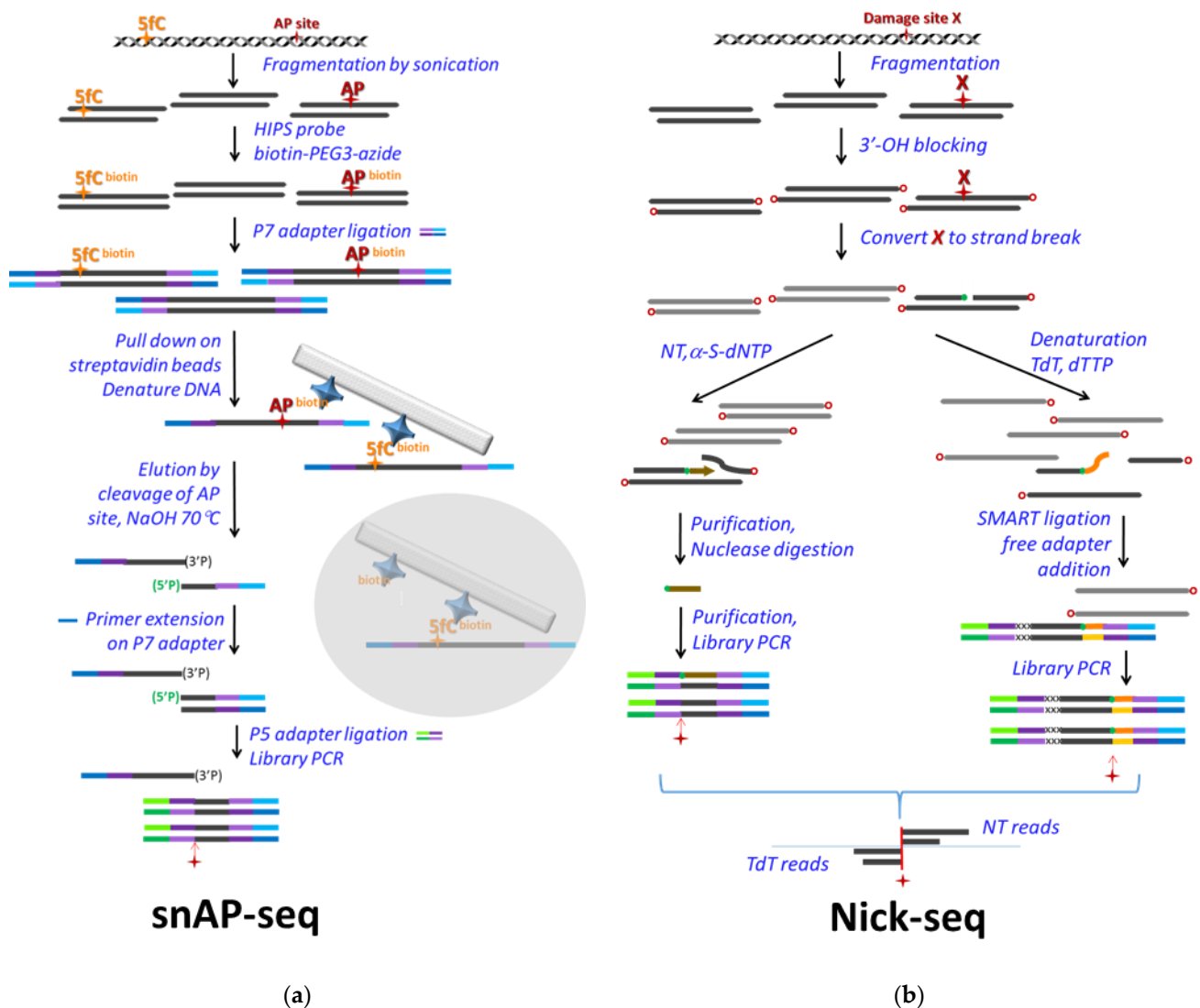


Figure 4. (a) snAP-seq. Genomic DNA fragments are chemically labelled by an aldehyde-reactive HIPS probe [116], either at the AP sites (red star) or at formylcytidine (yellow “5fC”) or formyluridine. The HIPS probe is labelled by biotin via a PEG3 linker (biotin–PEG3–azide) which allows efficient pull-down of the P7 adapter ligated and labelled fragments on streptavidin beads (grey rectangle represents the solid phase, blue star: streptavidin tetramers). After denaturation of the dsDNA bound on the beads, AP site fragments are selectively eluted by alkaline cleavage, when only one of the DNA strands downstream from the AP site will be appropriate for ligation. Upon primer extension from the P7 adapter, the resulted dsDNA will also be ligated to the P5 sequencing adapter. The PCR amplified library is sequenced, where the first base of the P5 reads will report on the AP sites (red star). (b) Nick-seq. First, 3'-OH groups are blocked by dideoxy nucleotide addition (red “o”) to the fragmented genomic DNA. Then, damage sites (red “X”) are converted to ssDNA breaks with free 3'-OH (green dot). Further processing is performed parallelly in two portions of the sample. NT: nick translation by polymerase I using α -thio-deoxynucleoside triphosphates (α -S-NTP) precursors that will result in phosphorothioate oligonucleotides resistant for nuclease treatment. These purified fragments are used for sequencing adapter ligation, library PCR and sequencing. TdT: terminal transferase is used to tail the free 3'-OH with poly(dT) using the commercially available SMARTER Chip-seq kit (Clontech). Addition of the sequencing adapters then depends on the poly(dT) tail, and applies a template-switching polymerase reaction instead of ligation, ensuring higher specificity of the library preparation. The sequencing results from NT and TdT processing are compared, and reads are positioned next to the damage site downstream and upstream, respectively.

The other similarly creative method is Nick-seq, developed by Dedon’s group for single-nucleotide resolution genomic maps of different DNA modifications and damage [120] (Figure 4b). The method relies on conversion of the modified bases into single

strand breaks on which two different types of polymerase reactions are performed separately. One portion of the sample is subjected to nick translation using α -thio-dNTPs to produce hydrolysis-resistant oligonucleotides downstream of the single strand break. Hence, the rest of the DNA can be selectively removed by exonuclease III and RecJ, and the resistant phosphorothioate-containing oligos can be sequenced. The second portion of the sample is used for poly(dT) tailing of the 3' end at the strand break by terminal transferase (TdT). Then, this tail is used for library preparation. By this approach, sequencing of the two separately processed samples can confirm the position of a base modification from two directions.

A novel genome-wide uracil detection method, dU-seq, was also combined with the pull-down technique [112] (Figure 5a). This method applies an enzymatic cascade including *E. coli* UDG to convert genomic uracils into AP sites that are cleaved by ENDO IV, and the gaps are resynthesized by Bst DNA polymerase in the presence of biotinylated nucleotide triphosphate. The biotinylated DNA fragments then pull down on streptavidin beads, where the Y adapter for sequencing is ligated before the elution is conducted for 3 min at 95 °C in distilled water. The eluted DNA fragments are amplified by PCR and sequenced by Illumina. Prior to the enzymatic treatments, repair of AP sites, ssDNA breaks, and ssDNA ends were performed. The input and the enriched samples were sequenced, and peak calling was performed by model-based analysis of ChIP-seq (MACS2) software. Only peaks uniquely present in the pull-down versus the control were considered in the consequent analysis. Uracil enrichment within the centromeres was reported, which was also confirmed to some extent by independent methods including pre-digestion Excision-seq, LC-MS/MS, and 3D-PCR.

The UDP-seq method, another DNA-IP-seq application, is quite similar to dU-seq except that the introduction of the biotin label to the uracil sites is performed in a non-enzymatic chemical reaction [86] (Figure 5b). The alkoxyamine moiety of the commercially available reagent EZ-Link Alkoxyamine-PEG4-SS-Biotin (ssARP) (Thermo Scientific) covalently labels the opened ring of the base-free deoxyribose at the AP site. The S-S bridge allows efficient elution by reducing agents such as dithiothreitol (DTT). Chemical blocking of pre-existing AP sites that otherwise could interfere with uracil detection is necessary. Application of UDP-seq in bacterial systems showed, on one hand, that upon dUTPase and UNG deficiency, the elevated uracil incorporation occurs mostly at the replication origin. On the other hand, the ectopically expressed APOBEC3A (A3A) catalyzed cytosine deamination patterns were addressed in *E. coli*, where both UDGs are mutated. The control pull-down without UNG treatment to check for non-specific binding was omitted (Figure 5b vs. Figure 5a), but control samples were introduced: active A3A-expressing cells were compared to either inactive A3A or empty plasmid-containing cells. In the data analysis, peak calling was performed with MACS, and a normalized differential coverage (NDC) for 100 bp moving window was calculated. A uracilation index (UI) was introduced to measure the frequency of TC to TT transitions specific for the A3A, and a preference of A3A activity was detected for the lagging strand during replication, as well as in short hairpin loops, tRNA and rRNA genes, and in the 5' termini of some protein coding genes.

In both dU-seq and UDP-seq, numerous enzymatic/chemical steps are involved, resulting in a complex arrangement with multiple potential pitfalls. Additionally, abasic sites independent from uracils need to be carefully considered. Moreover, sticky DNA ends could influence the polymerase-based labelling approach in dU-seq. Accordingly, both dU-seq and UDP-seq are based on well-established experimental setups and take advantage of the highly efficient biotin–streptavidin pull-down system. Processing and the interpretation of the NGS data involve several critical issues that will be detailed in Sections 3 and 4.

The most recent pull-down-based method, U-DNA-Seq, employs U-DNA-specific binding of the FLAG-tagged Δ UNG sensor (already described above) to pull down uracil-containing genomic DNA fragments. As such, it is a more direct method and involves less complex steps than dU-seq and UDP-seq (Figure 5c) [91]. It is independent from

the efficiency of different enzymatic/chemical reactions used in dU-seq and UDP-seq and relies on the specificity and affinity of the interactions between Δ UNG and U-DNA, and the anti-FLAG antibody and the FLAG-tag, respectively. In the published paper, U-DNA-Seq was applied in the human cancer cell line HCT116, and its mismatch repair proficient version, where the UNG inhibitor UGI was stably expressed [91]. The effects of two thymidylate synthase inhibitory drugs, 5FdUR and RTX, on the genomic uracil content and its distribution were addressed. Using their own analysis pipeline, remarkably high reproducibility among replicates (cf. Supplementary Materials in [91]) was presented even when results were compared to relevant samples from the published dU-seq data remapped and re-analyzed by the same pipeline.

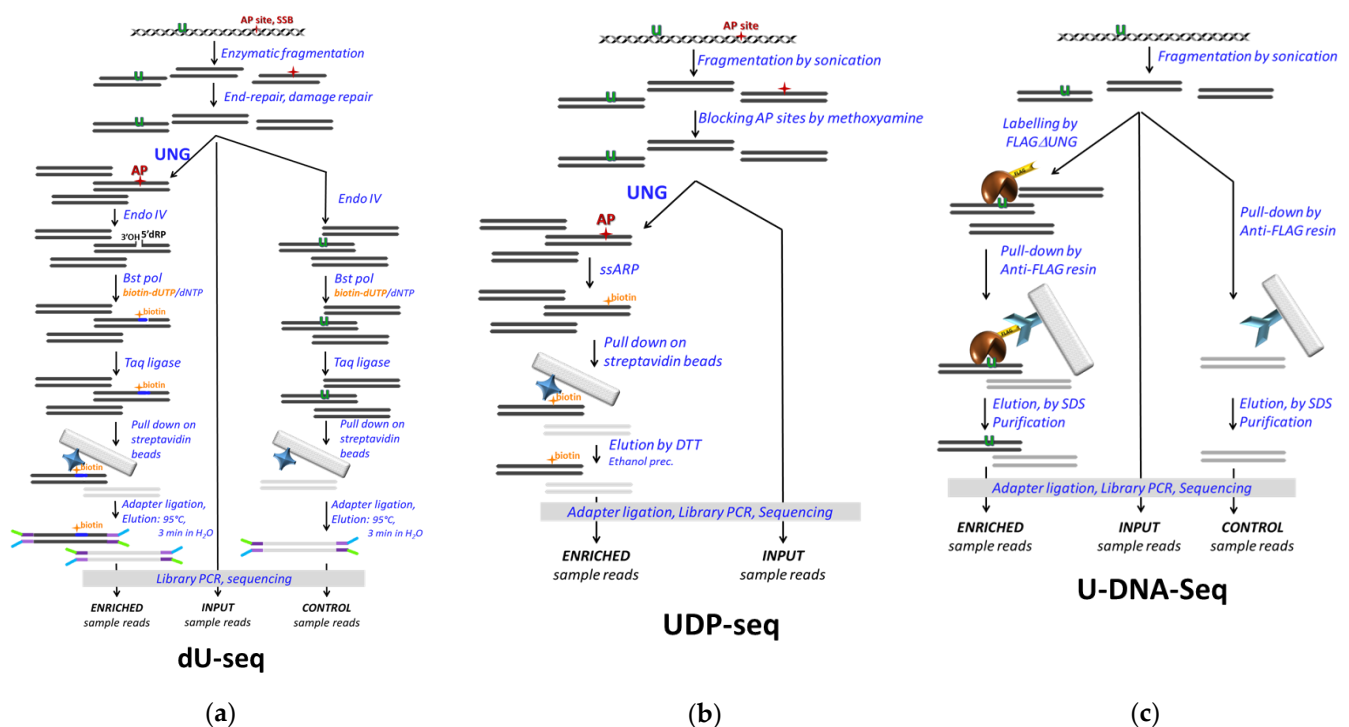


Figure 5. Pull-down-based methods for genome-wide mapping of U-DNA. **(a)** dU-seq. Pre-existing AP sites (red star) and single strand breaks (SSBs) require in vitro repair before the combined UNG/ENDO IV reaction on uracil (green U) containing genomic DNA fragments. Bst polymerase (Bst pol) is used to add nucleotides utilizing biotin-labelled dUTP to the free 3'-OH end and to remove the 5'-deoxyribosephosphate (5' drp). The biotin labelled and re-ligated fragments are pulled down by streptavidin beads. Sequencing results are compared to the control pull-down without UNG treatment. **(b)** UDP-seq. Pre-existing AP sites (red star) are blocked before the UNG treatment. Uracil-related AP sites are chemically labelled by an aldehyde reactive probe (ssARP) conjugated to biotin via a disulfide bridge-containing linker allowing efficient elution by dithiothreitol (DTT) upon pull-down. Sequencing results of pull-down sample is compared to the genomic input. **(c)** U-DNA-Seq. Uracils in genomic DNA fragments are specifically labelled by the FLAG-tagged Δ UNG sensor (FLAG Δ UNG) and pulled down using anti-FLAG agarose. Sequencing results of pull-down sample is compared to the genomic input, and non-specific binding is estimated by a control mock pull-down.

The experimental design, the applied biological models, and data analysis, as well as basic conclusions of the studies using the above described pull-down-based U-DNA mapping methods are summarized in Table 1.

It is interesting to consider if a similar pull-down coupled sequencing method could be developed based on UdgX. Sequencing over the crosslink might be challenging, although may not be impossible to solve. Considering that UdgX has strong preference towards uracils in ssDNA [85], and once it binds to it, other UDGs cannot initiate its conversion to AP sites anymore [92], such an approach could provide selective and safe enrichment of otherwise more vulnerable uracil-containing ssDNA fragments directly from cells.

Moreover, high throughput analysis of the crosslinked peptide–DNA fragments by mass spectrometry could also provide single base resolution data wherever it is interesting.

Table 1. Summary of studies based on dU-seq, UDP-seq and U-DNA-Seq. Abbreviations are as follows: *ung*, *dut*, *mug*, and *hmlh1* are genes encoding UNG, dUTPase, double stranded U-DNA specific UDG of *E. coli*, and human MLH1, respectively. MMR: mismatch repair, A3A and A3A*: APOBEC3A and its inactive mutant, EV: empty vector, NDC: normalized differential coverage, UI: uracilation index.

	dU-Seq	UDP-Seq	U-DNA-Seq
Genome	human	<i>E. coli</i>	human
Gene deficiency	<i>ung</i> ^{-/-} wt	<i>ung</i> ⁻ / <i>dut</i> ⁻ vs. wt, <i>ung</i> ⁻ / <i>mug</i> ⁻	<i>hmlh1</i> ^{-/-} Restored MMR
Transgene	- <i>ung</i> UDGs	- A3A, A3A* EV	<i>ugi</i> -
Treatment	- 5FdUR	- -	5FdUR, RTX -
Data pre-processing	Trim, Align	Trim, Align, Filter	Trim, Align, Blacklist, Filter
Enrichment analysis	Peak calling	Peak calling, NDC, UI	Coverage, log2 ratio, broad regions
Conclusion	Centromeric enrichment	<i>ung</i> ⁻ / <i>dut</i> ⁻ : replication origin A3A in <i>ung</i> ⁻ / <i>mug</i> ⁻ : lagging strand, hairpin loops, tRNA genes	Heterochromatin, upon treatments: shifted towards early replicating and active euchromatin

All of the methods described above potentially capable or used for genome-wide mapping of uracil moieties have advantages and limitations. The Excision-seq (pre- and post-digested versions), dU-seq, snAP-seq, UDP-seq, Nick-seq and U-DNA-Seq were used for genome-wide studies at different levels of genomic complexity, and for different biological samples with markedly different level and origin of uracil bases. Thus, the evaluation of the sequencing results as well as drawing conclusions requires appropriate considerations.

3. Factors to Consider in Analysis of NGS Data to Maximize Relevant Information While Avoiding Over- or Misinterpretation

We have presented a wide overview of genomic uracil detection methods applied to provide insights into profoundly different biological problems. Appropriate data processing and analysis are essential to avoid over- or misinterpretation. To promote the research interest in this field, we hereby discuss some potential pitfalls of data analysis that can easily lead to overinterpretation of the results. Carefully designed and performed experiments with appropriate controls are as crucial as the appropriate choice of analysis pipeline, including double-checking all calculation steps because even the widely used software often suffers from hidden bugs (as it is reflected in the open issues on GitHub).

Analysis of NGS data that are designed for mapping a feature (i.e., uracil) to the genome relies on an initial alignment of the short sequencing reads to a chosen reference genome set. Importantly, none of the reference genomes are fully relevant for any particular biological sample due to the unique character of the individual genomes with multiple polymorphisms including copy number variations. The current human reference genome (Genome Reference Consortium Human Reference 38 (GRCh Build 38 or hg38 in GeneBank), <https://www.ncbi.nlm.nih.gov/grc/human>, accessed on 7 April 2021) actually exists in multiple and somewhat different sets. These include not only the main chromosomes, but different sets of hundreds to thousands extra scaffolds, such as unlocalized (UnChr . . .) or unplaced (Un . . .) genomic segments, established genomic alterations (ALT . . .), viral, and other decoy sequences. Moreover, critical segments might or might not be masked in these sets, where certain gaps also occur. Therefore, the choice of the reference genome set

can influence the resulting read alignments. Alignment files usually cannot be shared in databases such as Gene Expression Omnibus (GEO); therefore, unequivocal definition of the used reference genome set is crucial in all publications to promote reproducibility.

The high frequency occurrence of repetitive regions in the human genome unfortunately interferes with mapping of many short sequence reads. Ambiguously mapped reads are indicated with zero mapping quality value within the alignment files but are not automatically excluded from downstream analysis. There is always a trade-off between loss of information and allowing the possibility of incorrect alignment. As we have already demonstrated [91], applying a sample specific blacklist and filtering for ambiguously mapped reads might be a solution to avoid misinterpretation of the data. A blacklist appropriate for the given biological sample should be defined by low mappability and by ultra-high signal (UHS) regions in addition to the otherwise quite narrow universal blacklist recommended by the ENCODE [121].

Certainly, if the research question focuses on such critical low mappability regions, blacklisting the whole region of interest is not an option. In such cases, application of strict measures and/or alternative approaches is required. The human centromeres are among the best examples for such critical regions with megabase-size higher order repeats consisting of several short alpha-satellite sequences [122], for which model sequences have already been implemented in the GRCh38 reference genome [123]. Although these models allow alignments to the centromeres, it should be done with extra care. In the dU-seq method [112], although the whole experimental design reflects supreme care and accuracy, the details of bioinformatic analysis are under-documented and may suffer from some artefacts. Namely, ambiguously mapped reads were most probably not excluded from the downstream analysis that was completely based on the results of the peak calling. In a highly repetitive region with potential massive copy number variations, the read alignment necessarily suffers from high amount of ambiguously mapped reads and UHS segments. Under these circumstances, local increases and decreases in the read coverage are not reliable; hence, peak calling algorithms can certainly detect more false positive peaks. Therefore, peak calling is not an appropriate approach in this case. Instead, a chromosome-wise comparison of the normalized read counts (RPKM) within the centromeres versus the non-centromere regions would have been more reliable, which was performed in the evaluation of the complementary Excision-seq experiment, but not for the dU-seq data. Based on the distribution of detected peaks in dU-seq data, a high centromeric uracil enrichment was suggested (up to 20–25-fold in wt and *ung*^{-/-} HEK293T, reported in Figure 5c in [112]). This result was confirmed to some extent by LC-MS/MS quantification (2.2-, 5- and 7-fold centromeric uracil enrichment in wt HEK293T, K562 and WPMY-1 cells, Figure 3e in [112]), and also by Excision-seq (approximately 1.6-fold centromeric uracil enrichment in K562 and WPMY-1 cells, Supplementary Figure S7 in [112]). The much lower centromeric uracil enrichment detected by Excision-seq and LC-MS/MS quantification as compared to dU-seq was contributed to the limited sensitivity of the method and with lower efficiency of the centromeric DNA isolation protocol, respectively. The dU-seq data were reanalyzed and this issue was discussed in detail [91].

To judge the reliability of any analysis, the best measures are the correlations among the replicates for which different statistical tests are available. We suggest that figures in the main text should represent merged data from the replicates; however, the same analysis for the individual replicates and correlation measures should definitely be reported at least as supplementary materials. In case of U-DNA-Seq performed on HCT116 cells, such analysis showed that the called peaks are less reproducible and have lower description value for the genomic uracil patterns as compared to the proposed broad genomic regions derived from log₂ enrichment tracks [91]. Individual replicates have an inherent variance due to natural alterations among biological samples and the limited depth of the sequencing as compared to the “depth” of the actual DNA sample. If genomic DNA is sequenced from a cell line, the starting material typically represents about 10 million diploid cells, while the average sequencing depth is only 6–7-fold in the case of typical 20 Gb, ~130 M raw reads, 150 PE

Illumina sequencing. Depending also on the original genetic heterogeneity of the given cell line, such sampling in the sequencing itself leads to somewhat limited significance of the results. This can be controlled by the statistical analysis of the individual replicates.

4. Single Base Resolution—When Is It Truly Relevant?

The recent literature in the field reflects high demand on gaining base-resolution data on genomic distribution of certain DNA damage and base modifications [72,73]. Among potential uracil-DNA sequencing approaches, pre-digestion Excision-seq, snAP-seq and Nick-seq are potentially capable of identifying the exact position of individual uracils.

From these methods, only pre-digestion Excision-seq was used directly for mapping genomic uracil in *E.coli* and yeast models with UNG and dUTPase deficiency, where a high rate of thymine-replacing uracil incorporation is expected [110]. Although the capability for single-base detection of the pre-digestion Excision-seq was demonstrated, it was not really exploited in these systems because of the stochastic distribution of the incorporated uracils. Excision-seq was further used for 10 kb-size HIV genome where a thymine-replacing incorporation mechanism was also suggested, and the single base resolution was not relevant to describe the relatively homogeneous distribution of uracils [100]. The only case when Excision-seq was used in the context of complex vertebrate genomes was a validation experiment for the dU-seq data, where a modest (~1.6-fold) centromeric enrichment of reads was observed as compared to the genomic average [112]. The potential benefit of the single-base data was not addressed in this study, i.e., there was no attempt to extract consensus sequence motifs around these uracils; however, the conclusion assumes some regulatory mechanisms behind the proposed centromeric uracil enrichment [112].

In contrast, snAP-seq was used for the mapping of hmU sites with truly single base resolution in the *L. major* genome, where hmUs could be selectively converted to AP sites using SMUG1 uracil-DNA glycosylase [103]. A considerable number (3200) of sites were identified with high confidence within four replicates that were in good agreement with the previously measured extremely high frequency data (~100/10⁶ bases) in the 33 Mb genome and also with the previously reported low resolution genome-wide data [124]. It was confirmed that these sites corresponded to hmU and not U or other U derivatives by using a negative control experiment using UNG (not capable of excising hmU). In *L. major*, hmU was shown to serve as a precursor of epigenetic mark base J involved in transcription termination; hence, it is supposed to be introduced to the genome most probably by dedicated enzymes [117–119]. This amazing performance was further rationalized by the analysis of the sequence context that revealed some preferences indicative for an enzymatic reaction by which hmU is introduced to the *L. major* genome. In contrast to this representative single-base resolution detection in *L. major*, the same snAP-seq method failed to detect high confidence AP sites in HeLA cells, even upon silencing APE1 [103]. Based on this experience, it was concluded that there are no well-defined APE1 substrate AP site positions within the genome, but wider stretches were found where the probability of AP site occurrence was higher. The genome-wide co-localization analysis of these regions with known genomic features revealed new information about typical genomic location where APE1 is active (promoter, UTR, and exon segments); however, a lower resolution method would have also been appropriate to provide the same information.

This study excellently demonstrated that the single-base resolution detection has relevance only in those cases where uracils are introduced by targeted enzyme reactions to preferred positions that should then also imply some regulatory role. Hence, application of such techniques can be extremely beneficial to explore new biological contexts, where uracil might serve as an epigenetic signal. If such properties are assumed, it is an obvious opportunity to check the immediate sequence context of the mapped sites searching for conserved recognition motifs. Detection of such motifs on the one hand strengthens the reliability of the base-resolution distribution data, while on the other hand it might provide a strategy to identify new enzymes responsible to introduce uracil into the specific genomic loci.

In the case of genomic uracil, intriguing questions arise on the widely studied field of cytidine deaminases; AID/APOBECs involved in basic processes of acquired immunity and in antiviral responses. The sequence preferences and other regulatory factors in the targeting of AID/APOBECs were extensively addressed by reporter systems involving mutational analysis [125] and even a base-resolution U-DNA sequencing technique (LM-PCR [98]). However, genome-wide studies were performed only with indirect mutational analysis [25], or with the lower resolution UDP-seq within the context of a limited-size bacterial genome [86]. To the best of our knowledge, no genome-wide and single-base resolution study has yet been published in this field.

Notably, even with a lower resolution genome-wide method, it is possible to gain relevant new information if uracils are introduced in a more targeted way to the genome. The UDP-seq method was used to analyze genomic DNA from APOBEC3A expressing *ung⁻ / mug⁻ E. coli* (representing a veritable lack of UDG function). Results showed somewhat different patterns for the wt APOBEC3A as compared to the controls (inactive APOBEC3A or empty vector). Focusing on the known minimal target motif of APOBEC3A (TC) and calculating a uracilation index (UI) for each TC in the genome revealed that APOBEC3A acts preferentially on the lagging strand [86].

It is important to consider that both spontaneous cytosine deamination and uracil incorporation through dUTP pool increase are basically stochastic processes. In these cases, single base resolution may have true impact only in single-cell sequencing, because the actual positions of uracils are expected to be variable in every single cell. However, it is well demonstrated that even in these cases, the genomic uracil distribution is not fully random, but is influenced by several circumstances. In the case of drug-induced uracil incorporation, the main determinant might be the local rate of ongoing DNA synthesis, either replicative or repair-related [91]. Genomic patterns of uracil enrichment upon the effect of two TS inhibitory drugs indicated a strong correlation with replication timing and, to a lesser extent, with the more actively transcribed euchromatin. Interestingly, clear differences were shown between the uracil enrichment patterns induced by the two drugs. In contrast in the non-treated cells, the low amount of uracils showed primarily heterochromatic enrichment [91]. In the case of spontaneous cytosine deamination, altered accessibility of DNA within differently packed chromatin structures (from heterochromatin to the actively transcribing genes), or in double-stranded/single-stranded segments [126], and the diverse protein bound states (nucleosomes, transcription factors, DNA manipulating enzymes, etc.) might also influence the local frequency of deamination reactions. Similarly, the genome-wide distribution of distinct UDGs and other repair enzymes including polymerases must have a strong impact on the actual genomic uracil patterns. Therefore, a statistical approach might provide higher descriptive value while avoiding misinterpretation of the data. Identifying genomic regions (or peaks) where uracil can occur with higher probability, and comparison of these segments to other known genomic features might shed light on the underlying cellular processes.

Author Contributions: Conceptualization, A.B. and B.G.V.; Funding acquisition, B.G.V.; Investigation, A.B., E.H. and H.L.P.; Visualization, A.B., E.H. and H.L.P.; Writing—original draft preparation, A.B., and B.G.V.; Writing—review and editing, A.B., E.H., H.L.P., and B.G.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Research, Development and Innovation Office of Hungary, grant numbers K 119493, NKP-2018-1.2.1-NKP-2018-00005, and the BME-Biotechnology FIKP grant of EMMI (BME FIKP-BIO).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chon, J.; Field, M.S.; Stover, P.J. Deoxyuracil in DNA and disease: Genomic signal or managed situation? *DNA Repair* **2019**, *77*, 36–44. [[CrossRef](#)]
2. Krokan, H.E.; Drabløs, F.; Slupphaug, G. Uracil in DNA—occurrence, consequences and repair. *Oncogene* **2002**, *21*, 8935–8948. [[CrossRef](#)]
3. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **1993**, *362*, 709–715. [[CrossRef](#)] [[PubMed](#)]
4. Muramatsu, M.; Kinoshita, K.; Fagarasan, S.; Yamada, S.; Shinkai, Y.; Honjo, T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **2000**, *102*, 553–563. [[CrossRef](#)]
5. 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*, 99–103. [[CrossRef](#)] [[PubMed](#)]
6. Neuberger, M.S.; Harris, R.S.; Di Noia, J.; Petersen-Mahrt, S.K. Immunity through DNA deamination. *Trends Biochem. Sci.* **2003**, *28*, 305–312. [[CrossRef](#)]
7. Beale, R.C.L.; Petersen-Mahrt, S.K.; Watt, I.N.; Harris, R.S.; Rada, C.; Neuberger, M.S. Comparison of the Differential Context-dependence of DNA Deamination by APOBEC Enzymes: Correlation with Mutation Spectra In Vivo. *J. Mol. Biol.* **2004**, *337*, 585–596. [[CrossRef](#)] [[PubMed](#)]
8. Honjo, T.; Muramatsu, M.; Fagarasan, S. Aid: How does it aid antibody diversity? *Immunity* **2004**, *20*, 659–668. [[CrossRef](#)] [[PubMed](#)]
9. Harris, R.S.; Dudley, J.P. APOBECs and virus restriction. *Virology* **2015**, *479–480*, 131–145. [[CrossRef](#)] [[PubMed](#)]
10. Siriwardena, S.U.; Chen, K.; Bhagwat, A.S. Functions and Malfunctions of Mammalian DNA-Cytosine Deaminases. *Chem. Rev.* **2016**, *116*, 12688–12710. [[CrossRef](#)]
11. Maul, R.W.; Gearhart, P.J. Refining the Neuberger model: Uracil processing by activated B cells. *Eur. J. Immunol.* **2014**, *44*, 1913–1916. [[CrossRef](#)]
12. Caradonna, S.J.; Cheng, Y.C. The role of deoxyuridine triphosphate nucleotidohydrolase, uracil-DNA glycosylase, and DNA polymerase α in the metabolism of FUdR in human tumor cells. *Mol. Pharmacol.* **1980**, *18*, 513–520.
13. Tanaka, M.; Yoshida, S.; Saneyoshi, M.; Yamaguchi, T. Utilization of 5-fluoro-2'-deoxyuridine triphosphate and 5-fluoro-2'-deoxycytidine triphosphate in DNA synthesis by DNA polymerases alpha and beta from calf thymus. *Cancer Res.* **1981**, *41*, 4132–4135. [[PubMed](#)]
14. Hirmondó, R.; Szabó, J.E.; Nyíri, K.; Tarjányi, S.; Dobrotka, P.; Tóth, J.; Vértessy, B.G. Cross-species inhibition of dUTPase via the Staphylococcal StI protein perturbs dNTP pool and colony formation in Mycobacterium. *DNA Repair* **2015**, *30*. [[CrossRef](#)]
15. Muha, V.; Horváth, A.; Békési, A.; Pukáncsik, M.; Hodoscsek, B.; Merényi, G.; Róna, G.; Batki, J.; Kiss, I.; Jankovics, F.; et al. Uracil-containing DNA in Drosophila: Stability, stage-specific accumulation, and developmental involvement. *PLoS Genet* **2012**, *8*, e1002738. [[CrossRef](#)]
16. Andersen, S.; Heine, T.; Sneve, R.; König, I.; Krokan, H.E.; Epe, B.; Nilsen, H. Incorporation of dUMP into DNA is a major source of spontaneous DNA damage, while excision of uracil is not required for cytotoxicity of fluoropyrimidines in mouse embryonic fibroblasts. *Carcinogenesis* **2005**, *26*, 547–555. [[CrossRef](#)] [[PubMed](#)]
17. Xue, J.H.; Xu, G.F.; Gu, T.P.; Chen, G.D.; Han, B.B.; Xu, Z.M.; Bjørås, M.; Krokan, H.E.; Xu, G.L.; Du, Y.R. Uracil-DNA glycosylase UNG promotes tet-mediated DNA demethylation. *J. Biol. Chem.* **2016**, *291*. [[CrossRef](#)] [[PubMed](#)]
18. Serebrenik, A.A.; Starrett, G.J.; Leenen, S.; Jarvis, M.C.; Shaban, N.M.; Salamango, D.J.; Nilsen, H.; Brown, W.L.; Harris, R.S. The deaminase APOBEC3B triggers the death of cells lacking uracil DNA glycosylase. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 22158–22163. [[CrossRef](#)]
19. Chandra, V.; Bortnick, A.; Murre, C. AID targeting: Old mysteries and new challenges. *Trends Immunol.* **2015**, *36*, 527–535. [[CrossRef](#)]
20. Pavri, R.; Nussenzweig, M.C. AID targeting in antibody diversity. *Adv. Immunol.* **2011**, *110*, 1–26. [[CrossRef](#)] [[PubMed](#)]
21. Wang, J.H. The role of activation-induced deaminase in antibody diversification and genomic instability. *Immunol. Res.* **2013**, *55*, 287–297. [[CrossRef](#)]
22. Stenglein, M.D.; Burns, M.B.; Li, M.; Lengyel, J.; Harris, R.S. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat. Struct. Mol. Biol.* **2010**, *17*, 222–229. [[CrossRef](#)] [[PubMed](#)]
23. Law, E.K.; Levin-Klein, R.; Jarvis, M.C.; Kim, H.; Argyris, P.P.; Carpenter, M.A.; Starrett, G.J.; Temiz, N.A.; Larson, L.K.; Durfee, C.; et al. APOBEC3A catalyzes mutation and drives carcinogenesis in vivo. *J. Exp. Med.* **2020**. [[CrossRef](#)] [[PubMed](#)]
24. Burns, M.B.; Leonard, B.; Harris, R.S. APOBEC3B: Pathological consequences of an innate immune DNA mutator. *Biomed. J.* **2015**, *38*, 102–110. [[CrossRef](#)] [[PubMed](#)]
25. Nikkilä, J.; Kumar, R.; Campbell, J.; Brandsma, I.; Pemberton, H.N.; Wallberg, F.; Nagy, K.; Scheer, I.; Vértessy, B.G.; Serebrenik, A.A.; et al. Elevated APOBEC3B expression drives a kataeic-like mutation signature and replication stress-related therapeutic vulnerabilities in p53-defective cells. *Br. J. Cancer* **2017**, *117*, 113–123. [[CrossRef](#)]
26. Periyasamy, M.; Singh, A.K.; Gemma, C.; Kranjec, C.; Farzan, R.; Leach, D.A.; Navaratnam, N.; Pálinkás, H.L.; Vértessy, B.G.; Fenton, T.R.; et al. p53 controls expression of the DNA deaminase APOBEC3B to limit its potential mutagenic activity in cancer cells. *Nucleic Acids Res.* **2017**, *45*, 11056–11069. [[CrossRef](#)]
27. Horváth, A.; Vértessy, B.G. A one-step method for quantitative determination of uracil in DNA by real-time PCR. *Nucleic Acids Res.* **2010**, *38*, e196. [[CrossRef](#)]

28. Róna, G.; Scheer, I.; Nagy, K.; Pálinkás, H.L.; Tihanyi, G.; Borsos, M.; Békési, A.; Vértessy, B.G. Detection of uracil within DNA using a sensitive labeling method for in vitro and cellular applications. *Nucleic Acids Res.* **2016**, *44*, e28. [[CrossRef](#)] [[PubMed](#)]
29. Wyatt, M.D.; Wilson, D.M. Participation of DNA repair in the response to 5-fluorouracil. *Cell. Mol. Life Sci.* **2009**, *66*, 788–799. [[CrossRef](#)]
30. Visnes, T.; Doseeth, B.; Pettersen, H.S.; Hagen, L.; Sousa, M.M.L.; Akbari, M.; Otterlei, M.; Kavli, B.; Slupphaug, G.; Krokan, H.E. Uracil in DNA and its processing by different DNA glycosylases. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **2009**, *364*, 563–568. [[CrossRef](#)]
31. Krokan, H.E.; Otterlei, M.; Nilsen, H.; Kavli, B.; Skorpen, F.; Andersen, S.; Skjelbred, C.; Akbari, M.; Aas, P.A.; Slupphaug, G. Properties and functions of human uracil-DNA glycosylase from the UNG gene. *Prog. Nucleic Acid Res. Mol. Biol.* **2001**, *68*, 365–386.
32. Slupphaug, G.; Eftedal, I.; Kavli, B.; Bharati, S.; Helle, N.M.; Haug, T.; Levine, D.W.; Krokan, H.E. Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase. *Biochemistry* **1995**, *34*, 128–138. [[CrossRef](#)]
33. Kavli, B.; Sundheim, O.; Akbari, M.; Otterlei, M.; Nilsen, H.; Skorpen, F.; Aas, P.A.; Hagen, L.; Krokan, H.E.; Slupphaug, G. hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup. *J. Biol. Chem.* **2002**, *277*, 39926–39936. [[CrossRef](#)] [[PubMed](#)]
34. Caradonna, S.; Muller-weeks, S. The Nature of Enzymes Involved in Uracil-DNA Repair: Isoform Characteristics of Proteins Responsible for Nuclear and Mitochondrial Genomic Integrity. *Curr. Protein Pept. Sci.* **2005**, *2*. [[CrossRef](#)] [[PubMed](#)]
35. Wibley, J.E.A.; Waters, T.R.; Haushalter, K.; Verdine, G.L.; Pearl, L.H. Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. *Mol. Cell* **2003**, *11*. [[CrossRef](#)]
36. Masaoka, A.; Matsubara, M.; Hasegawa, R.; Tanaka, T.; Kurisu, S.; Terato, H.; Ohyama, Y.; Karino, N.; Matsuda, A.; Ide, H. Mammalian 5-formyluracil-DNA glycosylase. 2. Role of SMUG1 uracil-DNA glycosylase in repair of 5-formyluracil and other oxidized and deaminated base lesions. *Biochemistry* **2003**, *42*. [[CrossRef](#)]
37. Boorstein, R.J.; Cummings, A.; Marenstein, D.R.; Chan, M.K.; Ma, Y.; Neubert, T.A.; Brown, S.M.; Teebor, G.W. Definitive Identification of Mammalian 5-Hydroxymethyluracil DNA N-Glycosylase Activity as SMUG1. *J. Biol. Chem.* **2001**, *276*. [[CrossRef](#)]
38. Hashimoto, H. Structural and mutation studies of two DNA demethylation related glycosylases: MBD4 and TDG. *Biophys J.* **2014**, *106*, 63–68. [[CrossRef](#)]
39. Liu, P.; Burdzy, A.; Sowers, L.C. Substrate recognition by a family of uracil-DNA glycosylases: UNG, MUG, and TDG. *Chem. Res. Toxicol.* **2002**, *15*, 1001–1009. [[CrossRef](#)]
40. Gallinari, P.; Jiricny, J. A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. *Nature* **1996**, *383*. [[CrossRef](#)]
41. Cortellino, S.; Xu, J.; Sannai, M.; Moore, R.; Caretti, E.; Cigliano, A.; Le Coz, M.; Devarajan, K.; Wessels, A.; Soprano, D.; et al. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **2011**, *146*, 67–79. [[CrossRef](#)]
42. Cortázar, D.; Kunz, C.; Selfridge, J.; Lettieri, T.; Saito, Y.; MacDougall, E.; Wirz, A.; Schuermann, D.; Jacobs, A.L.; Siegrist, F.; et al. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **2011**, *470*, 419–423. [[CrossRef](#)] [[PubMed](#)]
43. Hassan, H.M.; Kolendowski, B.; Iovic, M.; Bose, K.; Dranse, H.J.; Sampaio, A.V.; Underhill, T.M.; Torchia, J. Regulation of Active DNA Demethylation through RAR-Mediated Recruitment of a TET/TDG Complex. *Cell Rep.* **2017**, *19*. [[CrossRef](#)] [[PubMed](#)]
44. Bellacosa, A.; Drohat, A.C. Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites. *DNA Repair* **2015**, *32*. [[CrossRef](#)]
45. Krokan, H.E.; Sætrum, P.; Aas, P.A.; Pettersen, H.S.; Kavli, B.; Slupphaug, G. Error-free versus mutagenic processing of genomic uracil-Relevance to cancer. *DNA Repair* **2014**, *19*. [[CrossRef](#)]
46. Krokan, H.E.; Bjørås, M. Base excision repair. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a012583. [[CrossRef](#)]
47. Wilson, D.M., III. *The Base Excision Repair Pathway*; World Scientific: Singapore, 2017.
48. Schanz, S.; Castor, D.; Fischer, F.; Jiricny, J. Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 5593–5598. [[CrossRef](#)]
49. Bellacosa, A. Functional interactions and signaling properties of mammalian DNA mismatch repair proteins. *Cell Death Differ.* **2001**, *8*, 1076–1092. [[CrossRef](#)] [[PubMed](#)]
50. Hu, J.; Li, W.; Adebali, O.; Yang, Y.; Oztas, O.; Selby, C.P.; Sancar, A. Genome-wide mapping of nucleotide excision repair with XR-seq. *Nat. Protoc.* **2019**. [[CrossRef](#)] [[PubMed](#)]
51. Weiner, K.X.B.; Weiner, R.S.; Maley, F.; Maley, G.F. Primary structure of human deoxycytidylate deaminase and overexpression of its functional protein in *Escherichia coli*. *J. Biol. Chem.* **1993**, *268*. [[CrossRef](#)]
52. Mancini, W.R.; Cheng, Y.C. Human deoxycytidylate deaminase: Substrate and regulator specificities and their chemotherapeutic implications. *Mol. Pharmacol.* **1983**, *23*, 159–164. [[PubMed](#)]
53. Vértessy, B.G.; Tóth, J. Keeping uracil out of DNA: Physiological role, structure and catalytic mechanism of dUTPases. *Acc. Chem. Res.* **2009**, *42*, 97–106. [[CrossRef](#)]
54. Islam, Z.; Gurevic, I.; Strutzenberg, T.S.; Ghosh, A.K.; Iqbal, T.; Kohen, A. Bacterial versus human thymidylate synthase: Kinetics and functionality. *PLoS ONE* **2018**, *13*, e0196506. [[CrossRef](#)] [[PubMed](#)]

55. Longley, D.B.; Boyer, J.; Allen, W.L.; Latif, T.; Ferguson, P.R.; Maxwell, P.J.; McDermott, U.; Lynch, M.; Harkin, D.P.; Johnston, P.G. The role of thymidylate synthase induction in modulating p53-regulated gene expression in response to 5-fluorouracil and antifolates. *Cancer Res.* **2002**, *62*, 2644–2649. [[PubMed](#)]
56. Chu, E.; Voeller, D.M.; Jones, K.L.; Takechi, T.; Maley, G.F.; Maley, F.; Segal, S.; Allegra, C.J. Identification of a thymidylate synthase ribonucleoprotein complex in human colon cancer cells. *Mol. Cell. Biol.* **1994**, *14*. [[CrossRef](#)] [[PubMed](#)]
57. Garg, D.; Beribisky, A.V.; Ponterini, G.; Ligabue, A.; Marverti, G.; Martello, A.; Paola Costi, M.; Sattler, M.; Wade, R.C. Translational repression of thymidylate synthase by targeting its mRNA. *Nucleic Acids Res.* **2013**, *41*. [[CrossRef](#)]
58. Brunn, N.D.; Dibrov, S.M.; Kao, M.B.; Ghassemian, M.; Hermann, T. Analysis of mRNA recognition by human thymidylate synthase. *Biosci. Rep.* **2014**, *34*. [[CrossRef](#)] [[PubMed](#)]
59. Kawazoe, A.; Takahari, D.; Keisho, C.; Nakamura, Y.; Ikeno, T.; Wakabayashi, M.; Nomura, S.; Tamura, H.; Fukutani, M.; Hirano, N.; et al. A multicenter phase II study of TAS-114 in combination with S-1 in patients with pretreated advanced gastric cancer (EPOC1604). *Gastric Cancer* **2020**, *1*, 3. [[CrossRef](#)]
60. Yano, W.; Yokogawa, T.; Wakasa, T.; Yamamura, K.; Fujioka, A.; Yoshisue, K.; Matsushima, E.; Miyahara, S.; Miyakoshi, H.; Taguchi, J.; et al. TAS-114, a first-in-class dual dUTPase/DPD inhibitor, demonstrates potential to improve therapeutic efficacy of fluoropyrimidine-based chemotherapy. *Mol. Cancer Ther.* **2018**, *17*. [[CrossRef](#)]
61. Yokogawa, T.; Yano, W.; Tsukioka, S.; Osada, A.; Wakasa, T.; Ueno, H.; Hoshino, T.; Yamamura, K.; Fujioka, A.; Fukuoka, M.; et al. dUTPase inhibition confers susceptibility to a thymidylate synthase inhibitor in DNA-repair-defective human cancer cells. *Cancer Sci.* **2021**, *112*, 422–432. [[CrossRef](#)]
62. Nyíri, K.; Harris, M.J.; Matejka, J.; Ozohanics, O.; Vékey, K.; Borysik, A.J.; Vértessy, B.G. HDX and native mass spectrometry reveals the different structural basis for interaction of the staphylococcal pathogenicity island repressor stl with dimeric and trimeric phage dUTPases. *Biomolecules* **2019**, *9*, 488. [[CrossRef](#)]
63. Surányi, É.V.; Hírmondó, R.; Nyíri, K.; Tarjányi, S.; Köhegyi, B.; Tóth, J.; Vértessy, B.G. Exploiting a phage-bacterium interaction system as a molecular switch to decipher macromolecular interactions in the living cell. *Viruses* **2018**, *10*, 168. [[CrossRef](#)]
64. Nyíri, K.; Mertens, H.D.T.; Tihanyi, B.; Nagy, G.N.; Kohegyi, B.; Matejka, J.; Harris, M.J.; Szabó, J.E.; Papp-Kádár, V.; Németh-Pongrácz, V.; et al. Structural model of human dUTPase in complex with a novel proteinaceous inhibitor. *Sci. Rep.* **2018**, *8*. [[CrossRef](#)]
65. Benedek, A.; Pölöskei, I.; Ozohanics, O.; Vékey, K.; Vértessy, B.G. The Stl repressor from *Staphylococcus aureus* is an efficient inhibitor of the eukaryotic fruitfly dUTPase. *FEBS Open Bio* **2018**, *8*. [[CrossRef](#)]
66. Papp-Kádár, V.; Szabó, J.E.; Nyíri, K.; Vértessy, B.G. In Vitro analysis of predicted DNA-binding sites for the Stl repressor of the *Staphylococcus aureus* SaPIBov1 pathogenicity island. *PLoS ONE* **2016**, *11*. [[CrossRef](#)] [[PubMed](#)]
67. Nyíri, K.; Kohegyi, B.; Micsonai, A.; Kardos, J.; Vértessy, B.G. Evidence-based structural model of the staphylococcal repressor protein: Separation of functions into different domains. *PLoS ONE* **2015**, *10*. [[CrossRef](#)] [[PubMed](#)]
68. Szabó, J.E.; Németh, V.; Papp-Kádár, V.; Nyíri, K.; Leveles, I.; Bendes, Á.; Zagyva, I.; Róna, G.; Pálinkás, H.L.; Besztercei, B.; et al. Highly potent dUTPase inhibition by a bacterial repressor protein reveals a novel mechanism for gene expression control. *Nucleic Acids Res.* **2014**, *42*. [[CrossRef](#)] [[PubMed](#)]
69. Merényi, G.; Kovári, J.; Tóth, J.; Takács, E.; Zagyva, I.; Erdei, A.; Vértessy, B.G. Cellular response to efficient dUTPase RNAi silencing in stable hela cell lines perturbs expression levels of genes involved in thymidylate metabolism. *Nucleosides Nucleotides Nucleic Acids* **2011**, *30*. [[CrossRef](#)]
70. Koehler, S.E.; Ladner, R.D. Small interfering RNA-mediated suppression of dUTPase sensitizes cancer cell lines to thymidylate synthase inhibition. *Mol. Pharmacol.* **2004**, *66*. [[CrossRef](#)]
71. Wilson, P.M.; Danenberg, P.V.; Johnston, P.G.; Lenz, H.-J.; Ladner, R.D. Standing the test of time: Targeting thymidylate biosynthesis in cancer therapy. *Nat. Rev. Clin. Oncol.* **2014**, *11*, 282–298. [[CrossRef](#)]
72. Li, W.; Sancar, A. Methodologies for detecting environmentally induced DNA damage and repair. *Environ. Mol. Mutagen.* **2020**, *61*, 664–679. [[CrossRef](#)]
73. Mingard, C.; Wu, J.; McKeague, M.; Sturla, S.J. Next-generation DNA damage sequencing. *Chem. Soc. Rev.* **2020**, *49*, 7354–7377. [[CrossRef](#)]
74. Galashevskaya, A.; Sarno, A.; Vågbo, 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* **2013**, *12*, 699–706. [[CrossRef](#)]
75. Blount, B.C.; Ames, B.N. Analysis of uracil in DNA by gas chromatography-mass spectrometry. *Anal. Biochem.* **1994**, *219*, 195–200. [[CrossRef](#)]
76. Mashiyama, S.T.; Courtemanche, C.; Elson-Schwab, I.; Crott, J.; Lee, B.L.; Ong, C.N.; Fenech, M.; Ames, B.N. Uracil in DNA, determined by an improved assay, is increased when deoxynucleosides are added to folate-deficient cultured human lymphocytes. *Anal. Biochem.* **2004**, *330*, 58–69. [[CrossRef](#)]
77. Atamna, H.; Cheung, I.; Ames, B.N. A method for detecting abasic sites in living cells: Age-dependent changes in base excision repair. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 686–691. [[CrossRef](#)] [[PubMed](#)]
78. Ren, J.; Ulvik, A.; Refsum, H.; Ueland, P.M. Uracil in human DNA from subjects with normal and impaired folate status as determined by high-performance liquid chromatography-tandem mass spectrometry. *Anal. Chem.* **2002**, *74*. [[CrossRef](#)]
79. Kubo, K.; Ide, H.; Wallace, S.S.; Kow, Y.W. A Novel, Sensitive, and Specific Assay for Abasic Sites, the Most Commonly Produced DNA Lesion. *Biochemistry* **1992**, *31*. [[CrossRef](#)] [[PubMed](#)]

80. Kow, Y.W.; Dare, A. Detection of abasic sites and oxidative DNA base damage using an ELISA-like assay. *Methods* **2000**, *22*. [[CrossRef](#)] [[PubMed](#)]
81. Lari, S.-U.; Chen, C.-Y.; Vértessy, B.G.; Morré, J.; Bennett, S.E. Quantitative determination of uracil residues in Escherichia coli DNA: Contribution of ung, dug, and dut genes to uracil avoidance. *DNA Repair* **2006**, *5*, 1407–1420. [[CrossRef](#)] [[PubMed](#)]
82. Wei, S.; Shalhout, S.; Ahn, Y.H.; Bhagwat, A.S. A versatile new tool to quantify abasic sites in DNA and inhibit base excision repair. *DNA Repair* **2015**, *27*. [[CrossRef](#)]
83. Wei, S.; Perera, M.L.W.; Sakhtemani, R.; Bhagwat, A.S. A novel class of chemicals that react with abasic sites in DNA and specifically kill B cell cancers. *PLoS ONE* **2017**, *12*. [[CrossRef](#)]
84. Siriwardena, S.U.; Perera, M.L.W.; Senevirathne, V.; Stewart, J.; Bhagwat, A.S. A Tumor-Promoting Phorbol Ester Causes a Large Increase in APOBEC3A Expression and a Moderate Increase in APOBEC3B Expression in a Normal Human Keratinocyte Cell Line without Increasing Genomic Uracils. *Mol. Cell. Biol.* **2018**, *39*. [[CrossRef](#)] [[PubMed](#)]
85. 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*, 118. [[CrossRef](#)] [[PubMed](#)]
86. 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*, 15037–15051. [[CrossRef](#)] [[PubMed](#)]
87. Nagelhus, T.A.; Haug, T.; Singh, K.K.; Keshav, K.F.; Skorpen, F.; Otterlei, M.; Bharati, S.; Lindmo, T.; Benichou, S.; Benarous, R.; et al. A sequence in the N-terminal region of human uracil-DNA glycosylase with homology to XPA interacts with the C-terminal part of the 34-kDa subunit of replication protein A. *J. Biol. Chem.* **1997**, *272*, 6561–6566. [[CrossRef](#)] [[PubMed](#)]
88. Torseth, K.; Doseth, B.; Hagen, L.; Olaisen, C.; Liabakk, N.-B.; Græsmann, H.; Durandy, A.; Otterlei, M.; Krokan, H.E.; Kavli, B.; et al. The UNG2 Arg88Cys variant abrogates RPA-mediated recruitment of UNG2 to single-stranded DNA. *DNA Repair* **2012**, *11*, 559–569. [[CrossRef](#)]
89. Green, D.A.; Deutsch, W.A. Direct determination of uracil in [32P,uracil-3H]poly(dA-dT) and bisulfite-treated phage PM2 DNA. *Anal. Biochem.* **1984**, *142*. [[CrossRef](#)]
90. Duthie, S.J.; McMillan, P. Uracil misincorporation in human DNA detected using single cell gel electrophoresis. *Carcinogenesis* **1997**, *18*. [[CrossRef](#)]
91. Palinkas, H.L.; Bekesi, A.; Rona, G.; Pongor, L.; Papp, G.; Tihanyi, G.; Holub, E.; Poti, A.; Gemma, C.; Ali, S.; et al. Genome-wide alterations of uracil distribution patterns in human DNA upon chemotherapeutic treatments. *eLife* **2020**, *9*, e60498. [[CrossRef](#)]
92. 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*, 8452–8463. [[CrossRef](#)]
93. Tu, 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*. [[CrossRef](#)] [[PubMed](#)]
94. 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*, 607–614. [[CrossRef](#)] [[PubMed](#)]
95. Jia, Q.; Zeng, H.; Tu, J.; Sun, L.; Cao, W.; Xie, W. Structural insights into an MsmUdgX mutant capable of both crosslinking and uracil excision capability. *DNA Repair* **2021**, *97*. [[CrossRef](#)]
96. Datta, M.; Aroli, S.; Karmakar, K.; Dutta, S.; Chakravorty, 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*, 38–43. [[CrossRef](#)]
97. Suspene, R.; Henry, M.; Guillot, S.; Wain-Hobson, S.; Vartanian, J.-P. Recovery of APOBEC3-edited human immunodeficiency virus G→A hypermutants by differential DNA denaturation PCR. *J. Gen. Virol.* **2005**, *86*, 125–129. [[CrossRef](#)] [[PubMed](#)]
98. Martomo, S.A.; Fu, D.; Yang, W.W.; Joshi, N.S.; Gearhart, P.J. Deoxyuridine Is Generated Preferentially in the Nontranscribed Strand of DNA from Cells Expressing Activation-Induced Cytidine Deaminase. *J. Immunol.* **2005**, *174*, 7787–7791. [[CrossRef](#)]
99. Maul, R.W.; Saribasak, H.; Martomo, S.A.; McClure, R.L.; Yang, W.; Vaisman, A.; Gramlich, H.S.; Schatz, D.G.; Woodgate, R.; Wilson, D.M.; et al. Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nat. Immunol.* **2011**, *12*, 70–76. [[CrossRef](#)]
100. Hansen, E.C.; Ransom, M.; Hesselberth, J.R.; Hosmane, N.N.; Capoferri, A.A.; Bruner, K.M.; Pollack, R.A.; Zhang, H.; Drummond, M.B.; Siliciano, J.M.; et al. Diverse fates of uracilated HIV-1 DNA during infection of myeloid lineage cells. *Elife* **2016**, *5*. [[CrossRef](#)]
101. Connolly, B.A. Recognition of deaminated bases by archaeal family-B DNA polymerases. *Biochem. Soc. Trans.* **2009**, *37*, 65–68. [[CrossRef](#)]
102. Fogg, M.J.; Pearl, L.H.; Connolly, B.A. Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat. Struct. Biol.* **2002**. [[CrossRef](#)] [[PubMed](#)]
103. Liu, Z.J.; Martínez Cuesta, S.; van Delft, P.; Balasubramanian, S. Sequencing abasic sites in DNA at single-nucleotide resolution. *Nat. Chem.* **2019**. [[CrossRef](#)]
104. Keszthelyi, A.; Daigaku, Y.; Ptasinska, K.; Miyabe, I.; Carr, A.M. Mapping ribonucleotides in genomic DNA and exploring replication dynamics by polymerase usage sequencing (Pu-seq). *Nat. Protoc.* **2015**. [[CrossRef](#)]
105. Rauluseviciute, I.; Drablos, F.; Rye, M.B. DNA methylation data by sequencing: Experimental approaches and recommendations for tools and pipelines for data analysis. *Clin. Epigenetics* **2019**, *11*, 193. [[CrossRef](#)]
106. Adar, S.; Hu, J.; Lieb, J.D.; Sancar, A. Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. *Proc. Natl. Acad. Sci. USA* **2016**. [[CrossRef](#)] [[PubMed](#)]

107. Hu, J.; Adar, S.; Selby, C.P.; Lieb, J.D.; Sancar, A. Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. *Genes Dev.* **2015**. [[CrossRef](#)]
108. Fleming, A.M.; Ding, Y.; Burrows, C.J. Sequencing DNA for the Oxidatively Modified Base 8-Oxo-7,8-Dihydroguanine. *Methods Enzymol.* **2017**, *591*, 187–210.
109. Wu, J.; McKeague, M.; Sturla, S.J. Nucleotide-Resolution Genome-Wide Mapping of Oxidative DNA Damage by Click-Code-Seq. *J. Am. Chem. Soc.* **2018**, *140*, 9783–9787. [[CrossRef](#)] [[PubMed](#)]
110. 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*, 1534–1542. [[CrossRef](#)]
111. Ransom, M.; Bryan, D.S.; Hesselberth, J.R. High-resolution mapping of modified DNA nucleobases using excision repair enzymes. In *Methods in Molecular Biology*; Humana Press Inc.: New York, NY, USA, 2018; Volume 1672, pp. 63–76.
112. 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*, 680–687. [[CrossRef](#)] [[PubMed](#)]
113. Riedl, J.; Ding, Y.; Fleming, A.M.; Burrows, C.J. Identification of DNA lesions using a third base pair for amplification and nanopore sequencing. *Nat. Commun.* **2015**. [[CrossRef](#)]
114. Riedl, J.; Fleming, A.M.; Burrows, C.J. Sequencing of DNA Lesions Facilitated by Site-Specific Excision via Base Excision Repair DNA Glycosylases Yielding Ligatable Gaps. *J. Am. Chem. Soc.* **2016**, *138*. [[CrossRef](#)]
115. Craig, J.M.; Laszlo, A.H.; Derrington, I.M.; Ross, B.C.; Brinkerhoff, H.; Nova, I.C.; Doering, K.; Tickman, B.I.; Svet, M.T.; Gundlach, J.H. Direct detection of unnatural DNA nucleotides dNaM and d5SICS using the MspA nanopore. *PLoS ONE* **2015**, *10*, e0143253. [[CrossRef](#)] [[PubMed](#)]
116. Agarwal, P.; Kudirka, R.; Albers, A.E.; Barfield, R.M.; De Hart, G.W.; Drake, P.M.; Jones, L.C.; Rabuka, D. Hydrazino-pictet-spengler ligation as a biocompatible method for the generation of stable protein conjugates. *Bioconjug. Chem.* **2013**, *24*. [[CrossRef](#)] [[PubMed](#)]
117. Van Luenen, H.G.A.M.; Farris, C.; Jan, S.; Genest, P.A.; Tripathi, P.; Velds, A.; Kerkhoven, R.M.; Nieuwland, M.; Haydock, A.; Ramasamy, G.; et al. Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. *Cell* **2012**, *150*. [[CrossRef](#)] [[PubMed](#)]
118. Reynolds, D.; Cliffe, L.; Förstner, K.U.; Hon, C.C.; Siegel, T.N.; Sabatini, R. Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in Leishmania major and Trypanosoma brucei. *Nucleic Acids Res.* **2014**, *42*. [[CrossRef](#)]
119. Bullard, W.; Lopes Da Rosa-Spiegler, J.; Liu, S.; Wang, Y.; Sabatini, R. Identification of the glucosyltransferase that converts hydroxymethyluracil to base J in the trypanosomatid genome. *J. Biol. Chem.* **2014**, *289*. [[CrossRef](#)]
120. Cao, B.; Wu, X.; Zhou, J.; Wu, H.; Liu, L.; Zhang, Q.; Demott, M.S.; Gu, C.; Wang, L.; You, D.; et al. Nick-seq for single-nucleotide resolution genomic maps of DNA modifications and damage. *Nucleic Acids Res.* **2020**, *48*, 6715–6725. [[CrossRef](#)]
121. Amemiya, H.M.; Kundaje, A.; Boyle, A.P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **2019**, *9*, 9354. [[CrossRef](#)]
122. McNulty, S.M.; Sullivan, B.A. Alpha satellite DNA biology: Finding function in the recesses of the genome. *Chromosom. Res.* **2018**, *26*, 115–138. [[CrossRef](#)]
123. Miga, K.H.; Newton, Y.; Jain, M.; Altemose, N.; Willard, H.F.; Kent, W.J. Centromere reference models for human chromosomes X and Y satellite arrays. *Genome Res.* **2014**, *24*, 697–707. [[CrossRef](#)]
124. Kawasaki, F.; Beraldi, D.; Hardisty, R.E.; McInroy, G.R.; van Delft, P.; Balasubramanian, S. Genome-wide mapping of 5-hydroxymethyluracil in the eukaryote parasite Leishmania. *Genome Biol.* **2017**, *18*, 23. [[CrossRef](#)]
125. Nabel, C.S.; Schutsky, E.K.; Kohli, R.M. Molecular targeting of mutagenic AID and APOBEC deaminases. *Cell Cycle* **2014**, *13*, 171–172. [[CrossRef](#)]
126. Bhagwat, A.S.; Hao, W.; Townes, J.P.; Lee, H.; Tang, H.; Foster, P.L. Strand-biased cytosine deamination at the replication fork causes cytosine to thymine mutations in Escherichia coli. *Proc. Natl. Acad. Sci. USA* **2016**, *113*. [[CrossRef](#)] [[PubMed](#)]