

# **Genomic uracil occurrence and the role of dUTPase in transgenic models**

## **PhD Thesis**

Hajnalka Laura Pálincás

Genome Metabolism and Repair Research Group

Institute of Enzymology

Research Centre for Natural Sciences

Doctoral School of Multidisciplinary Medical Science

Faculty of Medicine

University of Szeged

Supervisor: Beáta G. Vértessy PhD, DSc



Szeged

2020

## List of scientific publications

### This thesis is based on the following peer-reviewed publications:

- I.** Róna G, Scheer I, Nagy K, Pálincás HL, Tihanyi G, Borsos M, Békési A, Vértessy BG. Detection of uracil within DNA using a sensitive labeling method for *in vitro* and cellular applications. *Nucleic Acids Res.* 2016; 44(3):e28. doi: 10.1093/nar/gkv977. **IF: 9.202**
- II.** Pálincás HL, Rácz GA, Gál Z, Hoffmann OI, Tihanyi G, Róna G, Gócza E, Hiripi L, Vértessy BG. CRISPR/Cas9-Mediated Knock-Out of dUTPase in Mice Leads to Early Embryonic Lethality. *Biomolecules.* 2019; 9(4). pii: E136. doi: 10.3390/biom9040136. **IF: 4.694**

**Cumulative impact factor of publications directly related to the thesis: 13.896**

### Preprint manuscript directly related to the subject of the thesis:

- III.** Pálincás HL, Békési A, Róna G, Pongor L, Tihanyi G, Holub E, Póti Á, Gemma C, *et al.* Genome-wide alterations of uracil distribution patterns in human DNA upon chemotherapeutic treatments. *bioRxiv preprint server*, 2020.  
doi: <https://doi.org/10.1101/2020.03.04.976977>

### Additional peer-reviewed publications, not included in this thesis:

- IV.** Szabó JE, Németh V, Papp-Kádár V, Nyíri K, Leveles I, Bendes AÁ, Zagyva I, Róna G, Pálincás HL, *et al.* Highly potent dUTPase inhibition by a bacterial repressor protein reveals a novel mechanism for gene expression control. *Nucleic Acids Res.* 2014; 42(19):11912-20. doi: 10.1093/nar/gku882. **IF: 9.112**
- V.** Róna G, Pálincás HL, Borsos M, Horváth A, Scheer I, Benedek A, Nagy GN, Zagyva I, Vértessy BG. NLS copy-number variation governs efficiency of nuclear import--case study on dUTPases. *FEBS J.* 2014; 281(24):5463-78. doi: 10.1111/febs.13086. **IF: 4.001**
- VI.** Periyasamy M, Singh AK, Gemma C, Kranjec C, Farzan R, Leach DA, Navaratnam N, Pálincás HL, Vértessy BG, *et al.* p53 controls expression of the DNA deaminase APOBEC3B to limit its potential mutagenic activity in cancer cells. *Nucleic Acids Res.* 2017; 45(19):11056-11069. doi: 10.1093/nar/gkx721. **IF: 11.561**
- VII.** Molnár P, Marton L, Izrael R, Pálincás HL, Vértessy BG. Uracil moieties in *Plasmodium falciparum* genomic DNA. *FEBS Open Bio.* 2018; 8(11):1763-1772. doi: 10.1002/2211-5463.12458. **IF: 1.05**

**Cumulative impact factor of publications which are not included in the thesis: 25.724**

## The author's contribution

My PhD thesis is based on two peer-reviewed publications and on a manuscript in preparation listed on page 1. Most of the data obtained in my PhD thesis is the result of successful teamwork where I have been an active participant involved in all the design, performance and evaluation of scientific work. Some experiments closely related to the subject of the thesis were carried out in a collaborative work with the co-authors of the publications. Therefore, I have presented these experiments and results in 'first person plural' in the text. Below I enlist and hereby gratefully thank my colleagues who have contributed for the thesis related research.

Contributor	Contribution to the thesis
Dr. Gergely Róna	creation of 1x/3x-FLAG-ΔUNG(-DsRed) constructs performing dot-blot assays optimization of uracil-staining method using bacterial cells and plasmid aggregates dSTORM microscopy
Dr. Ildikó Scheer, Dr. Kinga Nagy	performing dot-blot assays
György Várady	fluorescence activated cell sorting
György Török	STED microscopy
Dr. Orsolya Hoffmann	microinjection of fertilized oocytes
Zoltán Gál	maintenance, breeding and genotyping of mouse models
Gergely Rác	optimization of genotyping PCR off-target analysis
Dr. Elen Gócza	analysis of dissected mouse embryos blastocyst outgrowth assay

<b>Acknowledgement</b> .....	4
<b>Abbreviations</b> .....	5
<b>1. Introduction</b> .....	8
1.1. Maintenance of genome integrity.....	8
1.2. Genomic uracil occurrence.....	8
1.3. Thymidylate biosynthesis and its importance in chemotherapy .....	9
1.4. The role of dUTPase in preventing genomic uracil accumulation.....	11
1.5. Pathways and enzymes specialized for processing uracil-containing DNA .....	14
1.6. The physiological significance of uracil occurrence in DNA .....	17
1.7. Methods to analyze genomic uracil content .....	19
1.8. The CRISPR/Cas9-mediated gene editing technique.....	20
<b>2. Aims</b> .....	21
<b>3. Materials and methods</b> .....	22
3.1. Materials and methods described in our peer-reviewed publications .....	22
3.2. Additional description related to unpublished experiments presented in section 4.2. ..	23
3.2.1. Plasmid constructs and cloning of the FLAG- $\Delta$ UNG-SNAP construct.....	23
3.2.2. Generation of UGI-expressing stable cell line and treatment of cells.....	23
3.2.3. Immunofluorescent staining and imaging of uracil residues within DNA.....	24
3.2.4. U-DNA immunoprecipitation .....	25
<b>4. Results</b> .....	26
4.1. Uracil-DNA detection applying catalytically inactive UNG constructs in <i>in vitro</i> and cellular studies .....	26
4.1.1. A dot-blot based quantitative assay for detection of genomic uracil <i>in vitro</i> .....	26
4.1.2. <i>In situ</i> cellular detection of uracil residues in DNA .....	29
4.2. Promising applications of $\Delta$ UNG-based U-DNA labeling techniques for analysis of uracil distribution pattern in human genomic DNA .....	31
4.2.1. Development of a novel U-DNA sensor construct appropriate for super-resolution imaging.....	31
4.2.2. Imaging of endogenous genomic uracils in human cells by confocal and super-resolution microscopy .....	32
4.2.3. Development of U-DNA-Seq method to unravel genome-wide distribution of uracil residues.....	35
4.3. A critical investigation of the physiological role of dUTPase in a mammalian model, using CRISPR/Cas9-mediated gene editing.....	37
4.3.1. CRISPR/Cas9-mediated knock-out of dUTPase in mice .....	37
4.3.2. Homozygote knock-out of dUTPase leads to early embryonic lethality .....	40
4.3.3. Alterations in embryonic development of <i>dut</i> mutants .....	42
<b>5. Discussion</b> .....	45
<b>6. Conclusion</b> .....	50
<b>References</b> .....	51
<b>Annex</b> .....	58



## Acknowledgement

First of all, I would like to express my deepest gratitude to my supervisor, Prof. Beáta G. Vértessy, for her motivating professional guidance during my PhD study and related research. I am truly grateful for given me the opportunity to be a member of her exceptional research group.

I wish to thank László Buday, György Keserű, and György Pokol, for allowing me to conduct my PhD research at the Institute of Enzymology in the Research Centre for Natural Sciences. I also owe gratitude to Prof. László Dux for supporting my doctoral studies at the Doctoral School of Multidisciplinary Medical Science at the Faculty of Medicine University of Szeged.

I would like to express most sincere appreciation to Dr. Gergely Róna for his expertise. I will forever be grateful for his continuous support of my work with both practical experimental advices and inspiring ideas, from the very beginning of my studies. I also wish to thank Dr. Angéla Békési for the motivating discussions and her helpful advices on my thesis work. I am thankful for Nóra Kucsma who introduced me to the basics of working in the virology lab.

Also I thank my colleagues, György Várady and György Török for their help with FACS experiments and STED imaging, respectively. I would also like to offer my regards to our collaborators, Dr. László Hiripi, Dr. Elen Gócza, Dr. Orsolya Hoffmann and Zoltán Gál, for their work at the Department of Animal Biotechnology in the National Agricultural Research and Innovation Centre.

I owe a lot to my fellow labmates in the Research Group of Genome Metabolism and Repair, for creating an exceptionally helpful, friendly and lively working environment. Particularly, I am honoured to Dr. Ildikó Scheer, Dr. Kinga Nagy and Gergely Rác for the great time we worked together, along with Gergely Tihanyi and Eszter Holub, former undergraduate students who assisted me in several experiments with great care and enthusiasm. I truly feel blessed to have such great coworkers.

Finally, I do not even have the words to thank my family for supporting me in every conceivable way throughout the challenging doctoral years. Thank all of you from the bottom of my heart.

## Abbreviations

ΔUNG	inactive (D154N, H268N), truncated (Δ84) UNG2
5FdUMP	5-fluoro-deoxyuridine monophosphate
5FdUR	5-fluoro-2'-deoxyuridine
5FU	5-fluorouracil
5hmU	5-hydroxymethyluracil
AID	activation-induced cytidine deaminase
AP	abasic (apurinic/apyrimidinic)
APE	AP-endonuclease
APOBEC	apolipoprotein B editing complex
ARP	aldehyde reactive probe
BER	base excision repair
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
Cas9	CRISPR-associated protein 9
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CSR	class-switch recombination
DAPI	4', 6-diamidino-2-phenylindole
dC	deoxycytidine
dCDP	deoxycytidine diphosphate
DCDT	dCMP deaminase
dCMP	deoxycytidine monophosphate
dCTP	deoxycytidine triphosphate
DCTPP1	dCTP pyrophosphatase 1
DIC	differential interference contrast
dNTP	deoxynucleoside triphosphates
dpc	<i>days post coitum</i>
DSB	double-stranded DNA break
dsDNA	double-stranded DNA
dSTORM	direct stochastic optical reconstruction microscopy
dT	deoxythymine
dTDP	deoxythymidine diphosphate
dTMP	deoxythymidine monophosphate, thymidylate
dTTP	deoxythymidine triphosphate
dUDP	deoxyuridine diphosphate

dUMP	deoxyuridine monophosphate
dUR	deoxyuridine
dUTP	deoxyuridine triphosphate
dUTPase	deoxyuridine triphosphatase
<i>E. coli</i>	<i>Escherichia coli</i>
FACS	fluorescence activated cell sorting
GC	gas chromatography
gDNA	genomic DNA
GFP	green fluorescent protein
gRNA	guide RNA
HCl	hydrogen chloride
HIV	human immunodeficiency virus
HPLC	high performance / high pressure liquid chromatography
HR	homologous recombination
ICM	inner cell mass
Ig	immunoglobulin
MBD4	methyl-CpG binding domain protein 4
mDut	mitochondrial dUTPase
MEF	mouse embryonic fibroblast
MMR	mismatch repair
mRNA	messenger RNA
MS	mass spectrometry
MTHF	5,10-methylene tetrahydrofolate
MTX	methotrexate
NDPK	nucleoside-diphosphate kinase
nDut	nuclear dUTPase
NGS	next-generation sequencing
NK	deoxynucleoside kinase
NLS	nuclear localization signal
NMPK	nucleoside monophosphate kinase
NT	non-treated
PAM	protospacer-adjacent motifs
PBS	phosphate-buffered saline
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction

PFA	paraformaldehyde
Pfu	<i>Pyrococcus furiosus</i>
qPCR	quantitative PCR
RNR	ribonucleotide reductase
RPA	replication protein A
RTX (TDX)	raltitrexed (tomudex)
<i>S. aureus</i>	<i>Staphylococcus aureus</i>
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SaPIs	<i>Staphylococcus aureus</i> pathogenicity islands
SE	standard errors of the mean
sgRNA	single-guide RNA
SHM	somatic hypermutation
SMRT	single molecule real-time sequencing
SMUG1	single-strand selective monofunctional uracil-DNA glycosylase 1
ssDNA	single-stranded DNA
STED	stimulated emission depletion
<i>T. brucei</i>	<i>Trypanosoma brucei</i>
TDG	thymine DNA glycosylase
TE	trophectoderm
TK	dTMP kinase
TK1	thymidine kinase
tracrRNA	trans-activating CRISPR RNA
TS	thymidylate synthase
UDG	DNA glycosylase
U-DNA	uracil-containing DNA
U-DNA-Seq	U-DNA immunoprecipitation coupled with NGS
UDP	uridine diphosphate
UGI	uracil-DNA glycosylase inhibitor protein
UNG	uracil-DNA N-glycosylase
WT	wild-type

# **1. Introduction**

## **1.1. Maintenance of genome integrity**

The faithful maintenance and transmission of genetic information is an essential task for all known living organisms. However, DNA, the carrier of hereditary information can be easily altered due to the chemical variability of its building blocks. Living organisms have developed several mechanisms to prevent the cytotoxic and mutagenic effects arising from the normal spontaneous chemical decay (such as reactive oxygen species, alkylating agents, etc.) or from external environmental effects (like chemicals, ultraviolet, ionizing radiation, etc.)<sup>1</sup>. Highly integrated cellular processes are involved in prevention, regulation or correction of DNA damages. These include a variety of regulatory enzymes that can sanitize and balance the cellular nucleotide pool. Still, if the amount of hazardous metabolites overloads the safeguarding performance of these enzymes, deleterious nucleotides can be incorporated into nascent DNA through replication. Moreover, incorporated nucleotides can also undergo enzymatic conversion resulting in abnormal base pairing. The potential distortion in the DNA structure due to the distinct chemical character of these nucleotides may affect the interaction with DNA binding proteins. Complex multi-component processes, mainly the base excision repair (BER) and the mismatch repair (MMR) machinery were specialized to recognize and repair such erroneous DNA lesions. In addition to these repair pathways, DNA damage can induce signaling responses such as cell cycle arrest, cellular senescence or even cell death that eventually helps avoiding a malignant transformation.

## **1.2. Genomic uracil occurrence**

Besides of the well-known Watson-Crick base pairs, DNA sequence might also contain a large variety of modified bases and DNA defects. In my research, I focus on uracil, one of the most frequently occurring non-canonical bases in DNA, implicated in several physiological processes. Uracil is usually classified as a nucleobase component of RNA, which is normally replaced by thymine in DNA. Still, two independent pathways may lead to the presence of uracil in DNA: non-mutagenic thymine replacing misincorporation and the potentially mutagenic hydrolytic cytosine deamination.

Most DNA polymerases cannot distinguish between deoxythymidine triphosphate (dTTP) and deoxyuridine triphosphate (dUTP) because thymine and uracil differ only in a single methyl group on the 5-position of the former. The incorporation of uracil instead of thymine does not

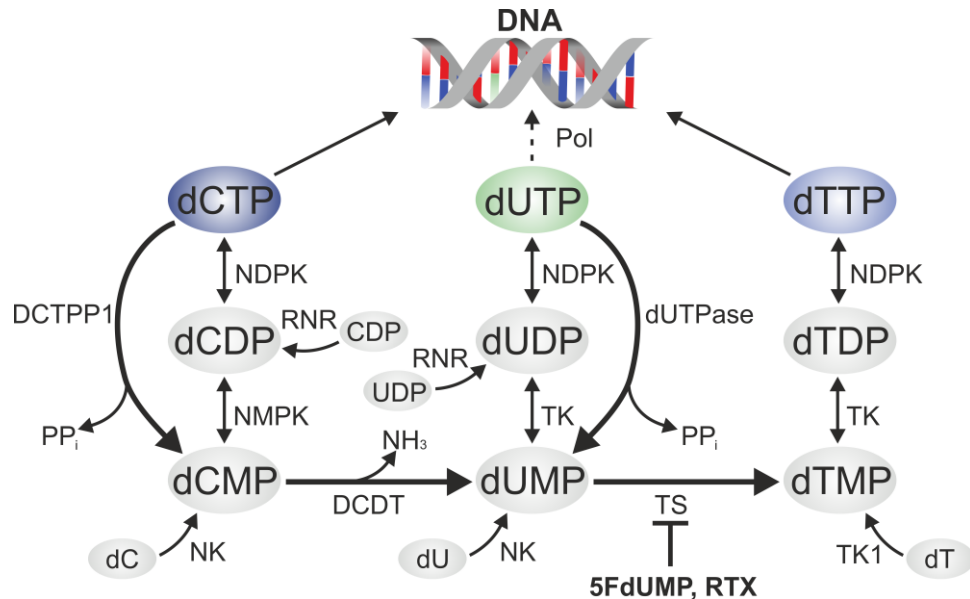
cause mutation, since uracil will form base pair with adenine in the same way as thymine. Uracil misincorporation depends on the cellular dUTP/dTTP ratio which is normally kept extremely low by two key enzymes: thymidylate synthase (TS) and deoxyuridine triphosphatase (dUTPase)<sup>2,3</sup>. Briefly, dUTPase can efficiently decrease dUTP level by hydrolyzing it to deoxyuridine monophosphate (dUMP), from which TS synthesises deoxythymidine monophosphate (dTMP), the precursor of dTTP. Increased thymine replacing uracil misincorporation events due to perturbation of the cellular nucleotide pool have to imply active DNA synthesis involved in either replication, transcription-coupled DNA repair or epigenetic reprogramming.

Under normal physiological conditions, oxidative deamination of cytosines within the context of the DNA double helix, is one of the most common spontaneous DNA-damaging processes. The reaction rate can be significantly increased by oxidative reagents (nitrogen oxides, reactive oxygen derivatives)<sup>1,4</sup>. Furthermore, the likelihood of deamination increases if the DNA is single-stranded, thus during replication and transcription. Overall, oxidative cytosine deamination occurs 70-200 times per day in a mammalian genome<sup>1</sup>. The emerging uracil:guanine mismatch can cause a stable point mutation if no DNA repair is performed until the next replication cycle, as opposite to the uracil adenine will be incorporated into the DNA instead of the original guanine (leading to C:G to T:A transition on one of the daughter DNAs)<sup>3</sup>. The mutation can alter the transcribed amino acid sequence of a particular protein, so it can even lead to complete loss of function. The facts that C to T transition mutation is the most common type of mutations in cancer<sup>5</sup>, and that the most common mutational signature in human cancers has been assigned to enzymatic cytosine deamination<sup>6</sup>, indicate the outstanding importance of eliminating uracil residues base paired with guanine from DNA.

### **1.3. Thymidylate biosynthesis and its importance in chemotherapy**

Thymidylate (deoxythymidine monophosphate, dTMP) biosynthesis is essential in proliferating cells, as it is the only source of dTTP synthesis which is one of the precursors of DNA synthesis. As shown in Figure 1, dTTP biosynthesis exclusively relies on dTMP because there is no corresponding ribonucleotide (cf. UDP and CDP)<sup>7</sup>. First, thymidylate synthase (TS) makes dTMP from dUMP precursor by the irreversible transfer of a methyl group. Then dTMP is further processed into deoxythymidine diphosphate (dTDP) and dTTP by dTMP kinase (TK) and nucleoside-diphosphate kinase (NDPK), respectively. Nevertheless, dUMP can also be

phosphorylated to dUTP by mono- and di-phosphate kinases (TK and NDPK), accordingly this non-canonical deoxynucleotide may be also available for DNA synthesis.



**Figure 1.** Overview of the synthesis pathways of pyrimidine nucleotides (dCTP, dTTP) and deoxyuridine triphosphate (dUTP) in eukaryotes. The main steps of pyrimidine nucleotide synthetic pathways are summarized, including the step catalyzed by TS that is essential to the dTTP production. TS inhibitors are shown in bold: 5FdUMP as the metabolite of 5FU and 5FdUR, and raltitrexed (RTX). Abbreviations are the following: Pol: DNA polymerase, DCDT: dCMP deaminase, DCTPP1: dCTP pyrophosphatase 1, RNR: ribonucleotide reductase, NDPK: nucleoside-diphosphate kinase, NK: nucleoside kinase, NMPK: nucleoside monophosphate kinase, TK1: thymidine kinase 1, TK: dTMP kinase, TS: thymidylate synthase, dUTPase: deoxyuridine triphosphatase, PPi: pyrophosphate.

In eukaryotes dUMP can be *de novo* generated through two different pathways. On the one hand, dUTPase catalyze the hydrolysis of dUTP into dUMP and pyrophosphate (releasing a proton). dUTP is the phosphorylated product (by NDPKs) of deoxyuridine diphosphate (dUDP) synthesized from uridine diphosphate (UDP) by ribonucleotide reductase (RNR). On the other hand, dCMP deaminase (DCDT) also produces dUMP from deoxycytidine monophosphate (dCMP) arisen from either deoxycytidine (dC), deoxycytidine diphosphate (dCDP) or deoxycytidine triphosphate (dCTP) by the action of deoxynucleoside kinases (NKs), nucleoside monophosphate kinases (NMPKs) and dCTP pyrophosphatase 1 (DCTPP1), respectively (Figure 1). Reducing the DCDT function in mammalian cell lines results in a drastic imbalance in deoxynucleoside triphosphate (dNTP) ratios, with an increased cellular dCTP and decreased dTTP level <sup>8</sup>; despite the existence of the so-called salvage pathway in eukaryotic cells that can

eventually generate dTMP from deoxythymine (dT) involving the action of thymidine kinase (TK1) <sup>9</sup>.

Based on this, the bottleneck of thymidylate synthesis seems to be the dUMP, whose methylation is catalyzed by TS into dTMP in the presence of the reducing cofactor, 5,10-methylene tetrahydrofolate (MTHF) <sup>3</sup>. Since maintenance of proper dNTP pool and thereby genomic integrity are essential for cellular function and survival especially in actively replicating cells such as cancer cells, TS is a common target for chemotherapeutic agents <sup>10</sup>.

Antifolate derivatives (e.g. methotrexate (MTX), aminopterin) inhibit the folate cycle required for methyl transfer from MTHF indirectly impeding thymidylate biosynthesis. Decrease in dTTP and increase in dUTP levels were observed when cells were treated with these drugs <sup>11,12</sup>. Other antifolates such as raltitrexed (RTX or also called tomudex, TDX) or pemetrexed act directly by forming complex with TS, resulting in the same effect on dTTP and dUTP levels <sup>12,13</sup>.

Fluoropyrimidines (5-fluorouracil (5FU) and 5-fluoro-2'-deoxyuridine (5FdUR)) belong to the oldest and most-effective chemotherapeutics that are commonly used today as well in anticancer therapies (capecitabine, tegafur, etc.). The exact mechanism of toxicity is not yet clear but they have a complex mode of action involving: i) irreversible inhibition of the enzyme TS by their metabolite 5-fluoro-dUMP (5FdUMP) and subsequent nucleotide pool imbalance (depletion of dTTP levels and elevated dUTP/dTTP ratio); and ii) misincorporation of dUMP and 5FdUMP into DNA by polymerases (from dUTP and 5-fluoro-dUTP, respectively) <sup>13</sup>. However, 5FU molecules are also able to perturb RNA synthesis and act primarily by misincorporation of 5-fluoro-UMP into RNA <sup>14,15</sup>. Treatment of 5FU and 5FdUR has a similar effect on mammalian cell lines as thymidine starvation in TS-deficient cells <sup>16,17</sup>. In addition, the increment of cellular dUTP level has been observed after 5FU and 5FdUR treatment in many cases <sup>13</sup>.

#### **1.4. The role of dUTPase in preventing genomic uracil accumulation**

Since dUMP has been shown to be a key intermediate molecule for thymidylate biosynthesis, dUTPase has remarkable function prior to the action of TS enzyme. As seen in Figure 1, dUTPase produces dUMP, the substrate for TS and dTTP synthesis, meanwhile eliminates dUTP therefore efficiently regulates dUTP/dTTP ratio and prevents undesirable genomic uracil accumulation.



Most eukaryotic organisms have two isoforms of dUTPase arising from alternative splicing of the mRNAs due to different promoter and 5' exon usage<sup>18</sup>. dUTPase isoforms in vertebrates including *Mus musculus* and *Homo sapiens*, are translocated into the nucleus (nDut) or into the mitochondria (mDut). The N-terminal of both isoforms contains the classical nuclear localization signal (NLS) sequence consisting of basic amino acids. However, the mitochondrial isoform also has a mitochondrial leader sequence ensuring strictly mitochondrial localization. The expression of mDut is constitutive, whereas nDut has a tightly regulated expression profile depending on cell cycle phase<sup>19</sup>. The fact that the expression of the nuclear isoform has been linked to S-phase, corroborates that dUTPase function is mainly required to control dUTP/dTTP levels during DNA replication in proliferating cells.

The enzyme dUTPase was found to be essential in *Escherichia coli* (*E. coli*)<sup>20</sup>, *Saccharomyces cerevisiae* (*S. cerevisiae*)<sup>21</sup>, *Caenorhabditis elegans* (*C. elegans*)<sup>22</sup>, *Trypanosoma brucei* (*T. brucei*)<sup>23</sup>, *Arabidopsis thaliana*<sup>24</sup> and *Mycobacterium smegmatis*<sup>25</sup>, where the deficiency of dUTPase led to cell death. However, thymidine supplement could compensate for the consequences of dUTPase depletion in many cases (in *E. coli*, *S. cerevisiae*, *C. elegans* and *T. brucei*, as well). Additionally, silencing of dUTPase resulted in a lethal phenotype in early pupal stages of *Drosophila melanogaster*, whereas under physiological conditions, larvae, pupae and imago maintain greatly elevated uracil content of DNA<sup>26,27</sup>. Importantly, the simultaneous depletion of the uracil-DNA N-glycosylase, Ung1 has been shown to rescue the lack of dUTPase in *E. coli*<sup>28</sup>, *S. cerevisiae*<sup>29</sup> and *C. elegans*<sup>30</sup>, and these double mutant organisms incorporate uracil into their DNA with a high frequency. Highlighting the potential of this phenomenon, Kunkel used the artificial *E. coli* strain for his site-directed mutagenesis method<sup>31</sup>. In contrast to the essentiality of dUTPase in eukaryotes, it has been also shown that several prokaryotic genomes naturally lack dUTPase, based on bioinformatic analysis of bacterial and archaeal genomes<sup>32</sup>. Diverse genetic distribution of proteins involved in uracil-DNA metabolism was observed and will be further discussed in detail in section 1.6. Recently, silencing of dUTPase in planarians was reported to cause regeneration failure and death in adult animals possibly due to genomic DNA (gDNA) fragmentation. Dual treatment with 5FU resulted in more DNA breaks and earlier lethality<sup>33</sup>.

Recently, a genetic study has been reported where a monogenic syndrome with diabetes and bone marrow failure could be associated with a missense mutation in dUTPase gene<sup>34</sup>. However, this clinical aspect was found to be extremely rare in patients known to date. Apparent

lack of genetic polymorphisms for dUTPase in the human population implies that deficiency may lead to severe consequences which are incompatible with life.

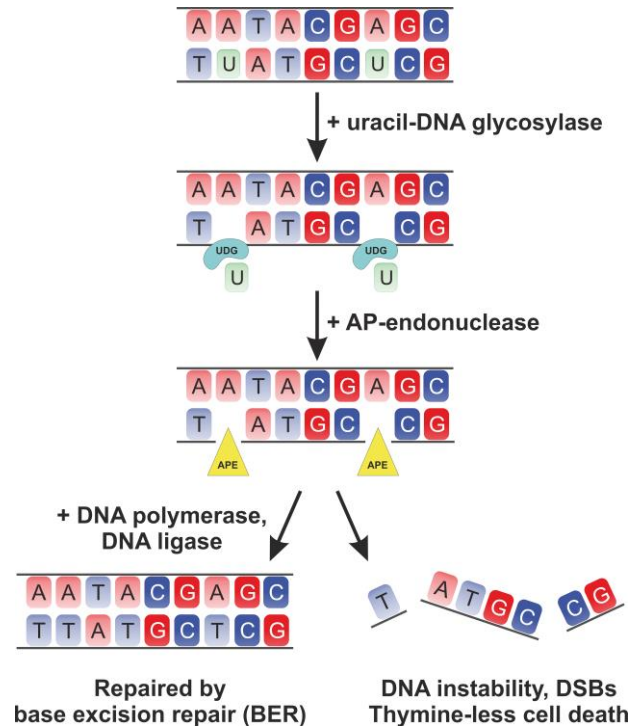
In mammalian cell line experiments, dUTPase disruption was only achieved so far by RNA silencing where the remaining amount of dUTPase could be sufficient for the viability of cells, particularly in the presence of thymidine in the cell culture medium, inducing the salvage pathway. However, several studies have demonstrated that the expression level of dUTPase significantly affects the response to chemotherapy targeting TS. On the one hand, overexpression of dUTPase induced resistance to 5FdUR in human cancer cells<sup>35,36</sup>. On the other hand, silencing or inhibition of dUTPase increased responses to TS inhibitors leading to dUTP pool increment, replication defects and cytotoxicity<sup>14,37,38,39,40</sup>. These findings may be explained by the fact that 5FdUTP also acts as a substrate for dUTPase<sup>40</sup>. A retrospective study including tumor specimens has shown that low levels of nuclear dUTPase protein expression is associated with higher response to 5FU-based treatment implicating greater overall survival, while high nuclear dUTPase expression indicates contrary prognosis with tumor resistance<sup>41</sup> and metastatic ability<sup>42</sup>. The combination of reduced dUTPase and high RNR subunit R2 expression level was also found to be associated with poor prognostic signal in colorectal and breast cancer patients. This study suggested that increased levels of dUTPase may compensate for the RNR-induced genomic stress through prevention of dUTP misincorporation into DNA, subsequently chromosomal instability and tumor evolution driving forces<sup>43</sup>. It has been also reported that the increased expression of dUTPase can reduce the cytotoxicity of TS inhibitors with other mode of action as well<sup>11,44</sup>. These data highlight the importance of dUTPase as a potential target in combination with other drugs to improve antitumor efficacy of commonly used nucleobase- and nucleoside-analogues<sup>45</sup>. Notably a novel small molecule dUTPase inhibitor, TAS-114 was successfully used to enhance the cytotoxicity of fluoropyrimidines in cancer cells and attained to the phase of clinical tests<sup>46</sup>. In addition, another potential route inhibiting dUTPase action was recently discovered in *Staphylococcus aureus* (*S. aureus*)<sup>47</sup>. A bacterial repressor protein, called Stl was reported to form a competitive, slow and tight binding with dUTPase from helper phages<sup>48</sup>. This specific interaction indicates a novel, special role of dUTPase in the mobilization of phage-related pathogenicity islands (SaPIs) or other mobile genetic elements. My colleagues found that Stl is a species-specific inhibitor of the different dUTPases, however this inhibitory effect is highly dependent on the actual dUTPase partner<sup>49</sup>. Investigation of the interaction between Stl and *E. coli* dUTPase revealed critical

structural differences<sup>50</sup> which may provide a promising basis for the development of specific dUTPase inhibitor peptides or peptide-mimicking small molecules in the future.

Moreover, inhibition of dUTPase could be also a relevant strategy to fight infectious diseases such as malaria and tuberculosis. In their pathogens (*Plasmodium falciparum* and *Mycobacterium tuberculosis*), DCDT and TK1 are absent while cytosine deamination occurs at the dCTP level catalyzed by dCTP deaminases. Therefore, the biosynthesis of dTMP relies exclusively on dUTPase activity highlighting a promising target<sup>3</sup>.

### 1.5. Pathways and enzymes specialized for processing uracil-containing DNA

Since most of the DNA polymerases cannot distinguish between dTTP and dUTP, they incorporate dTMP, dUMP and eventually 5FdUMP with similar efficacy. The nucleotide imbalance can lead to misincorporation of uracil and potentially fluoro-uracil into the DNA<sup>51</sup>. These DNA lesions are primarily subjected for post-replication base excision DNA repair (BER) and secondly the mismatch repair (MMR) pathway. Uracil-DNA glycosylases (UDGs) are responsible for the specific recognition of the non-canonical deoxynucleotides in DNA and they cleave the  $\beta$ -N-glycosidic bond between deoxyribose and the erroneous base<sup>52,53</sup> (see Figure 2). The created abasic (apurinic/apyrimidinic (AP)) site is recognized by the AP-endonuclease (APE) which incises the phosphodiester bond of DNA. After the removal of the remaining sugar fragment, single-stranded DNA undergoes gap filling by a DNA polymerase and sealing of the nick by a DNA ligase. However, under either high dUTP/dTTP ratio or total thymine deprivation, uracils presumably will be reincorporated during repair synthesis. Transformation of uracil-excision repair into a hyperactive futile cycle may lead to numerous AP sites, single- and double-stranded DNA breaks (DSBs), stalled replication forks and other potentially yet undiscovered consequences which trigger the so-called thymine-less cell death<sup>10,54,55</sup>. MMR machinery was also specialized to recognize DNA distortions and contribute to cytotoxicity induced by TS inhibitors based on the finding that inactivation of MMR (e.g. by the depletion of MLH1 enzyme) led to resistance to the cytotoxic effects of these agents<sup>56</sup>. After 5FdUR treatment of human cancer cells, it has been also shown that 5-fluorouracils form base pairs preferentially with guanines in MMR-deficient cells but not in MMR-proficient cells suggesting a significant role of MMR in the removal of 5FU<sup>57,58</sup>. Finally, activation of homologous recombination (HR) upon DSBs can be the last resort to repair these lesions promoting cell survival<sup>59,60</sup>.



**Figure 2.** Removal of uracil from genomic DNA by BER pathway. Initially uracil-DNA glycosylase recognizes and excises the non-canonical base from DNA. Then, the AP-endonuclease creates an abasic site and the remaining sugar molecule is eliminated. Finally, DNA polymerase incorporates the correct dNTP corresponding to the complement strand of DNA, and DNA ligase joins the free ends. Nevertheless, dUTP-enriched milieu can perturb the repair process leading to DNA instability, double-stranded DNA breaks and eventually thymine-less cell death.

Recognition of genomic uracil is carried out by the representatives of the UDG superfamily. These monofunctional glycosylases can be found in archaea, eubacteria, eukaryotes and large DNA viruses (herpes and pox) <sup>61</sup>. Mammals possess several UDGs, among them UNG, SMUG1, TDG, MBD4 are well characterized in the literature. The uracil-DNA N-glycosylase (UNG) protein is the most active enzyme in the removal of uracil from DNA in human cells <sup>15,62,63</sup>. Two different isoforms of UNG can be transcribed from the *ung* gene by alternative promoter usage leading to the mitochondrial UNG1 and nuclear UNG2, similarly to the case of human dUTPase isoforms. Both UNG isoforms can specifically excise uracil from both double-stranded (dsDNA) and single-stranded DNA (ssDNA) regardless to its origin i.e. cytosine deamination or thymine replacing misincorporation but with the preference ssU > U:G >> U:A <sup>61,62,63</sup>. In addition to its high uracil specificity, UNG is able to cleave to a lesser extent other non-cannonical bases modified at position 5 such as 5FU, 5-hydroxyuracil and 5-carboxyuracil <sup>64</sup>, or formed from cytosine oxidation such as isodialuric acid and alloxan <sup>65</sup>. Furthermore, it has been also shown *in vitro* that UNG is capable of binding AP sites,

but with a lower affinity as compared to uracil-containing DNA (U-DNA) <sup>66,67</sup>. According to the role of UNG2 in replication associated repair, its expression level is cell cycle regulated and the highest in S-phase <sup>63,68</sup>. Similarly to the nuclear isoform of dUTPase. Other proteins, namely proliferating cell nuclear antigen (PCNA) and replication protein A (RPA) involved in long patch BER were also demonstrated to interact and colocalize with UNG2 in replication foci <sup>68</sup>.

Several research groups have demonstrated the importance of UNG2 as the major replication-associated uracil-DNA glycosylase, and investigated the effects of its disruption for fluoropyrimidine-induced cytotoxicity. Surprisingly, some of these researchers found that depletion of UNG did not affect the sensitivity of TS inhibitors <sup>15,69,70,71</sup> or even increased resistance <sup>72</sup>. In contrast to this, recent studies demonstrated that 5FdUR treatment led to incorporation of uracil and 5FU into DNA resulting in a significantly enhanced cytotoxic response in UNG depleted, colon and ovarian cancer cells <sup>59,73</sup>. Furthermore, cell cycle arrest was also induced at late G1 and early S-phase indicating replication fork collapse and greatly enhanced DNA damages in the absence of UNG <sup>73</sup>. These ambiguous and contradictory results were reported studying a variety of cell lines with different genetic background which suggests a cell-type dependent mechanism of the commonly used anticancer agents.

Besides UNG, single-strand selective monofunctional uracil-DNA glycosylase 1 (SMUG1) is also an important member involved in the initiation of BER pathway and is more distributed in bacteria and eukaryotes than previously thought <sup>66</sup>. In contrast to its name, SMUG1 is also active on double-stranded uracil-DNA, although to a lesser extent <sup>63</sup>. In addition to uracil, it can also recognize and excise 5FU, 5-formyluracil, 5-hydroxymethyluracil (5hmU) and 3,N(4)-ethenocytosine. Increasing evidence suggests that SMUG1 can act as a broad specificity backup for UNG2 in human <sup>74</sup> and mouse cells <sup>75</sup>, especially in non-proliferating or proliferating cells outside the S-phase <sup>76</sup>.

The two last, minor UDGs identified in human cells, thymine DNA glycosylase (TDG) and methyl-CpG binding domain protein 4 (MBD4), are both mismatch-specific with the need of dsDNA substrates. TDG excises uracil and thymine from U:G and T:G mismatches, as well as 3,N(4)-ethenocytosine and 5FU, but with a strong preference for U:G pairs <sup>63,76</sup>. Similar to SMUG1 and UNG, it binds to AP sites, but in contrast, the highest expression of TDG is linked to G1-phase <sup>77</sup>. Compared to TDG, MBD4 has very similar substrate specificity and corrects U:G and T:G mismatches, generated by deamination of CpG and methylated CpG dinucleotides, respectively <sup>63,76</sup>. In addition, MBD4 presumably has a role in the MMR mechanism <sup>76</sup>. Interestingly, TDG and maybe MBD4 as well to some extent, has particular

importance in active DNA demethylation, thereby in epigenetic and transcriptional regulation<sup>78,79</sup>. UDGs including those already mentioned, are classified into families depending on their substrate specificity. Lately, a special UDG, called UdgX was discovered from *Mycobacterium smegmatis* and other organisms<sup>80</sup>. This enzyme is able to form a tight complex with uracil but surprisingly uracil excision is not detectable implying another non-repair function in U-DNA processing.

### 1.6. The physiological significance of uracil occurrence in DNA

However, genomic uracil occurrence was used to be considered as a mistake that must be repaired, accumulating evidence has been presented that uracils also appear in DNA under normal physiological conditions and may have key impact in highly diverse processes. To our knowledge, either maintenance of uracil enriched DNA, or the induction of uracil-excision repair at well-defined uracil sites can serve important cellular functions.

Long-term stabilization of uracilated DNA can be achieved by the efficient inactivation or absence of UDG enzymes involved in the main uracil-excision repair system. As already mentioned in section 1.4., screening numerous bacterial and archaeal sequence sets for the presence or absence of key enzymes involved in U-DNA metabolism has revealed a more diverse picture than previously thought<sup>32</sup>. It has been shown that a number of prokaryotes lack both genes for dUTPase and UNG (*dut*-, *ung*- genotype) which is expected to result in uracil-containing DNA<sup>32</sup>. However, even more numerous genomes lack the dUTPase gene but possess the UNG gene (*dut*-, *ung*+ genotype) threatening genetic stability<sup>32</sup>. Different survival strategies could be evolved such as yet undescribed proteins to supply dUTPase activity or UNG inhibition; exogenous supplement of dUTPases or UNG inhibitory proteins from prophages<sup>32</sup>.

Such specific protein inhibitors of UNG were identified in *Bacillus subtilis* bacteriophages PBS1/PBS2 and  $\Phi$ 29 encoding UGI<sup>81</sup> and p56<sup>82</sup>, respectively. Intriguingly, genome of these specific bacteriophages contains almost completely uracil instead of thymine, similarly to the *Yersinia enterocolitica*  $\Phi$ R1–37<sup>83</sup> and the general *Staphylococcace* S6<sup>84</sup> phages where potential UNG inhibitors are still undiscovered. Expression of UDG inhibitor proteins in the infected cells presumably serves as a defense mechanism to prevent degradation of uracil-enriched phage DNA caused by UDG activity, thereby facilitating viral DNA replication<sup>82,85</sup>. Notably all the members of family I UDGs except for poxvirus UNG are inhibited by the well-characterized UGI, which can mimic DNA backbone interactions by targeting the DNA binding

surface of UNG<sup>86</sup>. Not only phages have their own UNG inhibitor, but a similar DNA mimic protein was also identified in the genome of *S. aureus* itself, termed SAUGI<sup>87</sup>, encoded on a mobile genetic element. Interestingly, *S. aureus* does not have its own endogenous dUTPase, although dUTPase can be present from helper phages leading to sanitization of dUTP pool and modulation of horizontal transfer of SaPI *via* the Stl:dUTPase interaction<sup>48</sup> mentioned above in section 1.4. Parallel spreading of factors (dUTPase and UGI) involved in U-DNA metabolism with virulence elements by horizontal gene transfer may also have a prominent regulatory function determining the fate of both phages and host bacterial cells<sup>32</sup>. Another striking example is human immunodeficiency virus (HIV) which, like other primate lentiviruses, does not encode its own dUTPase or UNG, unlike other viruses. Following HIV infection, its reverse transcribed DNA is found to be highly uracilated<sup>88</sup>, consistently with the uracil-rich cellular environment of non-dividing immune cells, due to the low expression level of human dUTPase and UNG2<sup>88,89</sup>. This elevated uracil content of HIV DNA was suggested to protect the viral genome against suicidal autointegration, thereby promote favorable integration process into chromosomal DNA and increase viral infectivity<sup>88</sup>. Nevertheless, innate immunity pathway against viral integration was proved to be fully dependent on the presence of active UNG2 in the host cell, which may degrade viral U-DNA before its integration into the genome<sup>90</sup>.

Surprisingly, among eukaryotes there are also organisms or tissues lacking dUTPase or UNG enzymes. The most striking known example is the case of Holometabola insects which do not encode UNG. As mentioned in section 1.4., high genomic uracil level (200–2000 uracil/million bases) was found in larvae, pupae and imago of *Drosophila melanogaster*<sup>26</sup>. In these tissues uracil content of DNA correlated inversely with dUTPase expression. These findings suggested that uracil-DNA can be well tolerated and has potential function in a specific developmental period of the fruit fly. Nevertheless strict regulation of the cellular dUTP/dTTP ratio by dUTPase is required for early embryonic development as well as during the metamorphosis<sup>26,27</sup>.

Besides the cases presented so far, where I focused on various examples involving viability with increased U-DNA level, temporary uracil appearance in DNA at specific sites can be an efficient tool in mutagenesis to create gene diversity. The naturally occurring precedent is the somatic hypermutation (SHM) leading to immunoglobulin (Ig) gene diversification during antibody generation in B-cells. Uracil arisen by the performance of activation-induced cytosine deaminase (AID) is a normal intermediate in the variable regions of the Ig genes as part of not only SHM, but also of gene conversion and class-switch recombination (CSR) processes in adaptive immunity<sup>63,76,91</sup>. AID-triggered deamination of cytosine leads to C:G to T:A transition

mutation upon replication if left unrepaired. However, the activated BER and MMR pathways also contribute to the mutagenic effect, since subsequent to removal of uracil by UNG, altered bases can be incorporated by the action of translesion bypass polymerases<sup>92</sup>. DNA strand breaks can also be induced leading to exchange effector regions (in CSR). Consequently, AID generated uracils in specific regions of Ig loci serve as mutational and recombination hotspots in SHM and CSR, respectively. Interestingly, a recent study suggested that not only nuclear UNG2, but also mitochondrial UNG1 isoform may involved in CSR<sup>93</sup>. The proposed function of AID and UNG in Ig gene diversification was consistent with the observations reported so far based on enzyme deficiency studies. UNG-deficient mouse model developed lymphoid hyperplasia and B-cell lymphomas, although these malignancies manifested only later, in older mice<sup>94</sup>. Furthermore, lacking UNG activity was associated with hyper-IgM syndrome in human patients, in addition to increased genomic uracil incorporation<sup>95,96</sup>. AID deficiency has been also shown to cause similar immunological imbalances in human patients and mice<sup>91</sup>. Accordingly, enhanced AID action and/or the consecutive repair processes were demonstrated to correlate with elevated genomic uracil content, increased mutational rate and cancer progression in human B-cell malignancies<sup>97</sup>. Moreover, enzymatic cytosine deamination of foreign DNA to uracil by the members of the apolipoprotein B editing complex (APOBEC) family was proposed to have a key role in innate immunity defending against retroviral infections<sup>91,98</sup>. Nevertheless, it should be pointed out that the exact details of antibody maturation and antiviral defense processes of immune cells are still unclear to date, and therefore the complete mechanism remains an exciting research area to be explored.

### **1.7. Methods to analyze genomic uracil content**

As discussed above, genomic uracil occurrence was implicated as a key regulator factor in different fields of biology. Hence, various methods were developed to gain qualitative and quantitative information on uracil levels for *in vitro* studies.

Mass spectrometry (MS) based methods are probably the most sensitive methods suitable for direct U-DNA quantification. In these methods, DNA is enzymatically digested to 2'-deoxyribonucleosides or subjected to excision by UNG2, then a high pressure liquid chromatography (HPLC) step can be applied for further purification prior to the analysis by tandem LC/MS/MS<sup>99</sup>. Even though LC/MS/MS (and also gas chromatography coupled MS (GC/MS)) based methods are highly sensitive, they need laborious, excessive sample preparation that requires costly, specialized equipment and expertise. A special challenge



concerns the isobaric character of cytosine and uracil bases. Another sensitive method is UNG-ARP assay which utilizes the biotinylated aldehyde reactive probe (ARP) reagent to measure AP-sites in DNA resulting from UNG-mediated uracil removal <sup>100</sup>. Limiting factors of this kind of methods can be the need of excision of uracil moieties to achieve detection and the dependence of enzymatic multistep sample preparation. Thereafter, several published assays are DNA sequence-specific such as the real-time polymerase chain reaction (PCR) based technique. This method takes advantage of the specific binding incapability of *Pyrococcus furiosus* (Pfu) DNA polymerase against deaminated bases <sup>101</sup>. As a PCR method, it allows to evaluate uracil content only in a limited DNA fragment determined by specific primers, regardless how many uracils the region of interest contain. Furthermore, overall genomic uracil content is calculated as an extrapolation, thereby the selected and measured sequence should represent the entire genome of the organism assuming totally random U-DNA distribution. Each method presented above has their limitations and they do not allow *in situ* cellular detection of genomic uracil residues. Furthermore, no adequate antibody has yet been described for uracil moieties, in contrary to other non-orthodox DNA bases (e.g. 5-methylcytosine, 5-hydroxymethylcytosine, 5hmU, 8-oxo-guanine, etc.). Since uracil residues also escape detection by standard sequencing methods, even the highly powerful single molecule real-time sequencing (SMRT) technology, scientists have to develop new approaches and methods.

### **1.8. The CRISPR/Cas9-mediated gene editing technique**

The basics of the CRISPR/Cas system were discovered in 1987 <sup>102</sup>, but its real function and importance have only been deciphered in the last decade. Since then, it has revolutionized genome editing technology opening up new horizons for not only basic biological research, but also for the development of applied biotechnology and the treatment of diseases.

Clustered regularly interspaced short palindromic repeats (CRISPR) and accompanying CRISPR-associated (Cas) proteins, shortly CRISPR/Cas system originally serve as an adaptive immune system in prokaryotes, developed against bacteriophage infections or intake of plasmids <sup>103,104</sup>. Cas endonucleases has been specialized for the recognition and cleavage of complementary strands of DNA targeted by CRISPR sequences. These non coding short RNA sequences consist of CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA) sequences which are responsible for recruitment and specific guidance of the Cas enzyme, notably their established complex to the target site <sup>105</sup>. Interestingly, short (3–5 bp) DNA sequences, termed protospacer-adjacent motifs (PAM) were found to be important for excision.

Further research led to a highly efficient, programmable gene editing tool, including simplified guide RNA (gRNA) sequence and modified streptococcal Cas9 endonuclease to target any DNA sequence of interest <sup>106</sup>. CRISPR/Cas9 system through the introduction of targeted DSBs exploits endogenous DNA repair mechanisms to generate random insertions or deletions at the target locus. To create a more precise CRISPR/Cas9 system, a cytidine deaminase enzyme was fused to inactive Cas9 which converts cytidine to uridine within the targeted DNA segment <sup>107</sup>. The resulting C:G to T:A substitution can be utilized to gain a desired point mutation allowing more accurate genetic manipulation.

## 2. Aims

Considering the wide range of physiological and pathological conditions where U-DNA level is altered, expanding our knowledge about genomic uracil occurrence including causes and consequences is of particular importance.

Towards this end, novel, reliable detection methods providing either qualitative or quantitative knowledge of U-DNA metabolism are essential. Developing and using a catalytically inactive UNG2 ( $\Delta$ UNG) sensor protein, we aimed to elaborate a dot-blot based U-DNA measurement technique, and thereby quantify genomic uracil level in different biological samples. My goal was to prove that our U-DNA sensor-based method allows *in situ* cellular detection of uracils in DNA through an immunocytochemical approach.

Afterwards, I aimed to further develop this specific U-DNA sensor construct to enable microscopic visualization of genomic uracil residues in complex eukaryotic cells. After experimental validation of the optimized  $\Delta$ UNG-based sensor and U-DNA labeling method, I wished to investigate the nascent genomic uracil distribution, using confocal and superresolution microscopy. I also wished to develop an adequate U-DNA immunoprecipitation method to gain position and sequence-specific information by next-generation sequencing.

In another aspect, genomic uracil accumulation highly depends on the preventive function of dUTPase, which supports low cellular dUTP/dTTP ratio serving as an immunity against uracil incorporation into DNA. Until now, dUTPase function was only studied with gene silencing in cell cultures but gene knock-out strategy has not been published. Therefore, we aimed to investigate the physiological role of dUTPase *in vivo*, applying CRISPR/Cas9-mediated gene editing in mouse model. Importantly, I wished to reveal whether knocking-out of dUTPase could result in viable offspring or dUTPase deficiency may be incompatible with mammalian life.

### 3. Materials and methods

First, I present a list of the used materials and methods directly related to the results of my thesis, indicating also their detailed descriptions in our peer-reviewed publications<sup>108,109</sup> in Table 1. Second, I describe the materials and methods associated with unpublished results of my thesis.

#### 3.1. Materials and methods described in our peer-reviewed publications

Materials and methods	Related paper	Exact localization	Related section of the thesis
Plasmid constructs and cloning	Ref (108)	page 2	4.1.; 4.2.
Cell culture and transfection		page 2-3	4.1.2.; 4.2.
Treatment of cells, DNA isolation and purification		page 3	4.1.
Recombinant protein production		page 3	4.1.; 4.2.
Dot-blot based assay for quantification of DNA and related statistics		page 4	4.1.1.; 4.2.
Staining uracil residues in <i>E. coli</i> genomic DNA		page 4-5	4.1.2.
Staining uracil residues of extrachromosomal plasmids in MEF cells		page 5	4.1.2.
CRISPR constructs; CRISPR/Cas9 efficiency test in mouse embryonic fibroblast (MEF) cells	Ref (109)	2.1.; 2.2.	4.3.1.
Animals; Micromanipulation and detection of gene targeting		2.3.; 2.4.	4.3.1.
Cloning and sequencing; Off-target analysis		2.5.; 2.6.	4.3.1.
Genotyping		2.7.	4.3.2.
Analysis of dissected embryos and blastocyst outgrowth		2.8.; 2.9.	4.3.2.; 4.3.3.
Western blot and related statistical analysis		2.10.; 2.11.	4.3.3.

**Table 1.** Full list of the used materials and methods directly related to the results of my thesis. The related sections of the thesis, and the exact location in our peer-reviewed publications<sup>108,109</sup> containing their detailed descriptions are also indicated.

### 3.2. Additional description related to unpublished experiments presented in section 4.2.

#### 3.2.1. Plasmid constructs and cloning of the FLAG-ΔUNG-SNAP construct

To gain FLAG-ΔUNG-SNAP construct, SNAP encoding sequence from pSNAPf (New England Biolabs (NEB)), was amplified by PCR using primers SNAP-Fw (5'-TAA TGG TAC CGC GGG CCC GGG ATC CAC CGG TCG CCA CCA TGG ACA AAG ACT GCG AAA TG-3') and SNAP-Rev (5'-ATA TCT CGA GGC CTG CAG GAC CCA GCC CAG G-3'). The resulting fragments were digested by KpnI and XhoI and ligated into the KpnI/XhoI sites of the plasmid construct FLAG-ΔUNG-DsRed (in a pET-20b vector, presented in <sup>108</sup>). Primers were synthesized by Sigma-Aldrich and the constructs were verified by sequencing at Microsynth Seqlab GmbH. All UNG constructs were expressed in the *E. coli* BL21(DE3) *ung*-151 strain and purified using Ni-NTA affinity resin (Qiagen) as described in our publication <sup>108</sup>.

#### 3.2.2. Generation of UGI-expressing stable cell line and treatment of cells

The 293T cell line was a generous gift of Yvonne Jones (Cancer Research UK, Oxford, UK). 293T cells were grown in Dulbecco's modified Eagle's medium (Gibco, Life Technologies), supplemented with 50 µg/ml Penicillin-Streptomycin (Gibco) and 10% fetal bovine serum (FBS) (Gibco). Cells were cultured at 37°C in a humidified incubator with 5% CO<sub>2</sub> atmosphere.

The pLGC-hUGI/EGFP retroviral plasmid was kindly provided by Michael D. Wyatt (South Carolina College of Pharmacy, University of South Carolina, US). Retroviral packaging and stable cell line generation was done as described in <sup>110</sup>. Briefly, 293T cells ( $1.5 \times 10^6$  cells in T25 tissue culture flasks) were transfected with 1.5 µg pLGC-hUGI/EGFP, 0.5 µg pCMV-VSV-G envelope and 0.5 µg pGP packaging plasmids using Lipofectamine 3000 reagent (Invitrogen) according to the manufacturer's recommendation. The supernatant, containing lentiviral particles was collected and filtered through a 0.45 µm filter (Merck Millipore) 36 h after the transfection. Successfully transduced HCT116 cells were collected by fluorescence activated cell sorting (FACS) for GFP-positive cells using a BD FACSAria III Cell sorter (BD Biosciences). For comparison, HCT116 cells were also transfected transiently with FuGENE HD (Promega) according to the instructions of the manufacturer. UGI-expressing cells were treated with 20 µM 5FdUR (Sigma) or 100 nM RTX (Sigma) for 48 h before fixation for immunocytochemistry or collecting them for gDNA purification. For dot-blot measurements or U-DNA-Seq, DNA of eukaryotic HCT116 cells (as well as bacterial samples)

was purified using the Quick-DNA™ Miniprep Plus Kit (Zymo Research) using the recommendations of the manufacturer and elution was carried out in nuclease-free water.

### 3.2.3. Immunofluorescent staining and imaging of uracil residues within DNA

Staining of extrachromosomal plasmid DNA was done as described in our publication<sup>108</sup> with minor modifications for comparison of FLAG-ΔUNG-DsRed and FLAG-ΔUNG-SNAP sensor constructs. Briefly, uracil residues were visualized by applying 1.5 µg/ml of the FLAG-ΔUNG-DsRed or the FLAG-ΔUNG-SNAP, and then primary (anti-FLAG M2 antibody (1:10000, Sigma)) and secondary antibodies (Alexa 488 (1:1000, Molecular Probes)).

For immunofluorescent staining of genomic uracil residues, HCT116 cells stably expressing UGI were seeded onto 24-well plates containing cover glasses or onto µ-Slides (or their glass bottom derivatives) (ibidi GmbH) suitable for use in STED and single molecule applications, and treated as indicated. In case of dSTORM imaging, coverslips were coated with poly-D-lysine (Merck Millipore) before seeding the cells. Sub-confluent cultures of cells were fixed using 4% paraformaldehyde (PFA, pH = 7.4 in phosphate-buffered saline (PBS)) or Carnoy's fixative (ethanol: acetic acid: chloroform = 6:3:1) for 15 min. In case of dSTORM imaging, cells were pre-extracted with ice-cold CSK buffer (10 mM PIPES, pH = 6.8, 100 mM NaCl, 300 mM sucrose, 1 mM EGTA, 3 mM MgCl<sub>2</sub>, 0.25% Triton X-100) containing protease and phosphatase inhibitor tablets (Roche) for 5 min before PFA fixation. After washing or rehydration steps (1:1 ethanol:PBS, 3:7 ethanol:PBS, PBS), epitope unmasking was done by applying 2 M HCl, 0.5% Triton X-100 for 30 min. After neutralization with 0.1 M Na<sub>2</sub>B<sub>4</sub>O<sub>7</sub> (pH = 8.5) for 5 min followed by PBS washes, cells were incubated in blocking solution I (TBS-T (50 mM Tris-HCl, pH = 7.4, 2.7 mM KCl, 137 mM NaCl, 0.05% Triton X-100) containing 5% non-fat dried milk) for 15 min, followed by incubation in blocking buffer I supplemented with 200 µg/ml salmon sperm DNA (Invitrogen) for an additional 45 min. Uracil residues were visualized by applying 4 µg/ml of the FLAG-ΔUNG-SNAP construct for 1 h in blocking buffer I with 200 µg/ml salmon sperm DNA at room temperature. After several washing steps with TBS-T containing 200 µg/ml salmon sperm DNA, cells were labelled with 2.5 µM (0.5 µM for dSTORM imaging) SNAP-Surface Alexa Fluor 546 or 647 (NEB) for 20 min in blocking buffer II (5% fetal goat serum (FGS), 3% fetal bovine serum albumin (BSA) and 0.05% Triton X-100 in PBS), and optionally counterstained with 1 µg/ml DAPI (4', 6-diamidino-2-phenylindole, Sigma) nucleic acid stain, followed by several PBS washing steps before embedding in FluorSave™ Reagent (Calbiochem, Merck Millipore).

Confocal images were acquired on a Zeiss LSCM 710 microscope using a 63x (NA = 1.4) Plan Apo objective or a Leica TCS SP8 STED 3X microscope using a 100x (NA = 1.4) Plan Apo objective. STED images were acquired on the Leica TCS SP8 STED 3X microscope using 660 nm STED (1.5 W, continuous wave) laser for depletion (in combination with Alexa 546). Super-resolution images by dSTORM were obtained and reconstructed as previously described<sup>110</sup>. The imaging buffer, consisting of 1 mg/ml glucose oxidase, 0.02 mg/ml catalase, 10% glucose, 100 mM mercaptoethylamine in PBS, was mixed and added to samples (labelled with SNAP-Surface Alexa Fluor 647) just before imaging. The same image acquisition settings were applied on each corresponding sample to allow comparison. A moderate degree of deconvolution was applied to the recorded STED images using the Huygens STED Deconvolution Wizard (Huygens Software), based on theoretical point spread function values. Fluorescence images were processed using ZEN software.

### **3.2.4. U-DNA immunoprecipitation**

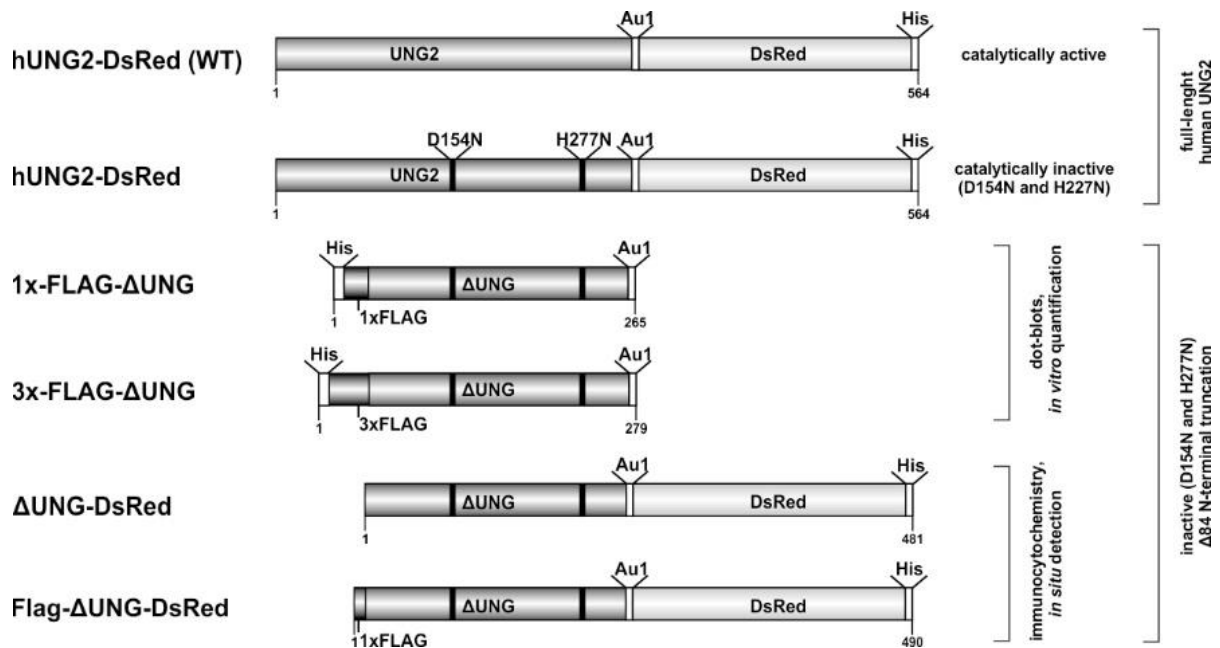
12 µg of gDNA was sonicated into fragments ranging between 100 and 500 basepairs (checked by agarose gel electrophoresis) with a BioRuptor (Diagenode). 25% of the samples was saved as input, and the remaining DNA was resuspended in the following IP buffer: 30 mM Tris-HCl, pH = 7.4, 140 mM NaCl, 0.01% Tween-20, 1 mM 0.5 mM ethylenediaminetetraacetic acid (EDTA), 15 mM β-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, 5 mM benzimidazole. Immunoprecipitation was carried out with 15 µg of 1xFLAG-ΔUNG construct for 2.5 h at room temperature with constant rotation. Anti-FLAG M2 agarose beads (Sigma) were equilibrated in IP buffer and then added to the IP mixture for 16 h at 4°C with constant rotation. Beads were washed three times for 10 min in IP buffer and resuspended in elution buffer containing 1% sodium dodecyl sulphate (SDS), 0.1 M NaHCO<sub>3</sub>. Elution of uracil sensor protein binding U-DNA was done by vortexing for 5 min with an additional incubation for 20 min with constant rotation. After centrifugation (13000 rpm for 3 min), supernatant was transferred to clean tubes. This procedure was repeated with the same amount of elution buffer and eluted protein/DNA complexes were combined in the same tube. Samples were incubated with 150 µg/ml RNase A (Epicentre) for 30 min, followed by the addition of 500 µg/ml Proteinase K (Sigma) for 1 h at 37°C for removal of residual RNA and proteins. Immunoprecipitated DNA was purified with NucleoSpin Gel and PCR Clean-up Kit (MACHEREY-NAGEL GmbH & Co. KG) according to the manufacturer's instructions. Enrichment of uracil in the DNA samples was examined by dot-blot assay, then input and enriched U-DNA samples were subjected to next-generation sequencing (NGS).

## 4. Results

### 4.1. Uracil-DNA detection applying catalytically inactive UNG constructs in *in vitro* and cellular studies

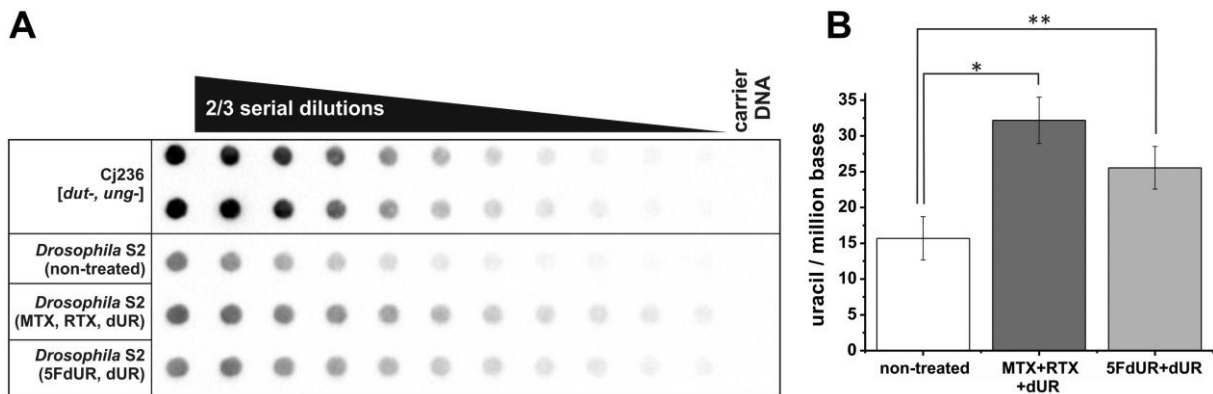
#### 4.1.1. A dot-blot based quantitative assay for detection of genomic uracil *in vitro*

For uracil-DNA detection, a double mutant UNG2 enzyme (D154N, H268N) was created that preserved the highly specific and strong binding capability with uracil-containing DNA, but its function of U-DNA excision was abolished<sup>108</sup>. In addition, the N-terminal 84 residues were removed from the enzyme in order to eliminate the recognition site for PCNA and RPA proteins to prevent protein-protein interactions. The  $\Delta$ UNG uracil-recognizing core was fused to different epitope tags: His-tag for affinity purification, 1x/3x-FLAG and Au1 tags for antibody-based detection, or DsRed-monomer for direct fluorescent detection, yielding various sensor constructs (Figure 3).



**Figure 3.** Scheme of the used constructs with their applications highlighted. The catalytically inactive (D154N and H277N, mutated sites indicated with black lines) and truncated  $\Delta$ UNG (lacking the N-terminal 84 residues) was created from human UNG2. The different epitope tags (His, 1x/3x-FLAG, DsRed) fused to the  $\Delta$ UNG uracil-recognizing core are also shown here. Figure was taken from our publication<sup>108</sup>.

Detection and relative quantification of genomic uracil relies on antigen-antibody interactions in the developed dot-blot assay. Briefly, uracil residues in the DNA samples can be specifically recognized with the catalytically inactive, 3x-FLAG- $\Delta$ UNG construct on a membrane surface, followed by anti-Flag primary and secondary antibody labelling (detailed in our publication<sup>108</sup>). Genomic uracil levels are then calculated in the unknown samples using interpolation of their normalized intensities to a standard calibration curve. To record a standard calibration curve, uracilated DNA isolated from CJ236 (*dut*<sup>-</sup>, *ung*<sup>-</sup>) *E. coli* cells were used in each dot-blot experiment since the U-DNA content of this specific cell line has been previously quantified as ~6580 uracil/million bases<sup>100,101</sup>. Applying dot-blot assay, U-DNA content of bacterial strains with (*dut*<sup>+</sup>, *ung*<sup>+</sup>), (*dut*<sup>-</sup>, *ung*<sup>-</sup>) and *ung*<sup>-</sup> genetic background was compared after deoxyuridine (dUR) and/or 5FdUR drug treatments. Since *ung* deficiency was proven to be a key factor for the increase of genomic uracil content upon drug treatment, *Drosophila* embryo-derived, *ung*<sup>-</sup> Schneider S2 cell line was then investigated as an eukaryotic cellular model (Figure 4A). Drug combination of MTX, RTX and dUR, similarly to treatments with 5FdUR and dUR, resulted in a significant increment of genomic uracil level compared to non-treated cells (Figure 4B).

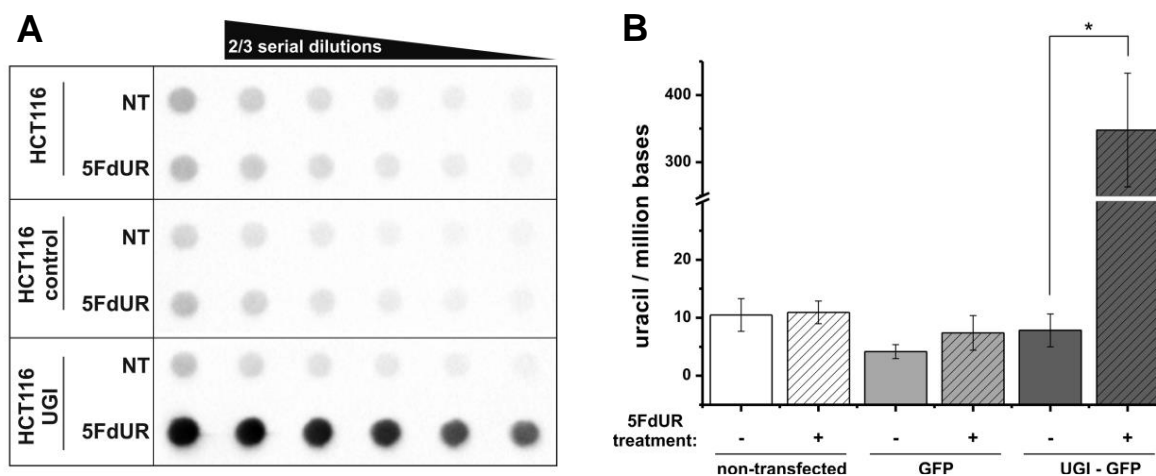


**Figure 4.** Measurement of genomic uracil content of *Drosophila* S2 cells after treatment with drugs targeting the *de novo* thymidylate biosynthesis pathway. **(A)** In the dot-blot assay two-third dilution series from CJ236 (*dut*<sup>-</sup>, *ung*<sup>-</sup>) *E. coli* gDNA were used and served as standard to quantify genomic U-DNA levels in different samples isolated from drug-treated and non-treated S2 cells. **(B)** Bar graph shows the number of uracil moieties/million bases of each sample. Error bars indicate standard errors of the mean (SE). \* =  $p < 0.01$ , \*\* =  $p < 0.05$ . Calculations were based on six independent datasets ( $n = 6$ ). Figure was taken from our publication<sup>108</sup>.



We proceeded to investigate U-DNA level alterations using a human colon carcinoma cell line (Figure 5A). HCT116 was chosen because of its favorable features such as chromosomal stability and near diploid chromosome set as compared to other cancer cell lines. HCT116 is MMR deficient (lacking hMLH1), thus in this cell line predominantly BER pathway is responsible for removal of uracil from DNA. Additionally, many isogenic cell lines are also available differing in their DNA repair protein repertoire for potential further studies.

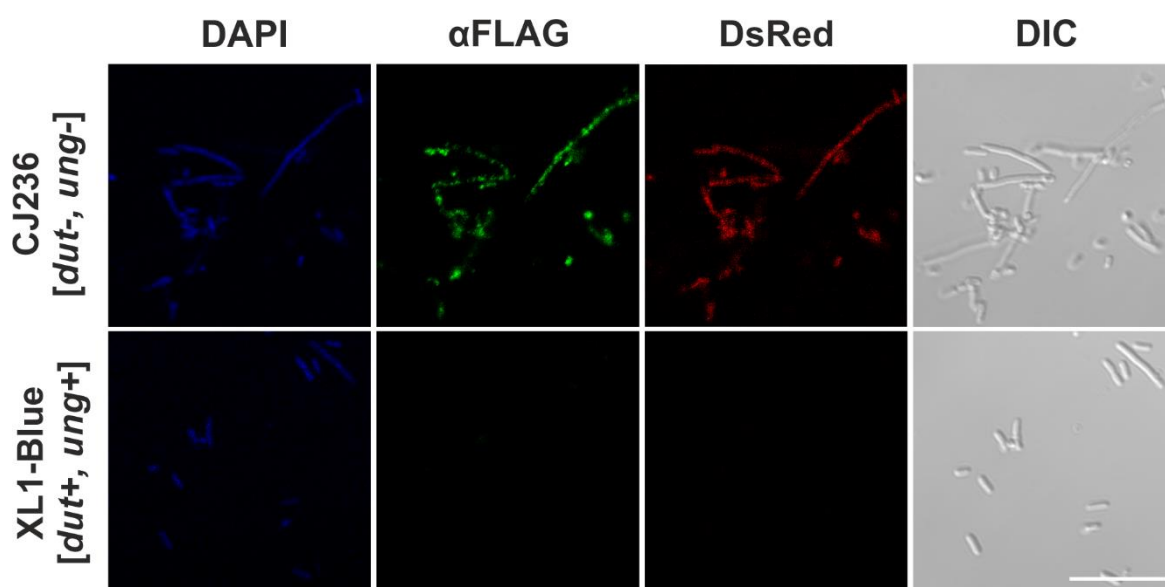
In our experiments, inhibition of the major uracil-recognizing factor, UNG was attained by introduction of UGI (derived from *Bacillus subtilis* bacteriophage PBS2). Expression of the human codon optimized UGI (along with GFP for tracking purposes) and 5FdUR treatment together significantly elevated the uracil content of genomic DNA compared to non-treated cells (Figure 5B). It is important to note that either UGI expression or drug treatment on their own did not lead to increased U-DNA level. Similarly, transfection with the empty vector expressing only GFP as a control neither affected genomic uracil content.



**Figure 5.** Quantification of genomic uracil levels in HCT116 cells after drug treatment and UNG inhibition by dot-blot assay. **(A)** Genomic uracil levels of non-transfected, GFP-transfected (control) and UGI-GFP-transfected HCT116 cells were measured in the presence or absence (NT) of 5FdUR treatment. **(B)** Bar graph shows the number of uracil moieties/million bases of each sample (mean values  $\pm$  SE). \* =  $p < 0.05$ . Calculations were based on four independent datasets ( $n = 4$ ). Figure was taken from our publication <sup>108</sup>.

#### 4.1.2. *In situ* cellular detection of uracil residues in DNA

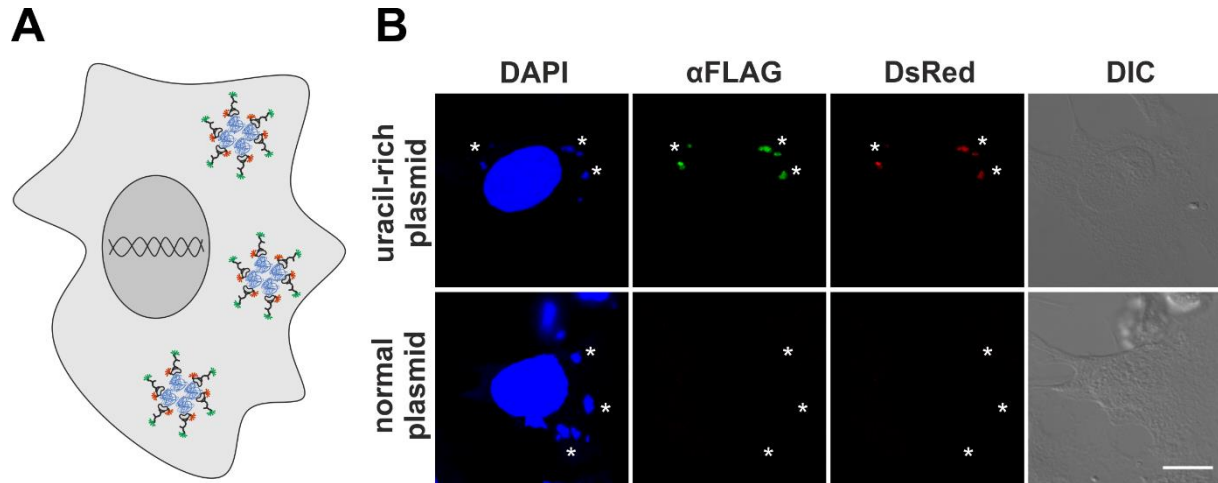
After we have successfully applied our novel uracil sensor construct in dot-blot experiments, we aimed to test the Flag-tagged and DsRed-tagged  $\Delta$ UNG protein for *in situ* visualization of genomic uracil residues. Staining of genomic uracils was achieved either indirectly *via* immunocytochemistry against the Flag epitope tag or directly *via* the fluorescent DsRed signal in CJ236 (*dut*<sup>-</sup>, *ung*<sup>-</sup>) cells. According to our expectations, using the same labeling procedure in uracil-free XL1-Blue (*dut*<sup>+</sup>, *ung*<sup>+</sup>) *E. coli* cells did not result in any visible signal (Figure 6). In the uracil-rich milieu, U-DNA staining colocalized with the signal of DAPI used to counterstain DNA.



**Figure 6.** Visualization of genomic uracil using the Flag- $\Delta$ UNG-DsRed sensor construct in uracil-containing CJ236 (*dut*<sup>-</sup>, *ung*<sup>-</sup>) *E. coli* cells as compared to uracil-free XL1-Blue (*dut*<sup>+</sup>, *ung*<sup>+</sup>) *E. coli* cells. Detected uracil residues are shown *via* immunostaining against Flag-tag (green) or direct signal of DsRed (red) protein. DAPI is a blue-fluorescent DNA stain. Differential interference contrast (DIC) image was also acquired. Scale bar represents 10  $\mu$ m. Figure was taken from our publication <sup>108</sup>.

Afterwards, in order to apply our uracil-staining method in mammalian cell lines, we set up an experiment taking advantage of the phenomenon that plasmid DNA could form cytoplasmic aggregates when being transfected in a high amount (Figure 7A). We used *ung*<sup>-/-</sup> MEF cells for transfection with uracil-rich or normal (uracil-free) plasmids produced in CJ236 (*dut*<sup>-</sup>, *ung*<sup>-</sup>) or in XL1-Blue (*dut*<sup>+</sup>, *ung*<sup>+</sup>) *E. coli* cells, respectively (Figure 7B). Such *ung*<sup>-/-</sup> background was required to avoid degradation of the highly uracilated plasmid DNA initiated by the endogenous UNG activity. Plasmid aggregates are clearly visible with DAPI staining in the

cytoplasm (shown by DIC images). These aggregates are marked with white asterisks for better visualization. Upon transfection with uracil-rich plasmids, aggregates can also be observed by Flag-tag or DsRed detection implying specific labeling with the U-DNA sensor construct, in contrast with the case of transfection with normal plasmids.

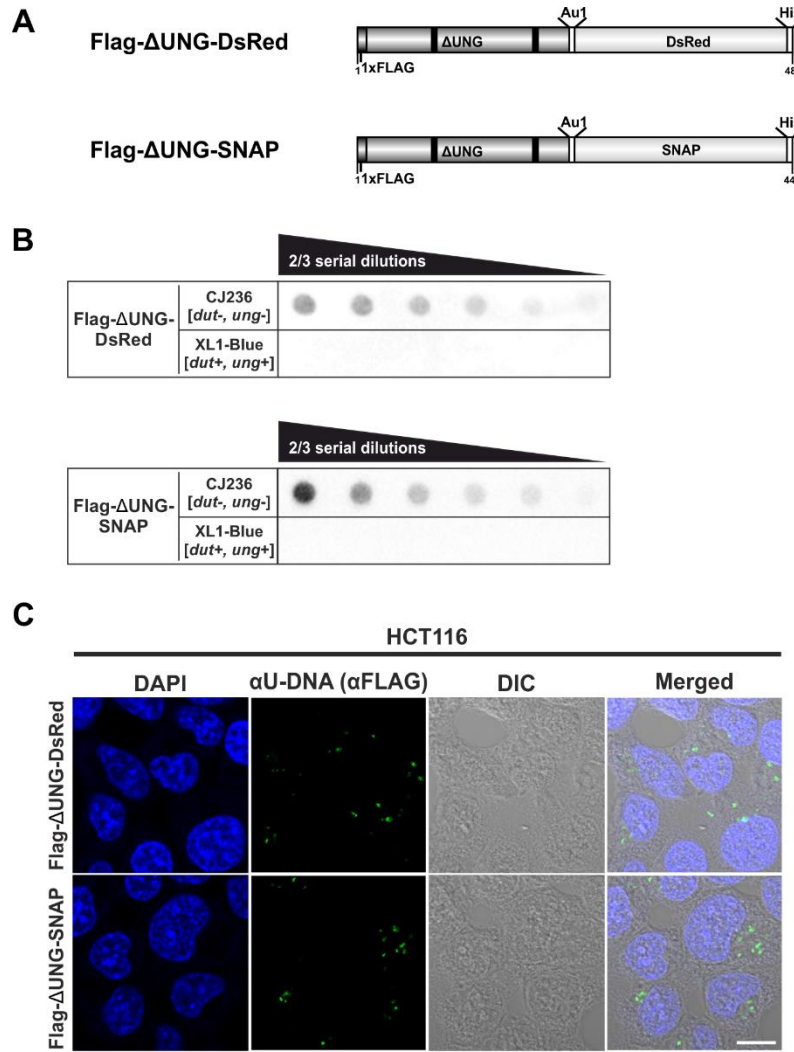


**Figure 7.** Imaging of uracil-containing DNA in (*ung*<sup>-/-</sup>) MEF cells transfected with uracil-rich plasmids. **(A)** Schematic image of the uracil-staining strategy relying on binding of Flag- $\Delta$ UNG-DsRed molecules to uracilated plasmid aggregates in the cytoplasm. **(B)** MEF cells were transfected with high amount of uracil-rich or normal (uracil-free) plasmids. Plasmid aggregates are indicated with white asterisks (\*) in DAPI (blue) stained, anti-Flag (green) immunostained or DsRed (red) labeled cells. The DAPI signal is oversaturated to strengthen the faint DAPI staining of the plasmid aggregates. Scale bar represents 10  $\mu$ m. Figure was partially taken from our publication<sup>108</sup>.

## **4.2. Promising applications of $\Delta$ UNG-based U-DNA labeling techniques for analysis of uracil distribution pattern in human genomic DNA**

### **4.2.1. Development of a novel U-DNA sensor construct appropriate for super-resolution imaging**

As our results showed that the presented  $\Delta$ UNG-based U-DNA labeling approach has the potential for *in situ* cellular detection, I aimed to further develop the U-DNA sensor constructs to allow visualization of genomic U-DNA in complex eukaryotic cells. Since these cells, according to the expectations, contain lower U-DNA level, *in situ* detection represented a significant challenge. The basis of specificity for U-DNA staining was still provided by the catalytically inactive  $\Delta$ UNG enzyme capable of binding but not hydrolyzing uracil moieties in DNA. Nevertheless, in order to achieve a versatile labeling technique to facilitate super-resolution imaging of U-DNA, DsRed was replaced to SNAP-tag in the C terminal end of  $\Delta$ UNG, yielding Flag- $\Delta$ UNG-SNAP, thus generating a novel sensor construct (Figure 8A). The SNAP-tag offers a flexible biorthogonal chemical labelling strategy as it reacts specifically and covalently with benzylguanine derivatives, permitting the irreversible labelling of SNAP fusion proteins with a wide variety of synthetic probes<sup>111</sup>. To check whether the functionality of this new construct is still preserved, I performed dot-blot and staining experiments. Results shown in Figure 8B indicate that the Flag- $\Delta$ UNG-SNAP construct is functional and shows reliable U-DNA detection using dot-blot assay, when compared to Flag- $\Delta$ UNG-DsRed protein. To check the possible use of this construct in immunocytochemistry, first I applied the Flag- $\Delta$ UNG-SNAP construct on HCT116 cells transfected with uracil-enriched plasmids. The new labeling construct specifically recognized the presence of extrachromosomal plasmid aggregates in the cytoplasm (Figure 8C).

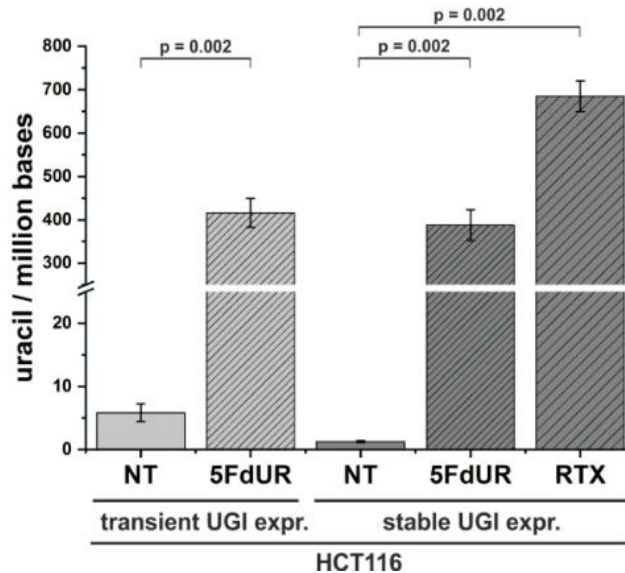


**Figure 8.** Validation of the new Flag-ΔUNG-SNAP uracil-sensor. **(A)** Scheme of Flag-ΔUNG-SNAP construct as compared to Flag-ΔUNG-DsRed. **(B)** Dot-blot assay was used to compare uracil binding capability of Flag-ΔUNG-DsRed and Flag-ΔUNG-SNAP sensor constructs on CJ236 (*dut-*, *ung-*) (positive control) and XL1-Blue (*dut-*, *ung+*) (negative control) *E. coli* DNA samples. **(C)** Comparison of these constructs in detecting uracil-rich plasmid aggregates within HCT116 cells. The sensors were visualized through the Flag epitope tag. DAPI was used to counterstain DNA. Scale bar represents 10 μm.

#### 4.2.2. Imaging of endogenous genomic uracils in human cells by confocal and super-resolution microscopy

Possessing this new sensor, I aimed to detect *in situ* endogenous genomic uracil in human cells. As we showed previously by dot-blot measurements (Figure 5), observation of uracil moieties is hampered by the efficient and fast action of UNG, the initiator of the BER process (see Figure 2). Hence, the UNG-inhibitor UGI was expressed in HCT116 cells to prevent the action of the major uracil-DNA glycosylase, thus allowing detection of nascent uracils. Besides transient transfection, I also established a stably UGI transfected HCT116 cell line by retroviral

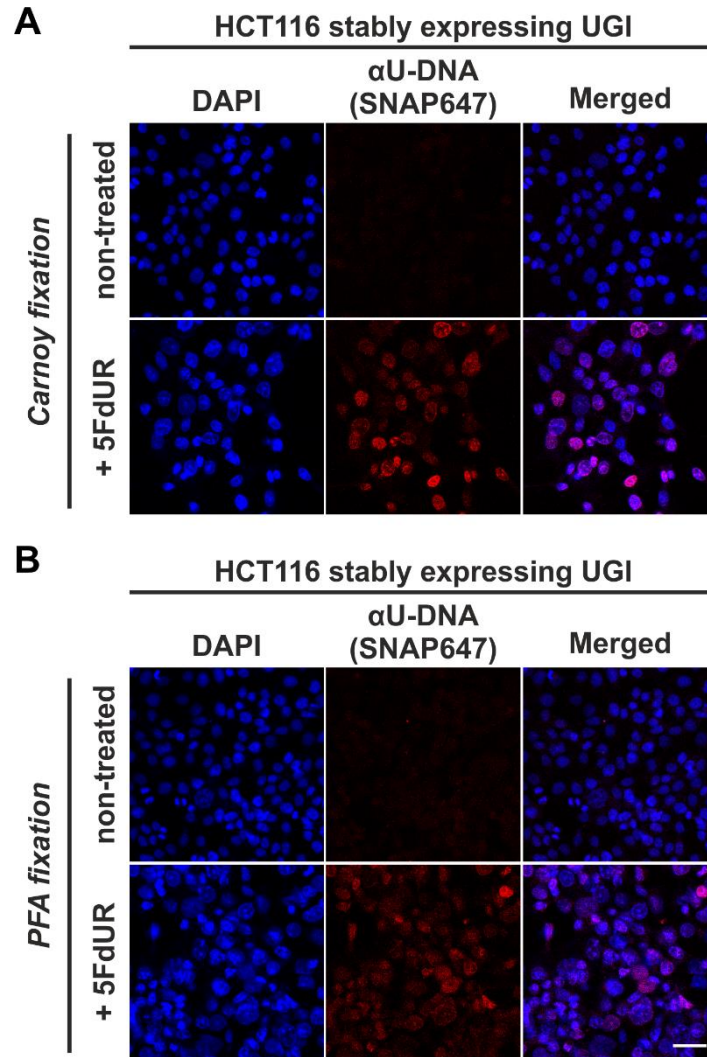
transduction of human codon optimized UGI along with GFP. GFP-positive population was then selected by FACS. Cells were then treated with 5FdUR or RTX. This dual approach (UGI expression and drug treatment) led to significantly elevated uracil levels in HCT116 cells either transiently or stably expressing UGI (~400 uracil moieties/million for 5FdUR and ~700 uracil moieties/million for RTX) as compared to non-treated (NT) cells (~2-5 uracil moieties/million), as shown on Figure 9.



**Figure 9.** Measurement of genomic uracil levels using dot-blot assays in non-treated (NT) and drug (5FdUR or RTX) treated HCT116 cells either transiently or stably expressing UGI. Bar graph shows the uracil moieties / million bases of each sample (mean values  $\pm$  SE). Calculations were based on six independent datasets ( $n = 6$ ). Statistical analysis was carried out using the non-parametric two-sided Mann-Whitney U test.  $p = 0.002$ .

Since dot-blot experiments also confirmed the specificity of the new Flag- $\Delta$ UNG-SNAP uracil-sensor, I moved forward to apply the new sensor to detect *in situ* endogenous genomic uracils in the generated UGI-expressing cell line. Cellular U-DNA level could be readily observable on conventional confocal microscopic images upon drug treatment in UNG-inhibited cells (Figure 10A). For visualization of our sensor, photostable SNAP-tag substrates (here SNAP647 or SNAP546) were used. It was also important to modify U-DNA staining conditions such that to allow further colocalization studies with other chromatin factors. Figure 10B demonstrates that our Flag- $\Delta$ UNG-SNAP sensor can be applied for straightforward staining of genomic uracil even with PFA fixation. Unlike Carnoy (containing mixture of ethanol: acetic acid: chloroform = 6:3:1), PFA fixative is compatible with most antibody-based staining procedures, thus it is suitable for multi-colour imaging allowing colocalization studies. Following either fixation method, treatment with HCl was found to be critical in order to increase DNA

accessibility for efficient U-DNA staining and to eliminate the potential of interaction between the expressing UGI and the applied UNG sensor proteins.

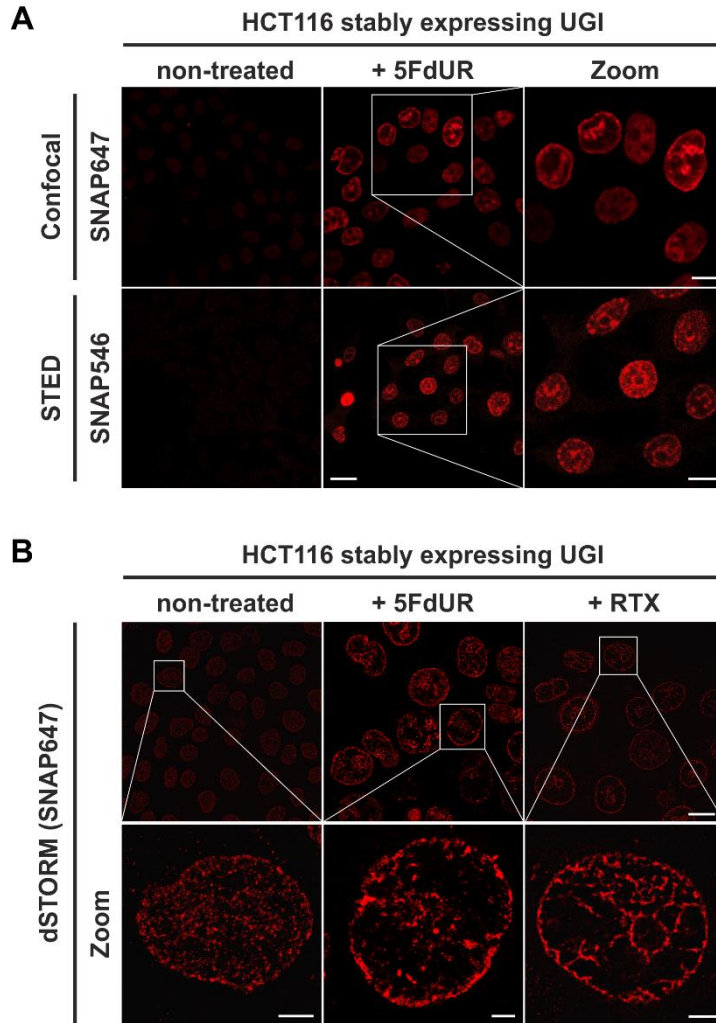


**Figure 10.** *In situ* detection of the cellular endogenous U-DNA content in HCT116 cells expressing UGI with or without 5FdUR treatment. The optimized staining method is capable of comparable, specific uracil detection in HCT116 cells after either Carnoy (**A**) or PFA (**B**) fixation, followed by HCl treatment. Uracil residues are labelled by our Flag- $\Delta$ UNG-SNAP sensor protein visualized by the SNAP647 substrate. DAPI was used for DNA counterstaining. Scale bar represents 40  $\mu$ m.

Next, we attempted to use super-resolution microscopy to have a better track of the uracil distribution pattern even in case of the low genomic uracil level found in the non-treated cells. Figure 11A-B compares confocal, STED and dSTORM microscopy techniques for U-DNA detection. The exquisite sensitivity of dSTORM is apparent from these experiments as it can detect the low level of genomic uracil in non-treated cells (Figure 11B). We observed different heterogeneous staining in the nucleus for uracil in non-treated and drug-treated cells.



Furthermore, images of drug-treated cells show uracil staining with signal enrichment at the nuclear membrane and areas surrounding the nucleoli.



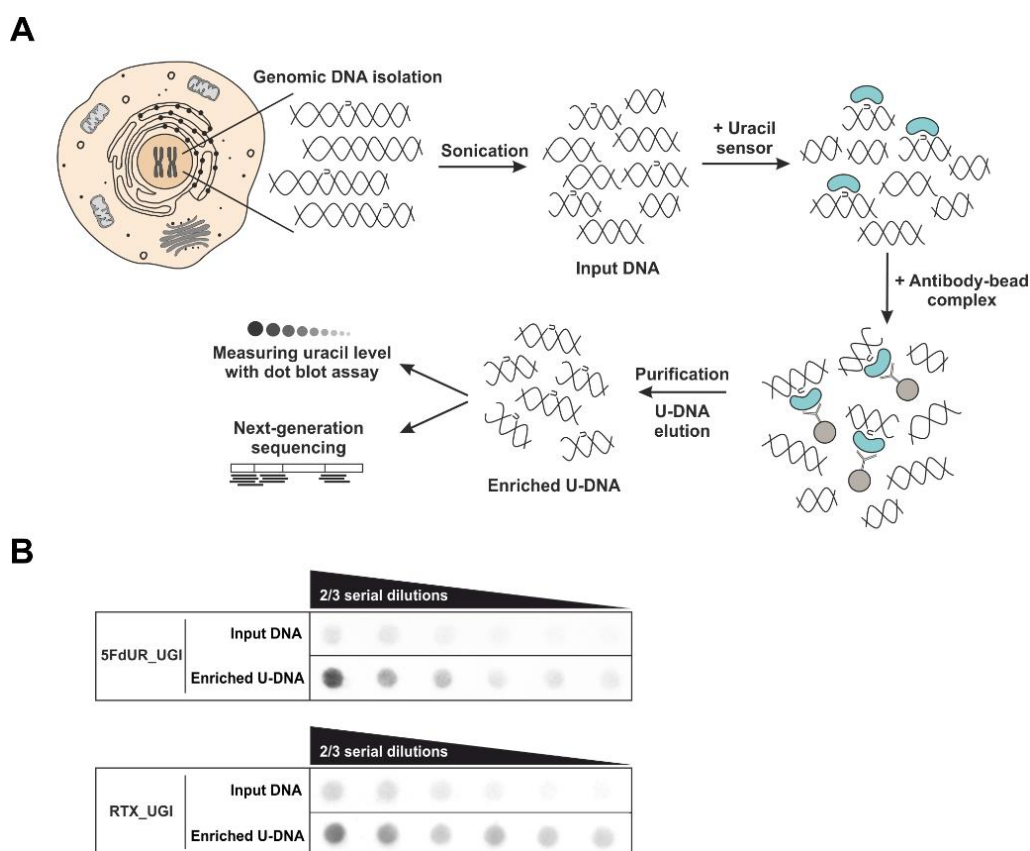
**Figure 11.** The Flag- $\Delta$ UNG-SNAP sensor enables super-resolution detection of genomic uracil by STED and dSTORM microscopy. Different SNAP-tag substrates, SNAP647 for confocal and dSTORM, or SNAP546 for STED were used to label Flag- $\Delta$ UNG-SNAP. **(A)** U-DNA staining was performed on non-treated or 5FdUR treated HCT116 cells stably expressing UGI. Scale bar represents 20  $\mu$ m for whole images and 10  $\mu$ m for zoomed sections. **(B)** dSTORM imaging was also performed on non-treated or drug-treated (5FdUR or RTX) HCT116 cells stably expressing UGI to compare the sensitivity of these imaging techniques. Scale bar represents 10  $\mu$ m for whole images and 2  $\mu$ m for zoomed sections.

#### 4.2.3. Development of U-DNA-Seq method to unravel genome-wide distribution of uracil residues

The observation of the non-random distribution pattern prompted us to address the characteristics of this pattern in a more quantitative manner and try to map it on the genomic sequence. Towards this aim, I designed an adequate DNA immunoprecipitation method that



can provide U-DNA specific genomic information by next-generation sequencing (NGS). This method, termed U-DNA-Seq, is based on the rationale of the well-established DNA-IP-seq technology. Figure 12A presents the scheme of the developed protocol leading to an enriched U-DNA sample that can be then subjected to NGS. Immunoprecipitation was carried out applying our Flag-tagged catalytically inactive  $\Delta$ UNG sensor to bind to uracils in purified and fragmented genomic DNA, followed by pull-down with anti-FLAG agarose beads (for details see section 3.2.4.). Successful enrichment of U-DNA was confirmed by dot-blot assay for 5FdUR or RTX treated, UGI-expressing cells (Figure 12B). Input DNA and enriched U-DNA samples could be subjected to library preparation and NGS.



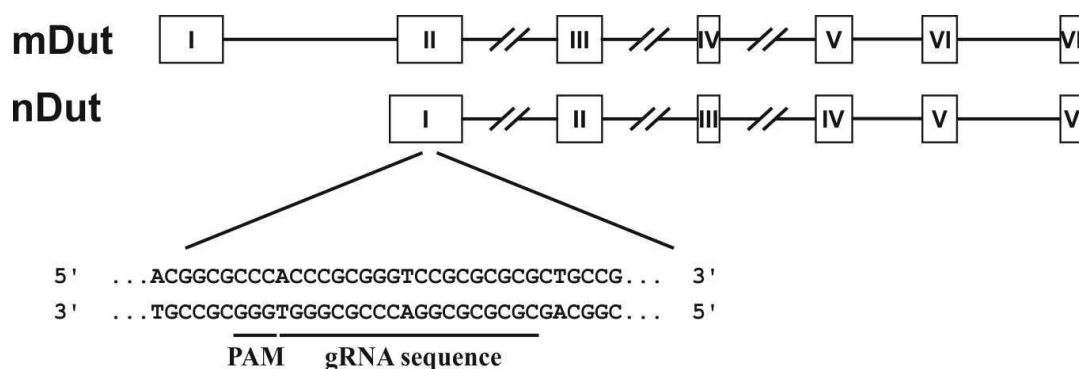
**Figure 12.** U-DNA-Seq provides genome-wide mapping of uracil-DNA distribution. **(A)** Schematic image of the novel U-DNA-Seq method. After sonication, enrichment of the fragmented U-DNA was carried out by the 1x-FLAG- $\Delta$ UNG sensor construct followed by pull-down with anti-FLAG agarose beads. **(B)** U-DNA enrichment compared to input DNA can be confirmed by dot-blot assay before sequencing. The samples are as follows: 5FdUR treated UGI-expressing (5FdUR\_UGI) and RTX treated UGI-expressing (RTX\_UGI) HCT116 cells. In case of the given treatment, the same amount of DNA was loaded from input and enriched U-DNA samples providing correct visual comparison of the dots.

### 4.3. A critical investigation of the physiological role of dUTPase in a mammalian model, using CRISPR/Cas9-mediated gene editing

#### 4.3.1. CRISPR/Cas9-mediated knock-out of dUTPase in mice

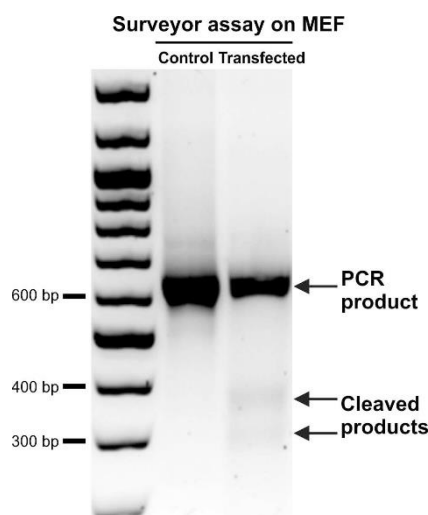
In my PhD thesis up to this point, I have focused on U-DNA detection harnessing the natural specificity of uracil-recognizing UNG. However, another factor has also a highly relevant role in genomic uracil accumulation. Namely, the cellular dUTP/dTTP ratio, regulated by dUTPase enzyme in most organisms, has a decisive influence on the resulting chemical composition of the genomic DNA (see Figure 1). To investigate the effects of dUTPase deficiency *in vivo*, specific knock-out of *dut* gene was addressed by CRISPR/Cas9-mediated gene editing strategy.

For the knock-out experiment, one guide RNA was adequate enough to target dUTPase because a single gene (*dut*) on chromosome 2 encodes both nuclear (nDut) and mitochondrial (mDut) isoforms of dUTPase in mouse (Figure 13). A synthetic, custom single-gRNA (sgRNA) was chosen, containing both the crRNA and tracrRNA sequences, in order to specifically and efficiently target *dut* gene by endonuclease activity of Cas9.



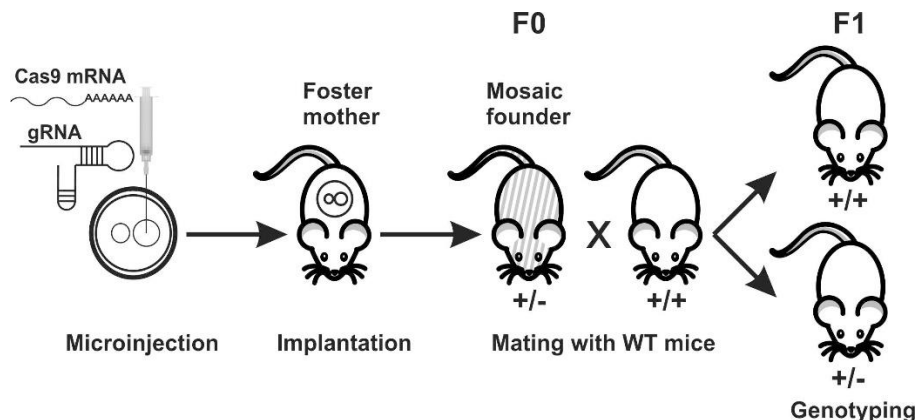
**Figure 13.** Schematic diagram of the target site of the applied sgRNA for dUTPase knock-out. Exons are numbered, while introns are indicated as simple lines (for longer introns lines are broken). Guide RNA (gRNA) and stand by PAM sequences are located in the first common exon of the two isoforms (exon 2 for mDut and exon 1 for nDut). Figure was taken from our publication <sup>109</sup>.

First, I have tested the chosen gRNA and Cas9 mRNA by the surveyor assay following co-transfection of mouse embryonic fibroblast (MEF) cell line. As seen in Figure 14, cleaved products could be observed confirming accomplished gRNA-directed genome modifications of Cas9 nuclease.



**Figure 14.** Surveyor assay showing efficient CRISPR genome editing events. CRISPR/Cas9-induced cleaved products were clearly visible in MEF cells transfected with target sgRNA (presented in Figure 13) and Cas9 mRNA. In the case of non-transfected (control) cells, DNA was not fragmented after digestion by Surveyor nuclease. Figure was taken from our publication <sup>109</sup>.

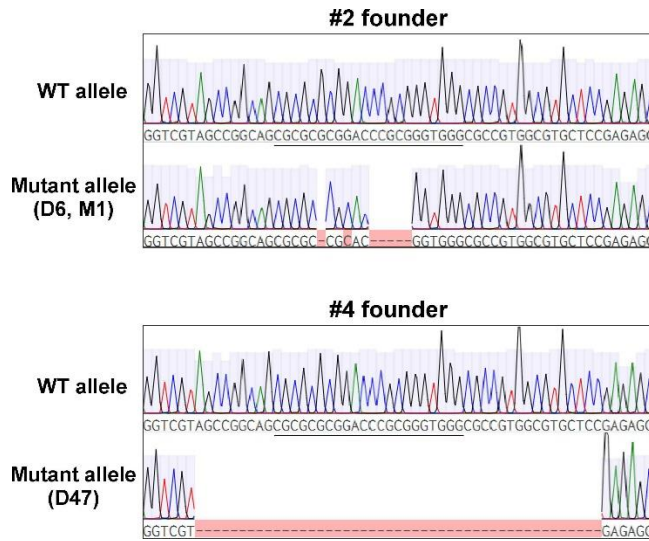
This promising result motivated the *in vivo* trials to produce CRISPR-targeted dUTPase knock-out mice (on FVB/N background). To this end, mouse zygotes were microinjected with Cas9 mRNA along with sgRNA, followed by implantation into foster mother animals, as the scheme shows in Figure 15. Their offspring, the so-called founders, were then cross-bred with wild-type (WT) mice to eliminate genetic mosaicism provided by germline transmission of the genome edited locus. This resulted in new-born animals that were wild-type (+/+) or heterozygous (+/-) for *dut* gene, separated by genotyping.



**Figure 15.** Scheme of the experimental strategy to establish dUTPase knock-out mice. The main steps are as follows: (1) microinjection of fertilized oocytes with gRNA and Cas9 mRNA; (2) implantation of the modified zygotes into foster mothers resulting in mosaic founders (F0); (3) mating them with WT mice leading to the F1 generation; (4) identification of *dut* +/- offspring by genotyping. Figure was taken from our publication <sup>109</sup>.

Using the herein described method of knock-out mouse model production led to two relevant founder animals, numbered as #2 and #4. Both these founders presented CRISPR/Cas9 editing events. To identify the mono-allelic insertion/deletion mutations, the different alleles had to be analysed separately. Therefore, target sites from isolated gDNA were PCR amplified and

transferred into empty cloning vectors. Following bacterial transformation, colonies were examined with Sanger sequencing. Sequencing results verified distinct random CRISPR events at the target locus, namely six bp deletion and one base substitution (D6, M1) in mice #2; and 47 bp deletion (D47) in mice #4 (Figure 16).



**Figure 16.** Sequencing traces encompassing the dUTPase target region clearly show the CRISPR/Cas9 induced genome modifications: 6 bp deletion and 1 bp substitution (D6, M1); and 47 bp deletion (D47) in founder animals #2 and #4, respectively. For comparison, WT allele is also presented including the CRISPR target site underlined. The CRISPR induced sequence differences are indicated by red highlight. Figure was taken from our publication <sup>109</sup>.

Figure 17 summarizes these sequencing results with the predicted consequences on the corresponding amino acid sequences. In the case of mice #2, (D6, M1) DNA alterations prospectively cause modified amino acid sequences, including the deletion of two amino acids and the change of other two. In the case of mice #4, deletion of 47 bp (D47) has a more severe impact on the genetic code implying frameshift mutation and early stop codons upon transcription.

DNA level	#2	5' GCGGCTCTCGGAGCACGCCACGGCGCCACCCGCGGGTCCGCGCGCGCTGCCGGCTACGACCTAT 3'	WT
	+/-	GCGGCTCTCGGAGCACGCCACGGCGCCACC-----GTGCG-GCGCGCTGCCGGCTACGACCTAT	D6,M1
	#4	5' GCGGCTCTCGGAGCACGCCACGGCGCCACCCGCGGGTCCGCGCGCGCTGCCGGCTACGACCTAT 3'	WT
	+/-	GCGGCTCTC-----ACGACCTAT	D47
Aminoacid level	#2	MetPCSEDAAAVSASKRARAEDGASLRFVRLSEHATAPTRGSARAAGYDLFSAYDYTISPMEKAI	WT
	+/-	MetPCSEDAAAVSASKRARAEDGASLRFVRLSEHATAPT--RRAAGYDLFSAYDYTISPMEKAI	D6,M1
	#4	MetPCSEDAAAVSASKRARAEDGASLRFVRLSRPIQCLStopLYNITHGESHREDRHSDSCPFVW	D47
	+/-	LWKSSSTFWLGCKALHRCRSWCHRRGLQKRWGRAVStopLWEREVStopSEKRStopSDCAHL StopADFLSRLRGSADPGStopHRERLRLRLHREEL	

**Figure 17.** CRISPR targeted DNA locus and its predicted amino acid transcript in the two heterozygous (*dut +/-*) founder mice (#2 and #4), compared to the WT. Mouse #2 carried 6 bp deletion and a C to G mutation (D6, M1) resulting in two deletion and two change of amino acids. Mouse #4 carried 47 bp deletion (D47) leading to frameshift mutation and early stop

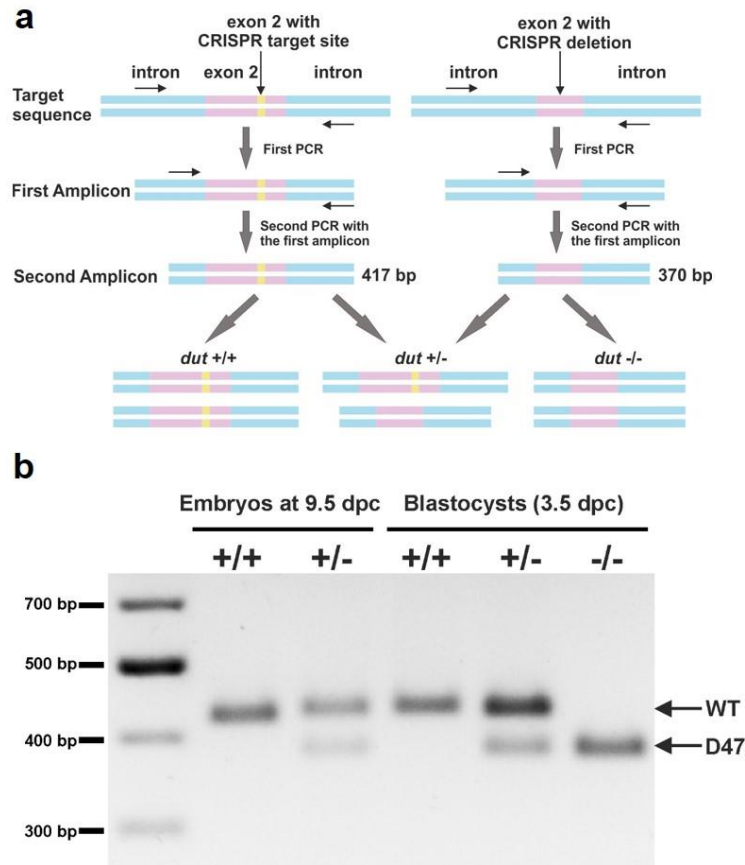
codons indicated with dashed lines. Target site for CRISPR is underlined in the WT DNA sequence. Figure was taken from our publication <sup>109</sup>.

One of the most critical limitation of CRISPR/Cas9 system is the possibility of off-target effects arisen from non-specific nuclease activity <sup>112</sup>. Therefore, we analysed the top twenty potential off-target sites of CRISPR/Cas9 for our on-target sgRNA using the online predictor software CCTop <sup>113</sup>. Among these genomic segments, ten selected possible off-target sites were successfully tested with Sanger sequencing in mouse #4, and compared to the WT. None of the sequenced sites had mutations, suggesting there were no off-target effects at the investigated genomic regions.

#### **4.3.2. Homozygote knock-out of dUTPase leads to early embryonic lethality**

Hereinafter, founder mouse #2 (D6, M1) and #4 (D47) were crossed with wild-type mice for germline transmission of the described *dut* mutations, yielding heterozygous (*dut* +/-) or wild-type (*dut* +/+) animals in F1 generation. Further breeding of heterozygous D47 siblings were necessary to have the possibility of gaining homozygous, *dut* -/- knock-out mouse models. It is important to note that progeny of (D6, M1) and D47 strains did not show any gross abnormalities or systemic diseases in adulthood, and were fertile through multiple generations.

To genotype the mouse progeny derived from numerous crosses, a genotyping PCR method had to be designed. As shown in Figure 18A, the optimized semi-nested PCR method relies on the differing length of PCR products amplified from the WT or the CRISPR targeted D47 allele. The two-step reaction allowed reliable genotyping not only from mouse tail DNA, but also from samples with low DNA content, such as early embryos. Figure 18B shows representative genotyping results using the specific semi-nested PCR on lysed embryos. After heterozygous crossing, we found all the three feasible genotypes (*dut* +/+, +/-, -/-) in 3.5-day-old embryos (blastocysts), although *dut* -/- embryos could not be detected at later developmental stages, for instance at 9.5 dpc (Figure 18B).



**Figure 18.** Genotyping of embryos produced by heterozygous D47 crossing. **(A)** Scheme of the applied semi-nested genotyping PCR where the first amplicon served as template for a second PCR. The resulting product length reflect the CRISPR/Cas9-mediated deletion event: 370 bp for the D47 allele while 417 bp for the WT allele. Exon 2 (pink) containing the CRISPR target site (yellow), adjacent introns (blue) and the applied primers (shown as arrows) are indicated. **(B)** Visualization of semi-nested PCR products by agarose gelelectrophoresis. The upper (417 bp) and the lower (370 bp) bands correspond to WT and D47 allele, respectively. Figure was taken from our publication <sup>109</sup>.

During our studies we have genotyped numerous additional generations derived from heterozygous D47 crossings. Nevertheless, we never found viable *dut* -/- offspring, implying that dUTPase deficiency led to prenatally lethal phenotype. Then, dissected embryos were analysed in order to pinpoint the developmental stage associated with the onset of lethality in *dut* -/- embryos. Summary of the numerous postnatal and embryonic samples generated by heterozygous D47 crossings is shown in Table 2. We found that homozygous *dut* -/- mutant embryos were only observable in such early embryonic stages as blastocyst stage (3.5 dpc). At later stages (at 8.5 dpc and 10.5 dpc), resorbed embryonic tissues were also seen which may be associated with the resorption of the *dut* -/- knock-out embryos. In these cases, the genotype could not be assessed because of isolation problems from maternal tissues.

DNA source	Genotype			Resorbed	No. Total
	+/+	+/-	-/-		
Postnatal	21	42	0	-	63
10.5 dpc	3	5	0	3	11
9.5 dpc	5	5	0	0	10
8.5 dpc	10	5	0	5	20
3.5 dpc	11	13	7	NA	31

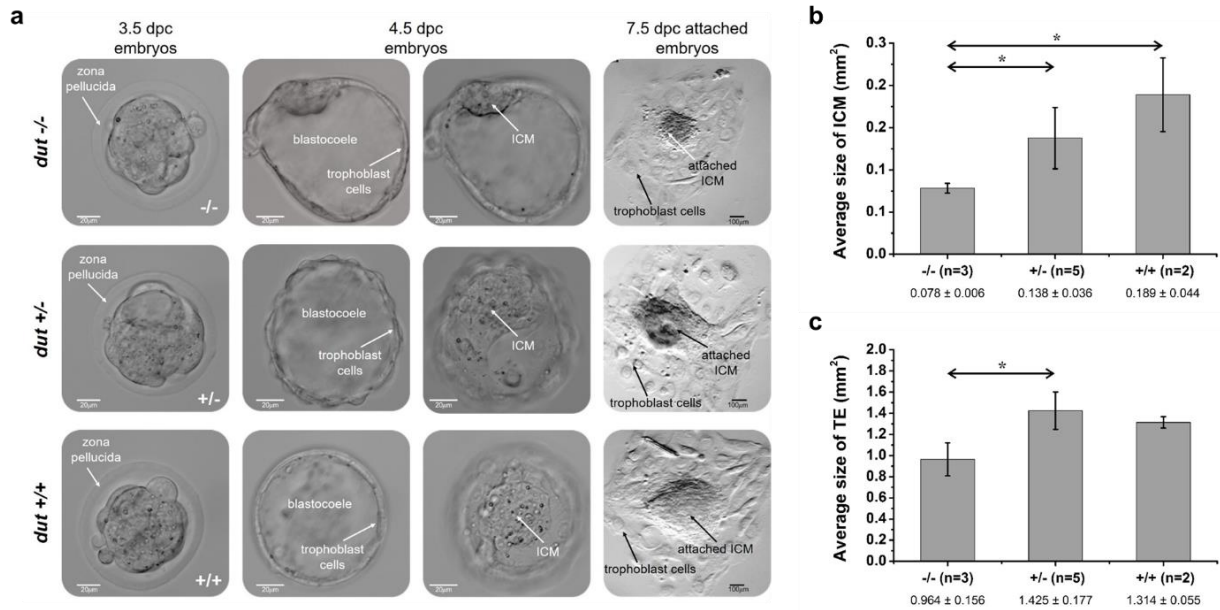
**Table 2.** Genotype analysis referring to *dut* gene after *dut* +/- D47 intercrossings. This table summarizes all the animals and embryos carrying *dut* +/+; *dut* +/- or *dut* -/- genotype, and also resorbed embryos with unknown genotype. NA, not applicable. Table was taken from our publication <sup>109</sup>.

#### 4.3.3. Alterations in embryonic development of *dut* mutants

Since new-born mice carrying the D47 frameshift mutation could not be obtained, dissected embryos had to be investigated. At post-implantation embryonic stages (8.5 dpc, 9.5 dpc, 10.5 dpc), heterozygotes did not show any visible developmental changes as compared to the wild-type. To gain information about *dut* -/- embryos, we started to examine blastocysts, which is a highly challenging experimental setup.

An outgrowth assay enabled us to get an insight into blastocyst formation, which represents the first cell lineage specification steps of embryo development, normally occurring before implantation into the endometrium of the uterus. Following heterozygous D47 breeding, blastocysts at 3.5 dpc were isolated and seeded to a gelatin-coated surface allowing *in vitro* culturing for some days. One day later (at 4.5 dpc), blastocysts were entirely attached and formed a characteristic structure including an inner cell mass (ICM) outgrowth surrounded by monolayer trophoblast cells (Figure 19A). Outgrowth of ICM and trophectoderm (TE) regions were calculated at 7.5 dpc, after four days in culture. In *dut* -/- embryos ICM clump size was significantly smaller than either *dut* +/+ WT or *dut* +/- variants (Figure 19B). Similarly, analysis of TE size revealed a significant decrease in *dut* -/- when compared to *dut* +/- embryos, while it substantially differed from *dut* +/+ blastocysts ( $p = 0.060$ ) (Figure 19C). In both aspects, either in ICM or TE size, there was no significant difference between *dut* +/- and *dut* +/+ embryos. Based on our results, formation of both ICM and trophectoderm are impeded in dUTPase deficient blastocysts.

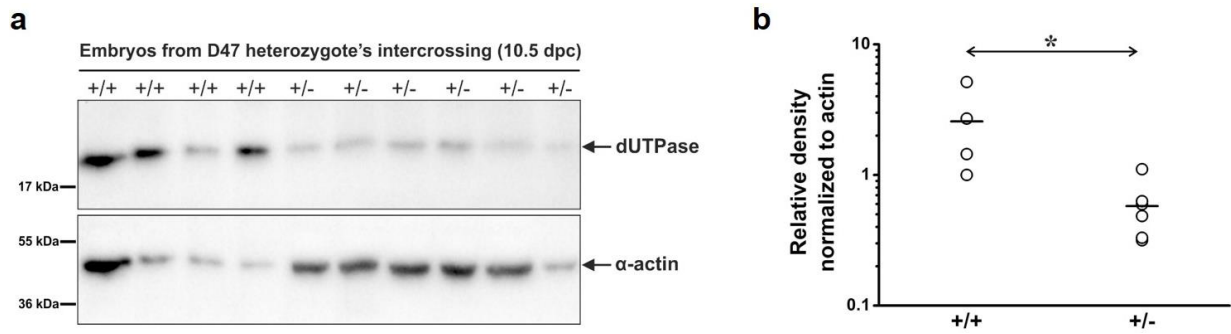




**Figure 19.** ICM outgrowth assay for blastocysts produced by D47 heterozygous crosses. (A) Representative phase contrast images of *dut*<sup>-/-</sup>; *dut*<sup>+/-</sup>; *dut*<sup>+/+</sup> embryos in *in vitro* culture. The first column shows the isolated embryos at 3.5 dpc surrounded by the zona pellucida. The second and third columns represent the attached embryos at 4.5 dpc, focusing on the trophoblast cells or the ICM in the blastocoele. In the last column, outgrowths at 7.5 dpc are seen. Scale bar represents 20  $\mu$ m, or 100  $\mu$ m for images in the last column. Average size of ICM (B) and trophectoderm (TE) (C) were quantified for embryos at 7.5 dpc. Error bars indicate standard deviation. Sample size is also provided with n number for all the genotypes. \*  $p < 0.05$ . Figure was taken from our publication<sup>109</sup>.

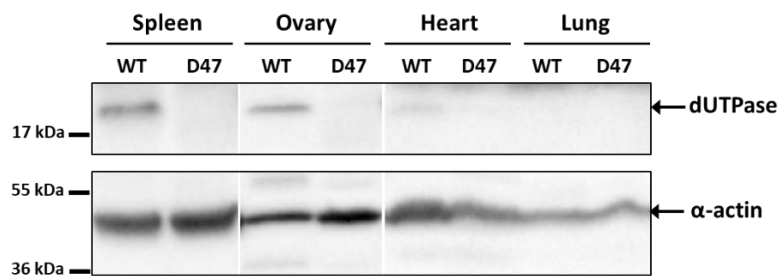
During our investigations, we did not observe any developmental differences or abnormalities between *dut*<sup>+/-</sup> heterozygotes and *dut*<sup>+/+</sup> wild-type mice regarding either embryos or adults. It was of immediate interest to address how the CRISPR-induced heterozygous genotype affects dUTPase protein level. Therefore, I performed Western blot analysis on total protein extracts prepared from *dut*<sup>+/+</sup> and *dut*<sup>+/-</sup> 10.5 dpc embryos produced by D47 heterozygous mating (Figure 20A). Densitometric analysis revealed that *dut*<sup>+/-</sup> heterozygous embryos possessed significantly reduced dUTPase protein content as compared to WT embryos (Figure 20B).





**Figure 20.** Expression level of dUTPase in 10.5 dpc embryos generated by D47 heterozygous intercrosses was examined by Western blot. **(A)** Membrane was probed with anti-dUTPase and anti- $\alpha$ -actin used as a loading control. **(B)** Relative density of dUTPase normalized to  $\alpha$ -actin. Every densitometric data point is shown corresponding to the presented blot, and mean values are indicated with horizontal lines. \*  $p < 0.05$ . Figure was taken from our publication <sup>109</sup>.

To investigate whether this difference is still observable in adult animals, additional Western blots were performed on organs derived from *dut* +/+ WT or *dut* +/- heterozygous D47 mice. In some tissues, such as spleen and ovary, dUTPase protein could be clearly detected in WT samples, while it was absent in D47 heterozygotes (Figure 21). However, in other mouse tissues, no signal (in case of lung) or only weak band (in case of heart) could be found hampering correct comparison of WT and D47 strains. These results are in good agreement with the recent finding that expression pattern of dUTPase is highly organ-specific <sup>114</sup>. Further details of gene expression may be explored using quantitative real-time PCR (qPCR) and proper statistical analysis.



**Figure 21.** Comparison of dUTPase protein level in WT and D47 mouse tissues such as spleen, ovary, heart and lung. Development against  $\alpha$ -actin was applied on the Western blot.

## 5. Discussion

Different fields of biology such as phage genetics, viral infection mechanisms, antibody affinity maturation or even chemotherapeutic approaches in cancer treatment heavily rely on genomic uracil occurrence and its effect on genomic stability (discussed in sections 1.3. and 1.6. in detail). It has been also argued that uracil accumulation in DNA may play a more decisive role in genomic instability and cytotoxicity than the induced uracil-excision repair<sup>59,73</sup>. Hence, specific detection of uracil containing DNA is of high importance for ongoing research of numerous, significant physiological and pathological events.

With my colleagues, we developed a new dot-blot based method for relative quantification of genomic uracil content using a catalytically inactive, flag-tagged  $\Delta$ UNG construct<sup>108</sup> (shown in Figure 3). After primary and secondary antibody labelling, U-DNA levels can be calculated *via* interpolation to a standard calibration curve, which could be obtained using uracilated DNA derived from (*dut*-, *ung*-) *E. coli* cells. We successfully validated our assay on different bacterial strains and also on the eukaryotic S2 cell line. Combination of *ung*- genetic background and treatment with well-known drugs targeting *de novo* thymidylate biosynthesis pathways (MTX, RTX, 5FdUR) led to significantly increased genomic uracil level compared to non-treated cells (Figure 4). When we proceeded to measure U-DNA level alterations using a human colon carcinoma cell line (HCT116), we found that similar drug treatments on their own do not lead to elevated U-DNA level. The molecular mechanism of action of these drugs is expected to generate increased amount of uracil residues *via* perturbations of cellular nucleotide pools, although efficient uracil excision repair could quickly erase the nascent uracils from DNA. Therefore, apart from the rare exceptions such as the *Drosophila* S2 embryonic tissues or other cases discussed in section 1.6., genomic uracil content is very low in cells with normal DNA repair and enzyme repertoire. To allow direct capture of the nascently appearing uracil moieties in DNA, inhibition of the major uracil-recognizing factor, UNG was important. Using dual treatment (UGI expression and 5FdUR drug addition) we could obtain and detect significantly elevated U-DNA content, as compared to non-treated cells (Figure 5). Based on our results, dot-blot based U-DNA quantification assay was proven to be robust, sensitive and does not require costly equipment. Hence, it shows several advantages over other quantification methods (using e.g. PCR<sup>101</sup> or LC/MS/MS<sup>99</sup>, described in section 1.7.) published earlier. Importantly, it examines the entire genome, thus independent on the distribution of uracil, and does not require enzymatic or complex, labor-intensive sample preparation which may influence the reliability of the results.

In addition, our  $\Delta$ UNG sensor-based method uniquely allows *in situ* microscopic visualization of U-DNA, either indirectly *via* immunocytochemistry against the Flag tag or directly *via* the fluorescent DsRed signal. To prove this issue, U-DNA staining was compared in CJ236 (*dut*<sup>-</sup>, *ung*<sup>-</sup>) and XL1-Blue (*dut*<sup>+</sup>, *ung*<sup>+</sup>) *E. coli* cells (Figure 6). Using *ung*<sup>-/-</sup> MEF cells for transfection with uracil-rich or normal plasmids, the herein presented U-DNA sensor protein successfully detected cellular uracil-enriched plasmid aggregates colocalizing with DAPI staining (Figure 7). Encouraged by these results, we further sought to adapt the  $\Delta$ UNG-based U-DNA labeling approach to visualize genomic uracil residues in complex eukaryotic cells. A versatile labeling technique could be achieved by cloning a SNAP-tag in the C terminal end of  $\Delta$ UNG, yielding Flag- $\Delta$ UNG-SNAP. A wide variety of synthetic probes are commercially available which can be applied for irreversible labelling of SNAP fusion proteins, thus facilitating imaging with different, state-of-the-art fluorescent microscopy techniques. Comparison of Flag- $\Delta$ UNG-SNAP and Flag- $\Delta$ UNG-DsRed constructs demonstrated that the novel sensor shows also reliable U-DNA detection (Figure 8). Afterwards, an UNG-inhibited, human-derived, model cell line was created by retroviral transduction of human codon optimized UGI into HCT116 cells. Drug treatment with 5FdUR or RTX in this cell line elevated U-DNA level, interestingly to different extent (Figure 9). The difference might correspond to drug-specific mechanism of action, involving potential alterations in signalling processes, transcription regulation and the timing of cell-cycle arrest. It is well-known that both drugs inhibit TS, thereby facilitating dUTP incorporation into DNA, while the nucleotide analogue 5FdUR also leads to direct incorporation of 5FdUMP into the DNA<sup>15,51</sup>, as described in section 1.3. Genomic uracil and fluorouracil may have different effects on transcription and epigenetic regulation processes. It should be noted that our  $\Delta$ UNG-based method detects both uracil and also fluorouracil within the DNA<sup>15,64</sup>, which might also contribute to our findings. Although, the combination of UGI expression and drug treatments did not result in any observable cell death, differences in cell-cycle progression were observed as compared to non-treated cells. Consistently, the 5FdUR treatment was reported to cause an S-phase arrest in the second cycle<sup>59,73</sup>, while the actual time point of cell-cycle arrest upon RTX treatment is still controversial<sup>115,116,117</sup>. Further work including cell cycle analysis with flow cytometry is needed to be done to elicit this question.

Using the generated UGI-expressing cell line, the Flag- $\Delta$ UNG-SNAP construct was successfully applied for *in situ* staining of genomic uracil in 5FdUR treated cells (Figure 10). In addition, U-DNA staining could be carried out with cross-linking PFA fixation of cells, instead of

Carnoy fixation. PFA is a routinely used fixative compatible with most antibody-based staining procedures, thus suitable for multi-color imaging in colocalization studies, in contrast to Carnoy. Genomic uracil distribution pattern was also investigated by super-resolution microscopy using appropriate SNAP substrates for either STED or dSTORM applications (Figure 11). A characteristic pattern could be observed in drug (5FdUR or RTX) treated cells, that differed from U-DNA signal of non-treated, UGI-expressing cells, as revealed by dSTORM imaging. Uracil staining showed signal enrichment at the nuclear membrane and areas surrounding the nucleoli in the case of drug-treated cells. To my knowledge, no previous technique has been reported for *in situ* microscopic detection of human genomic uracil. One similar method was published earlier, where a fluorescently tagged UdgX was applied for staining of uracils in *E. coli* DNA<sup>118</sup>, but this tool was not further extended for detection of uracils within the highly complex chromatin of human cells.

To gain sequence-specific information of genomic uracil distribution, I developed the U-DNA-Seq method. Briefly, U-DNA immunoprecipitation was carried out applying the Flag-tagged, catalytically inactive  $\Delta$ UNG sensor in purified and fragmented gDNA, followed by pull-down with anti-FLAG beads before sequencing (for details see section 3.2.4. and Figure 12). Genome-wide uracil distribution data could be obtained by NGS, and requires careful data analysis. U-DNA-Seq is a direct, feasible alternative to the recently published UPD-seq<sup>119</sup>, Excision-seq<sup>120</sup>, or dU-seq<sup>121</sup> methods, all of which allows only indirect detection requiring one or more auxiliary chemical or enzymatic step(s). Only these methods have the potential thus far to map genome-wide distribution of uracil within isolated gDNA based on NGS, and only dU-seq was used in the context of human genome. U-DNA-Seq might be beneficial for investigation of drug (such as 5FdUR or RTX) effects on genomic uracil distribution, in order to gain novel insights into the mechanism of their action. Moreover, colocalization studies with histone markers or other chromatin factors have the potential to reinforce the findings achieved by U-DNA-Seq, which may shed light on several important biological phenomena involving uracil-DNA related processes. Such research is currently being conducted in our team.

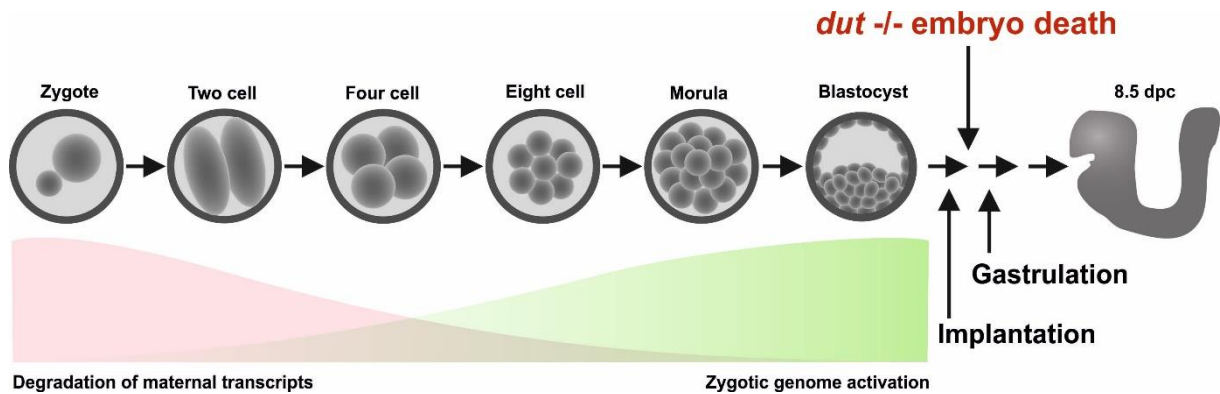
Generally, the cellular dUTP/dTTP ratio is under fine-tuned regulation of dUTPase action, predominantly linked to replication. Importance of efficient dUTPase enzymatic function was underlined by numerous gene disruption experiments in different organisms (detailed in section 1.4.). Although the essentiality of this enzyme was investigated by gene silencing approaches in cell cultures, knock-out studies on dUTPases have not yet been published for any

mammalian species. Therefore, we attempted to generate dUTPase knock-out model mice using CRISPR/Cas9-mediated gene editing to study the effects of dUTPase deficiency *in vivo*.

After validation of the chosen sgRNA and Cas9 mRNA by *in vitro* surveyor assay on cotransfected MEF cells, CRISPR/Cas9 system was used to produce dUTPase knock-out mice, according to the procedure presented in Figure 15. Two founder animals (#2 and #4) were generated harboring distinct CRISPR/Cas9 editing events, namely six bp deletion and one base substitution (D6, M1) in mice #2; and 47 bp deletion (D47) in mice #4. CRISPR/Cas9-induced D47 mutation resulted in frameshift mutation on *dut* gene and consequently early stop codons (Figure 17). In the case of such mutation, aberrant mRNA transcripts containing premature stop codons is expected to be eliminated by nonsense-mediated mRNA decay, to avoid translation to truncated proteins<sup>122</sup>. Therefore, further experiments were conducted on later generations of the mouse #4 with D47 mutation. Genotyping analysis using semi-nested PCR of numerous generations derived from heterozygous D47 crossings revealed that deficiency of dUTPase leads to embryonic lethality. Upon investigation of dissected embryos at different developmental stages, we found homozygous *dut* <sup>-/-</sup> mutant embryos only in blastocyst stage (3.5 dpc), but not at later stages, as shown in Table 2. Differences in the development of blastocysts with altered *dut* status were analysed in *in vitro* cultures. In *dut* <sup>-/-</sup> embryos, outgrowth of both ICM and TE regions were significantly reduced as compared to either *dut* <sup>+/+</sup> WT or *dut* <sup>+/-</sup> variants (Figure 19), indicating developmental defects in the absence of dUTPase. In contrast, no significant difference between *dut* <sup>+/-</sup> and *dut* <sup>+/+</sup> embryos were detected neither in blastocyst outgrowth assay, nor visually on dissected embryos at later embryonic stages (8.5 dpc, 9.5 dpc, 10.5 dpc). Examination by Western blots showed that *dut* <sup>+/-</sup> heterozygous embryos contain significantly reduced dUTPase protein level as compared to WT embryos (Figure 20). This difference might also apply for adult animals, but this needs further organ-specific experimental evidence. Regarding adulthood, progeny of D47 strains did not show any gross or fertility abnormalities, or systemic diseases during our study. Taken together, *dut* <sup>+/-</sup> heterozygosity and related potential suboptimal dUTPase protein level did not result in any visible phenotypic change in mice. Still, sensitivity of *dut* <sup>+/-</sup> mice to drug treatments targeting cellular dNTP pool have not yet been investigated.

Our results showed that knock-out of dUTPase leads to early embryonic lethality, which occurs in around or shortly after implantation into the uterus of mice (Figure 22). According to literature, the maternal-to-zygotic transition, including degradation of maternally deposited mRNA and protein sets, and the parallel transcriptional activation of zygotic genome, is

carried out by the blastocyst stage<sup>123,124</sup>. This argues for the hypothesis that the first several duplication cycles of embryogenesis can occur independently on dUTPase presence, while its absence impairs subsequent embryonic development. Based on *in vitro* outgrowth assays, the abnormal ICM and TE formation found in *dut*<sup>-/-</sup> blastocysts may lead to lethality around implantation. The fact that resorbed embryonic tissues were observed at later stages (at 8.5 dpc and 10.5 dpc), may argue for post-implantation death.



**Figure 22.** Schematic diagram shows the parallel elimination of maternal transcripts (red area) and activation of zygotic gene expression (green area) during mouse embryonic development; and the supposed time period when *dut*<sup>-/-</sup> embryo dies. Figure was taken from our publication<sup>109</sup>.

The herein presented *dut*<sup>+/-</sup> mouse models and early *dut*<sup>-/-</sup> embryos might be beneficial for understanding the exact mechanism underlying lethality. Importantly, genomic uracil occurrence could be investigated using our sensitive detection methods. We hypothesize that the complete loss of dUTPase enzymatic function (described in sections 1.3. and 1.4.) is the main cause responsible for the observed fatal outcome. Nevertheless, we can not exclude yet undescribed roles of dUTPase in protein-protein interactions which may also contribute. To elucidate this question, targeted point mutagenesis by cytidine deaminase coupled CRISPR/Cas9 system<sup>125</sup> (mentioned in section 1.8.) is needed to be done to gain a mouse model carrying catalytically inactive dUTPase protein (through exchange of aspartate to asparagine in the third conserved motif<sup>126</sup>). Targeting separately the different isoforms of dUTPase may also provide interesting results. Since depletion of Ung1 rescued lack of dUTPase in *E. coli*<sup>28</sup>, *S. cerevisiae*<sup>29</sup> and *C. elegans*<sup>30</sup>, multiple knock-outs may be worthwhile to be attempted. However, in the case of *T. brucei*, combination of dUTPase and UNG depletion led to worse *scenario* compared to cells deficient in only one of these activities<sup>127</sup>. Mouse models for UNG; SMUG1 and double knock-out have been previously reported<sup>94,128</sup>, thus their crossing with *dut*<sup>+/-</sup> mouse may lead to viable *dut*<sup>-/-</sup> mouse model for further investigations in the future.

## 6. Conclusion

In my PhD thesis, I have presented a novel, reliable detection method providing either qualitative or quantitative knowledge of U-DNA metabolism. It was demonstrated here that our dot-blot based U-DNA measurement technique relying on a catalytically inactive UNG2 ( $\Delta$ UNG) sensor protein enables quantification of genomic uracil level in different biological samples. I also verified that our U-DNA sensor-based method allows *in situ* cellular detection of uracils in bacterial or extrachromosomal DNA through an immunocytochemical approach.

This specific U-DNA sensor construct and staining method was further developed, which allowed imaging of genomic uracil residues within DNA of complex eukaryotic cells using confocal as well as super-resolution microscopy. I also managed to establish a novel HCT116 cell line stably expressing the UNG-inhibitor UGI protein, by retroviral transduction. Drug (5FdUR or RTX) treatment of the UNG-inhibited cells resulted in significantly elevated U-DNA level, thus gaining an adequate model system for investigation of nascent genomic uracil distribution. To my best knowledge this is the first time that a labeling technique applicable for *in situ* microscopic detection of human genomic uracil residues, is presented. Super-resolution imaging by dSTORM revealed that drug treated cells showed different U-DNA pattern than non-treated cells. Namely, signal enrichment could be observed at the nuclear membrane and areas surrounding the nucleoli in the case of drug-treated cells. Colocalization studies are needed to be done in the future to explore potential correlations between U-DNA distribution pattern and localization pattern of histone markers or other chromatin factors. Importantly, I have also described an adequate DNA immunoprecipitation method (U-DNA-Seq) providing sequence-specific information of genomic uracil distribution by next-generation sequencing.

We also aimed to investigate the physiological role of dUTPase *in vivo*, applying CRISPR/Cas9-mediated gene editing in mouse model. Based on our results, we could conclude that deficiency of dUTPase leads to early embryonic lethality in mice. Homozygous *dut*  $-/-$  mutant embryos were only found in blastocyst stage, and we suggest that further developmental abnormalities may lead to lethality occurring in around or shortly after implantation. Considering the timing of the maternal-to-zygotic transition in mice, the first several duplication cycles of embryogenesis may occur independently on dUTPase expression. Since U-DNA occurrence is involved in the wide range of physiological and pathological processes, the herein presented *dut*  $+/-$  mouse model and early *dut*  $-/-$  embryos might be beneficial to gain new insights into the details of U-DNA metabolism.

## References

1. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–15 (1993).
2. Grafstrom, R. H., Tseng, B. Y. & Goulian, M. The incorporation of uracil into animal cell DNA in vitro. *Cell* **15**, 131–40 (1978).
3. Vértessy, B. G. & Tóth, J. Keeping uracil out of DNA: physiological role, structure and catalytic mechanism of dUTPases. *Acc. Chem. Res.* **42**, 97–106 (2009).
4. Almatarneh, M. H., Flinn, C. G. & Poirier, R. A. Mechanisms for the deamination reaction of cytosine with H<sub>2</sub>O/OH(–) and 2H<sub>2</sub>O/OH(–): a computational study. *J. Chem. Inf. Model.* **48**, 831–43 (2008).
5. Greenman, C. *et al.* UKPMC Funders Group Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
6. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–21 (2013).
7. Mathews, C. K. DNA precursor metabolism and genomic stability. *FASEB J.* **20**, 1300–14 (2006).
8. Darè, E., Zhang, L. H., Jenssen, D. & Bianchi, V. Molecular analysis of mutations in the hprt gene of V79 hamster fibroblasts: effects of imbalances in the dCTP, dGTP and dTTP pools. *J. Mol. Biol.* **252**, 514–21 (1995).
9. Jackson, R. C. The regulation of thymidylate biosynthesis in Novikoff hepatoma cells and the effects of amethopterin, 5-fluorodeoxyuridine, and 3-deazauridine. *J. Biol. Chem.* **253**, 7440–6 (1978).
10. 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.* **11**, 282–98 (2014).
11. Tinkelenberg, B. A., Hansbury, M. J. & Ladner, R. D. dUTPase and uracil-DNA glycosylase are central modulators of antifolate toxicity in *Saccharomyces cerevisiae*. *Cancer Res.* **62**, 4909–15 (2002).
12. Chen, V. J. *et al.* Preclinical cellular pharmacology of LY231514 (MTA): a comparison with methotrexate, LY309887 and raltitrexed for their effects on intracellular folate and nucleoside triphosphate pools in CCRF-CEM cells. *Br. J. Cancer* **78 Suppl 3**, 27–34 (1998).
13. Grogan, B. C., Parker, J. B., Guminski, A. F. & Stivers, J. T. Effect of the thymidylate synthase inhibitors on dUTP and TTP pool levels and the activities of DNA repair glycosylases on uracil and 5-fluorouracil in DNA. *Biochemistry* **50**, 618–27 (2011).
14. Koehler, S. E. & Ladner, R. D. Small interfering RNA-mediated suppression of dUTPase sensitizes cancer cell lines to thymidylate synthase inhibition. *Mol Pharmacol* **66**, 620–626 (2004).
15. Pettersen, H. S. *et al.* UNG-initiated base excision repair is the major repair route for 5-fluorouracil in DNA, but 5-fluorouracil cytotoxicity depends mainly on RNA incorporation. *Nucleic Acids Res.* **39**, 8430–44 (2011).
16. Yoshioka, A. *et al.* Deoxyribonucleoside triphosphate imbalance. 5-Fluorodeoxyuridine-induced DNA double strand breaks in mouse FM3A cells and the mechanism of cell death. *J. Biol. Chem.* **262**, 8235–41 (1987).
17. Elstein, K. H. *et al.* Nucleoside-mediated mitigation of 5-fluorouracil-induced toxicity in synchronized murine erythroleukemic cells. *Toxicol. Appl. Pharmacol.* **146**, 29–39 (1997).
18. Ladner, R. D., McNulty, D. E., Carr, S. A., Roberts, G. D. & Caradonna, S. J. Characterization of distinct nuclear and mitochondrial forms of human deoxyuridine



- triphosphate nucleotidohydrolase. *J. Biol. Chem.* **271**, 7745–51 (1996).
19. Ladner, R. D. & Caradonna, S. J. The human dUTPase gene encodes both nuclear and mitochondrial isoforms. Differential expression of the isoforms and characterization of a cDNA encoding the mitochondrial species. *J. Biol. Chem.* **272**, 19072–80 (1997).
  20. el-Hajj, H. H., Zhang, H. & Weiss, B. Lethality of a dut (deoxyuridine triphosphatase) mutation in *Escherichia coli*. *J. Bacteriol.* **170**, 1069–75 (1988).
  21. Gadsden, M. H., McIntosh, E. M., Game, J. C., Wilson, P. J. & Haynes, R. H. dUTP pyrophosphatase is an essential enzyme in *Saccharomyces cerevisiae*. *EMBO J.* **12**, 4425–31 (1993).
  22. Erdélyi, P. *et al.* Shared developmental roles and transcriptional control of autophagy and apoptosis in *Caenorhabditis elegans*. *J. Cell Sci.* **124**, 1510–8 (2011).
  23. Castillo-Acosta, V. M., Estévez, A. M., Vidal, A. E., Ruiz-Perez, L. M. & González-Pacanowska, D. Depletion of dimeric all- $\alpha$  dUTPase induces DNA strand breaks and impairs cell cycle progression in *Trypanosoma brucei*. *Int. J. Biochem. Cell Biol.* **40**, 2901–13 (2008).
  24. Dubois, E. *et al.* Homologous recombination is stimulated by a decrease in dUTPase in *Arabidopsis*. *PLoS One* **6**, e18658 (2011).
  25. Pecsí, I. *et al.* The dUTPase enzyme is essential in *Mycobacterium smegmatis*. *PLoS One* **7**, e37461 (2012).
  26. Muha, V. *et al.* Uracil-containing DNA in *Drosophila*: stability, stage-specific accumulation, and developmental involvement. *PLoS Genet.* **8**, e1002738 (2012).
  27. Horváth, A., Békési, A., Muha, V., Erdélyi, M. & Vértessy, B. G. Expanding the DNA alphabet in the fruit fly: uracil enrichment in genomic DNA. *Fly (Austin)*. **7**, 23–7 (2013).
  28. Tye, B. K., Chien, J., Lehman, I. R., Duncan, B. K. & Warner, H. R. Uracil incorporation: a source of pulse-labeled DNA fragments in the replication of the *Escherichia coli* chromosome. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 233–7 (1978).
  29. Guillet, M., Van Der Kemp, P. A. & Boiteux, S. dUTPase activity is critical to maintain genetic stability in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **34**, 2056–66 (2006).
  30. Dengg, M. *et al.* Abrogation of the CLK-2 checkpoint leads to tolerance to base-excision repair intermediates. *EMBO Rep.* **7**, 1046–51 (2006).
  31. Kunkel, T. A., Bebenek, K. & McClary, J. Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.* **204**, 125–39 (1991).
  32. Kerepesi, C. *et al.* Life without dUTPase. *Front. Microbiol.* **7**, 1768 (2016).
  33. Alam, M. S., Moriyama, H. & Matsumoto, M. Inhibition of Dr-dut gene causes DNA damage in planarian. *Mol. Reprod. Dev.* **85**, 188–196 (2018).
  34. Dos Santos, R. S. *et al.* dUTPase (DUT) Is Mutated in a Novel Monogenic Syndrome With Diabetes and Bone Marrow Failure. *Diabetes* **66**, 1086–1096 (2017).
  35. Canman, C. E. *et al.* Induction of resistance to fluorodeoxyuridine cytotoxicity and DNA damage in human tumor cells by expression of *Escherichia coli* deoxyuridinetriphosphatase. *Cancer Res.* **54**, 2296–8 (1994).
  36. Canman, C. E., Lawrence, T. S., Shewach, D. S., Tang, H. Y. & Maybaum, J. Resistance to fluorodeoxyuridine-induced DNA damage and cytotoxicity correlates with an elevation of deoxyuridine triphosphatase activity and failure to accumulate deoxyuridine triphosphate. *Cancer Res.* **53**, 5219–24 (1993).
  37. Wilson, P. M., LaBonte, M. J., Lenz, H.-J., Mack, P. C. & Ladner, R. D. Inhibition of dUTPase induces synthetic lethality with thymidylate synthase-targeted therapies in non-small cell lung cancer. *Mol. Cancer Ther.* **11**, 616–28 (2012).
  38. Merényi, G. *et al.* 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* **30**, 369–90 (2011).

39. Studebaker, A. W., Lafuse, W. P., Kloesel, R. & Williams, M. V. Modulation of human dUTPase using small interfering RNA. *Biochem. Biophys. Res. Commun.* **327**, 306–10 (2005).
40. Hagenkort, A. *et al.* dUTPase inhibition augments replication defects of 5-Fluorouracil. *Oncotarget* **8**, 23713–23726 (2017).
41. Ladner, R. D. *et al.* dUTP nucleotidohydrolase isoform expression in normal and neoplastic tissues: association with survival and response to 5-fluorouracil in colorectal cancer. *Cancer Res.* **60**, 3493–503 (2000).
42. Kawahara, A. *et al.* Higher expression of deoxyuridine triphosphatase (dUTPase) may predict the metastasis potential of colorectal cancer. *J. Clin. Pathol.* **62**, 364–9 (2009).
43. Chen, C.-W. *et al.* The Impact of dUTPase on Ribonucleotide Reductase-Induced Genome Instability in Cancer Cells. *Cell Rep.* **16**, 1287–1299 (2016).
44. Webley, S. D., Hardcastle, A., Ladner, R. D., Jackman, A. L. & Aherne, G. W. Deoxyuridine triphosphatase (dUTPase) expression and sensitivity to the thymidylate synthase (TS) inhibitor ZD9331. *Br. J. Cancer* **83**, 792–9 (2000).
45. Wilson, P. M. *et al.* Novel opportunities for thymidylate metabolism as a therapeutic target. *Mol. Cancer Ther.* **7**, 3029–37 (2008).
46. Yano, W. *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.* **17**, 1683–1693 (2018).
47. Tormo-Más, M. A. *et al.* Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature* **465**, 779–82 (2010).
48. Szabó, J. E. *et al.* Highly potent dUTPase inhibition by a bacterial repressor protein reveals a novel mechanism for gene expression control. *Nucleic Acids Res.* **42**, 11912–20 (2014).
49. Nyíri, K. *et al.* Structural model of human dUTPase in complex with a novel proteinaceous inhibitor. *Sci. Rep.* **8**, 4326 (2018).
50. Benedek, A. *et al.* The Role of a Key Amino Acid Position in Species-Specific Proteinaceous dUTPase Inhibition. *Biomolecules* **9**, 1–15 (2019).
51. Longley, D. B., Harkin, D. P. & Johnston, P. G. 5-fluorouracil: mechanisms of action and clinical strategies. *Nat. Rev. Cancer* **3**, 330–8 (2003).
52. Kim, Y.-J. & Wilson, D. M. Overview of base excision repair biochemistry. *Curr. Mol. Pharmacol.* **5**, 3–13 (2012).
53. Jacobs, A. L. & Schär, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121**, 1–20 (2012).
54. Khodursky, A., Guzmán, E. C. & Hanawalt, P. C. Thymineless Death Lives On: New Insights into a Classic Phenomenon. *Annu. Rev. Microbiol.* **69**, 247–63 (2015).
55. Kunz, C. *et al.* Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil. *PLoS Biol.* **7**, e91 (2009).
56. Meyers, M., Wagner, M. W., Hwang, H. S., Kinsella, T. J. & Boothman, D. A. Role of the hMLH1 DNA mismatch repair protein in fluoropyrimidine-mediated cell death and cell cycle responses. *Cancer Res.* **61**, 5193–201 (2001).
57. Meyers, M. *et al.* DNA mismatch repair-dependent response to fluoropyrimidine-generated damage. *J. Biol. Chem.* **280**, 5516–5526 (2005).
58. Germann, M. W., Johnson, C. N. & Spring, A. M. Recognition of damaged DNA: structure and dynamic markers. *Med. Res. Rev.* **32**, 659–83 (2012).
59. Huehls, A. M. *et al.* Genomically Incorporated 5-Fluorouracil that Escapes UNG-Initiated Base Excision Repair Blocks DNA Replication and Activates Homologous Recombination. *Mol. Pharmacol.* **89**, 53–62 (2016).
60. Dubois, E. *et al.* Homologous recombination is stimulated by a decrease in dUTPase in

- Arabidopsis. *PLoS One* **6**, e18658 (2011).
61. Schormann, N., Ricciardi, R. & Chattopadhyay, D. Uracil-DNA glycosylases-structural and functional perspectives on an essential family of DNA repair enzymes. *Protein Sci.* **23**, 1667–85 (2014).
  62. Slupphaug, G. *et al.* Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase. *Biochemistry* **34**, 128–38 (1995).
  63. Krokan, H. E., Drabløs, F. & Slupphaug, G. Uracil in DNA--occurrence, consequences and repair. *Oncogene* **21**, 8935–48 (2002).
  64. Warner, H. R. & Rockstroh, P. A. Incorporation and excision of 5-fluorouracil from deoxyribonucleic acid in *Escherichia coli*. *J. Bacteriol.* **141**, 680–6 (1980).
  65. Dizdaroglu, M., Karakaya, A., Jaruga, P., Slupphaug, G. & Krokan, H. E. Novel activities of human uracil DNA N-glycosylase for cytosine-derived products of oxidative DNA damage. *Nucleic Acids Res.* **24**, 418–22 (1996).
  66. Pettersen, H. S. *et al.* Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. *Nucleic Acids Res.* **35**, 3879–92 (2007).
  67. Bharti, S. K. & Varshney, U. Analysis of the impact of a uracil DNA glycosylase attenuated in AP-DNA binding in maintenance of the genomic integrity in *Escherichia coli*. *Nucleic Acids Res.* **38**, 2291–301 (2010).
  68. Otterlei, M. *et al.* Post-replicative base excision repair in replication foci. *EMBO J.* **18**, 3834–44 (1999).
  69. Welsh, S. J., Hobbs, S. & Aherne, G. W. Expression of uracil DNA glycosylase (UDG) does not affect cellular sensitivity to thymidylate synthase (TS) inhibition. *Eur. J. Cancer* **39**, 378–87 (2003).
  70. Andersen, S. *et al.* 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* **26**, 547–55 (2005).
  71. Luo, Y., Walla, M. & Wyatt, M. D. Uracil incorporation into genomic DNA does not predict toxicity caused by chemotherapeutic inhibition of thymidylate synthase. *DNA Repair (Amst)*. **7**, 162–9 (2008).
  72. Fischer, J. A., Muller-Weeks, S. & Caradonna, S. J. Fluorodeoxyuridine modulates cellular expression of the DNA base excision repair enzyme uracil-DNA glycosylase. *Cancer Res.* **66**, 8829–8837 (2006).
  73. Yan, Y. *et al.* Inhibition of uracil DNA glycosylase sensitizes cancer cells to 5-fluorodeoxyuridine through replication fork collapse-induced DNA damage. *Oncotarget* **7**, 59299–59313 (2016).
  74. Kavli, B. *et al.* 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.* **277**, 39926–39936 (2002).
  75. Nilsen, H. *et al.* Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase. *EMBO J.* **20**, 4278–86 (2001).
  76. Slupphaug, G. *et al.* Uracil in DNA and its processing by different DNA glycosylases. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 563–568 (2008).
  77. Hardeland, U., Kunz, C., Focke, F., Szadkowski, M. & Schär, P. Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2. *Nucleic Acids Res.* **35**, 3859–67 (2007).
  78. Cortellino, S. *et al.* Demethylation by Linked Deamination-Base Excision Repair. *Cell* **146**, 67–79 (2011).
  79. Bellacosa, A. & Drohat, A. C. Role of base excision repair in maintaining the genetic

- and epigenetic integrity of CpG sites. *DNA Repair (Amst)*. **32**, 33–42 (2015).
80. Sang, P. B., Srinath, T., Patil, A. G., Woo, E. & Varshney, U. A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily. *Nucleic Acids Res.* **43**, 8452–63 (2015).
  81. Wang, Z. & Mosbaugh, D. W. Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. *J. Biol. Chem.* **264**, 1163–71 (1989).
  82. Serrano-Heras, G., Bravo, A. & Salas, M. Phage phi29 protein p56 prevents viral DNA replication impairment caused by uracil excision activity of uracil-DNA glycosylase. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 19044–9 (2008).
  83. Kiljunen, S. *et al.* Yersiniophage phiR1-37 is a tailed bacteriophage having a 270 kb DNA genome with thymidine replaced by deoxyuridine. *Microbiology* **151**, 4093–102 (2005).
  84. Uchiyama, J. *et al.* Intragenus generalized transduction in *Staphylococcus* spp. by a novel giant phage. *ISME J.* **8**, 1949–52 (2014).
  85. Katz, G. E., Price, a R. & Pomerantz, M. J. Bacteriophage PBS2-induced inhibition of uracil-containing DNA degradation. *J. Virol.* **20**, 535–8 (1976).
  86. Mol, C. D. *et al.* Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* **82**, 701–8 (1995).
  87. Wang, H.-C. *et al.* *Staphylococcus aureus* protein SAUGI acts as a uracil-DNA glycosylase inhibitor. *Nucleic Acids Res.* **42**, 1354–64 (2014).
  88. Yan, N., O'Day, E., Wheeler, L. A., Engelman, A. & Lieberman, J. HIV DNA is heavily uracilated, which protects it from autointegration. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9244–9 (2011).
  89. Kennedy, E. M. *et al.* Abundant non-canonical dUTP found in primary human macrophages drives its frequent incorporation by HIV-1 reverse transcriptase. *J. Biol. Chem.* **286**, 25047–55 (2011).
  90. Weil, A. F. *et al.* Uracil DNA glycosylase initiates degradation of HIV-1 cDNA containing misincorporated dUTP and prevents viral integration. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E448–57 (2013).
  91. Kavli, B., Otterlei, M., Slupphaug, G. & Krokan, H. E. Uracil in DNA--general mutagen, but normal intermediate in acquired immunity. *DNA Repair (Amst)*. **6**, 505–16 (2007).
  92. Waters, L. S. *et al.* Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol. Mol. Biol. Rev.* **73**, 134–54 (2009).
  93. Sarno, A. *et al.* Uracil-DNA glycosylase UNG1 isoform variant supports class switch recombination and repairs nuclear genomic uracil. *Nucleic Acids Res.* **47**, 4569–4585 (2019).
  94. Nilsen, H. *et al.* Gene-targeted mice lacking the Ung uracil-DNA glycosylase develop B-cell lymphomas. *Oncogene* **22**, 5381–6 (2003).
  95. Imai, K. *et al.* Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat. Immunol.* **4**, 1023–8 (2003).
  96. Kavli, B. *et al.* B cells from hyper-IgM patients carrying UNG mutations lack ability to remove uracil from ssDNA and have elevated genomic uracil. *J. Exp. Med.* **201**, 2011–21 (2005).
  97. Pettersen, H. S. *et al.* AID expression in B-cell lymphomas causes accumulation of genomic uracil and a distinct AID mutational signature. *DNA Repair (Amst)*. **25**, 60–71 (2015).
  98. 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.* **17**, 222–9 (2010).

99. Galashevskaya, A. *et al.* A robust, sensitive assay for genomic uracil determination by LC/MS/MS reveals lower levels than previously reported. *DNA Repair (Amst)*. **12**, 699–706 (2013).
100. Lari, S.-U., Chen, C.-Y., Vertéssy, 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 (Amst)*. **5**, 1407–20 (2006).
101. Horváth, A. & Vértessy, B. G. A one-step method for quantitative determination of uracil in DNA by real-time PCR. *Nucleic Acids Res.* **38**, 1–10 (2010).
102. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **169**, 5429–33 (1987).
103. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of Bacteria and Archaea. *Science (80-. )*. **327**, 167–170 (2010).
104. Deveau, H., Garneau, J. E. & Moineau, S. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* **64**, 475–93 (2010).
105. Barrangou, R. Diversity of CRISPR-Cas immune systems and molecular machines. *Genome Biol.* **16**, 247 (2015).
106. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–21 (2012).
107. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–4 (2016).
108. Róna, G. *et al.* Detection of uracil within DNA using a sensitive labeling method for in vitro and cellular applications. *Nucleic Acids Res.* **44**, e28 (2016).
109. Pálincás, H. L. *et al.* CRISPR/Cas9-Mediated Knock-Out of dUTPase in Mice Leads to Early Embryonic Lethality. *Biomolecules* **9**, (2019).
110. Rona, G. *et al.* PARP1-dependent recruitment of the FBXL10-RNF68-RNF2 ubiquitin ligase to sites of DNA damage controls H2A.Z loading. *Elife* **7**, 1–31 (2018).
111. Keppler, A. *et al.* A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **21**, 86–9 (2003).
112. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–6 (2013).
113. Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J. & Mateo, J. L. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLoS One* **10**, e0124633 (2015).
114. Rácz, G. A. *et al.* Evaluation of critical design parameters for RT-qPCR-based analysis of multiple dUTPase isoform genes in mice. *FEBS Open Bio* **9**, 1153–1170 (2019).
115. Blackledge, G. New developments in cancer treatment with the novel thymidylate synthase inhibitor raltitrexed ('Tomudex'). *Br. J. Cancer* **77 Suppl 2**, 29–37 (1998).
116. Ding, W.-X. *et al.* Raltitrexed increases radiation sensitivity of esophageal squamous carcinoma cells. *Cancer Cell Int.* **19**, 36 (2019).
117. Zhao, H., Zhang, Y., Sun, J., Zhan, C. & Zhao, L. Raltitrexed Inhibits HepG2 Cell Proliferation via G0/G1 Cell Cycle Arrest. *Oncol. Res.* **23**, 237–48 (2016).
118. Datta, M. *et al.* Development of mCherry tagged UdgX as a highly sensitive molecular probe for specific detection of uracils in DNA. *Biochem. Biophys. Res. Commun.* **518**, 38–43 (2019).
119. Sakhtemani, R. *et al.* Genome-wide mapping of regions preferentially targeted by the human DNA-cytosine deaminase APOBEC3A using uracil-DNA pulldown and sequencing. *J. Biol. Chem.* **294**, 15037–15051 (2019).
120. 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.* **24**, 1534–42 (2014).
121. Shu, X. *et al.* Genome-wide mapping reveals that deoxyuridine is enriched in the human centromeric DNA. *Nat. Chem. Biol.* **14**, 680–687 (2018).
  122. Baker, K. E. & Parker, R. Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Curr. Opin. Cell Biol.* **16**, 293–9 (2004).
  123. Li, L., Zheng, P. & Dean, J. Maternal control of early mouse development. *Development* **137**, 859–70 (2010).
  124. Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* **136**, 3033–42 (2009).
  125. Liang, P. *et al.* Effective gene editing by high-fidelity base editor 2 in mouse zygotes. *Protein Cell* **8**, 601–611 (2017).
  126. Barabás, O., Pongrácz, V., Kovári, J., Wilmanns, M. & Vértessy, B. G. Structural insights into the catalytic mechanism of phosphate ester hydrolysis by dUTPase. *J. Biol. Chem.* **279**, 42907–15 (2004).
  127. Castillo-Acosta, V. M. *et al.* Increased uracil insertion in DNA is cytotoxic and increases the frequency of mutation, double strand break formation and VSG switching in *Trypanosoma brucei*. *DNA Repair (Amst.)*. **11**, 986–95 (2012).
  128. Alsøe, L. *et al.* Uracil Accumulation and Mutagenesis Dominated by Cytosine Deamination in CpG Dinucleotides in Mice Lacking UNG and SMUG1. *Sci. Rep.* **7**, 7199 (2017).

**Annex**

- I.** Róna G, Scheer I, Nagy K, Pálincás HL, Tihanyi G, Borsos M, Békési A, Vértessy BG. Detection of uracil within DNA using a sensitive labeling method for *in vitro* and cellular applications. Nucleic Acids Res. 2016; 44(3):e28. doi: 10.1093/nar/gkv977.
- II.** Pálincás HL, Rác GA, Gál Z, Hoffmann OI, Tihanyi G, Róna G, Gócza E, Hiripi L, Vértessy BG. CRISPR/Cas9-Mediated Knock-Out of dUTPase in Mice Leads to Early Embryonic Lethality. Biomolecules. 2019; 9(4). pii: E136. doi: 10.3390/biom9040136.

## **Annex I.**



# Detection of uracil within DNA using a sensitive labeling method for *in vitro* and cellular applications

Gergely Róna<sup>1,2,\*</sup>, Ildikó Scheer<sup>1,2</sup>, Kinga Nagy<sup>1,2</sup>, Hajnalka L. Pálincás<sup>1,3</sup>,  
Gergely Tihanyi<sup>1,2</sup>, Máté Borsos<sup>1</sup>, Angéla Békési<sup>1</sup> and Beáta G. Vértessy<sup>1,2,\*</sup>

<sup>1</sup>Institute of Enzymology, RCNS, Hungarian Academy of Sciences, Magyar Tudósok Str. 2, H-1117 Budapest, Hungary, <sup>2</sup>Department of Applied Biotechnology and Food Sciences, Budapest University of Technology and Economics, Szt Gellért Square 4, H-1111 Budapest, Hungary and <sup>3</sup>Doctoral School of Multidisciplinary Medical Science, University of Szeged, H-6720 Szeged, Hungary

Received May 17, 2015; Revised September 10, 2015; Accepted September 15, 2015

## ABSTRACT

The role of uracil in genomic DNA has been recently re-evaluated. It is now widely accepted to be a physiologically important DNA element in diverse systems from specific phages to antibody maturation and *Drosophila* development. Further relevant investigations would largely benefit from a novel reliable and fast method to gain quantitative and qualitative information on uracil levels in DNA both *in vitro* and *in situ*, especially since current techniques does not allow *in situ* cellular detection. Here, starting from a catalytically inactive uracil-DNA glycosylase protein, we have designed several uracil sensor fusion proteins. The designed constructs can be applied as molecular recognition tools that can be detected with conventional antibodies in dot-blot applications and may also serve as *in situ* uracil-DNA sensors in cellular techniques. Our method is verified on numerous prokaryotic and eukaryotic cellular systems. The method is easy to use and can be applied in a high-throughput manner. It does not require expensive equipment or complex know-how, facilitating its easy implementation in any basic molecular biology laboratory. Elevated genomic uracil levels from cells of diverse genetic backgrounds and/or treated with different drugs can be demonstrated also *in situ*, within the cell.

## INTRODUCTION

Uracil, one of the nucleic acid bases present in RNA, is usually considered to be a mistake when appearing in DNA (1). Two independent pathways may lead to the presence of uracil in DNA. On the one hand, hydrolytic deamination of cytosine within the DNA, a rather frequent event,

results in numerous uracil moieties, which are mutagenic since these replacements, if left unrepaired, will lead to an exchange of a G:C base pair to a A:U (A:T) base pair (2). On the other hand, most polymerases cannot distinguish between deoxyuridine and deoxythymidine and will readily incorporate either of these two building blocks depending on the ratio of cellular dUTP and dTTP pools (3). The nucleotide pool is usually sanitized by dUTPases (in most cases, these enzymes are encoded by the *dut* gene), an enzyme family conserved from bacteria to human, to avoid such thymine-replacing uracil incorporation events (4–7). Genomic uracil is specifically recognized by representatives of the uracil-DNA glycosylase superfamily (UDG), cleaving the N-glycosidic bond between the pyrimidine ring and deoxyribose and resulting in apyrimidinic (AP) sites that are further processed by base excision repair (8,9). UNG, one of the four UDGs found in mammalian cells, specifically excise uracil bases from both double-stranded (dsDNA) and single-stranded DNA (ssDNA). *In vitro* the enzyme removes uracil in the order of preference ssU > U:G >> U:A (10–13). UNG activity is somewhat affected by the sequence context, having a slightly different affinity for uracil in A/T rich regions compared to G/C rich environment (12,14–16). To a lesser extent, bases formed from cytosine oxidation are also substrates of UNG (5-hydroxyuracil, isodialuric acid and alloxan; the latter only recognized by the human enzyme) (17,18). With a slower rate, 5-fluorouracil is also processed, however other larger 5-halouracils (like BrdU) are not recognized (19,20). A growing number of results show that UNG is somewhat capable of binding AP sites, but with a lower affinity compared to genomic uracil (14,21–23). No activity has been detected against normal DNA bases or against uracil in RNA (13) since RNA is excluded from the DNA-binding pocket due to unfavorable steric reasons (24,25).

Fine-tuned regulation of nucleotide pools is also of key importance for genomic stability. Inhibitors targeting path-

\*To whom correspondence should be addressed. Tel: +36 13 826 707; Fax: +36 14 665 465; Email: vertessy@mail.bme.hu  
Correspondence may also be addressed to Gergely Róna. Tel: +36 13 826 762; Fax: +36 14 665 465; Email: rona.gergely@ttk.mta.hu

ways involved in proper dUTP/dTTP pool maintenance, such as the *de novo* thymidylate biosynthesis pathways, induce thymine-less cell death and are a focus of cancer treatment (26). Importantly, genomic uracil may also appear under normal physiological conditions. In the most extreme cases of specific bacteriophages, such as in *Bacillus subtilis* PBS1 and PBS2 phages, and the *Yersinia enterocolitica*  $\Phi$ R1–37 phage, the phage DNA contains deoxyuridine but no deoxythymidine (27–30). Uracil in DNA was implicated as a key factor in B lymphocyte function during somatic hypermutation and class-switch recombination (31–33). The surprisingly high uracil content (estimated as >25 000 uracil/million bases) of reverse-transcribed HIV genomic DNA has been suggested to play an important role in the viral life cycle (34). DNA from fruit fly larvae and pupae also contains highly elevated levels of uracil (200–2000 uracil/million bases) (35,36).

As summarized above, several different fields in biology from phage genetics to lentiviral infection mechanisms, from antibody maturation and *Drosophila* development to chemotherapeutic approaches in cancer treatment heavily rely on genomic uracil occurrence. Hence, a reliable, fast, cheap and easy method to gain quantitative and qualitative information on uracil levels in DNA is of high importance for *in vitro* and *in vivo* studies. Currently available genomic uracil quantification methods vary in specificity, sensitivity and price. Even though LC/MS/MS based methods are sensitive, they need laborious, excessive sample preparation that involves nucleotide or uracil hydrolysis of the samples (37–42). Chemical modification of uracil moieties to enhance detection also provides a highly sensitive method but needs several steps in sample preparation (43,44). Real-time polymerase chain reaction (PCR)-based techniques reflect the uracil content only on a limited DNA fragment, and total genomic uracil content is calculated as an extrapolation based on the assumption that uracil residues are evenly distributed throughout the genome (45) — although this may not be always the case.

Most techniques reported to date that aim to quantify genomic uracil levels excise uracil from DNA during the process and do not allow *in situ* cellular detection. Interestingly, for detection of numerous other non-orthodox DNA bases such as 5-methylcytosine (46), 5-hydroxymethylcytosine (47), 5-hydroxymethyluracil (48), thymine dimers (49), 8-oxo-guanine (50) and 8-nitroguanine (51), antibodies have been described. To our knowledge, no such method has yet been reported for the uracil moieties in DNA.

In the present work, we therefore aimed at designing a uracil sensor by applying a catalytically inactive UNG uracil-DNA glycosylase, which is capable of binding to but not excising uracil (22). The uracil-recognizing UNG sensor was designed in such a way that it can be detected either with conventional antibodies in dot-blot applications or also *in situ* using an immunocytochemical approach. Our method is a relative quantification approach that delivers the sensitivity of MS based approaches, reaching a femtomol uracil detection limit. It may also be developed further as a ChIP like approach to gain position- and sequence-specific information on genomic uracil content. Performance of the herein described uracil sensor has been analyzed in dot-blot and immunocytochemical ap-

proaches, using prokaryotic CJ236 *Escherichia coli* [*dut*–, *ung*–] or BL21(DE3) *ung*-151 *E. coli* [*ung*–] and also eukaryotic (*Drosophila* and human) cell lines with altered base excision repair background (36,52). We have also analyzed cells treated with several different chemotherapeutic drugs, known to interfere with thymidylate biosynthesis and leading to increased uracil content in DNA. Our results are in good agreement with current reports from the literature verifying that our method is sensitive, cost-effective and adequate.

## MATERIALS AND METHODS

### Plasmid constructs and cloning

Human uracil-DNA glycosylase 2 (UNG2) cDNA was a generous gift of Professor Salvatore Caradonna (Department of Molecular Biology, University of Medicine and Dentistry of New Jersey) and was cloned into the XhoI/KpnI sites of the pDsRed-Monomer-N1 vector (Clontech, Mountain View, CA, USA), as described previously (53). DsRed-fused UNG2 was further PCR amplified and cloned into NdeI/XhoI sites of the vector pET-20b (Novagen, Merck Millipore, Billerica, MA, USA) with primers UNG\_F and UNG\_R (Supplementary Table S1). Point mutations (D154N and H277N) were created by the QuickChange mutagenesis method (Stratagene, Santa Clara, CA, USA) (with primers D154N\_F, D154N\_R and H277N\_F, H277N\_R respectively). UNG2 (D154N and H277N) lacking the first 84 amino acid from its N-terminus ( $\Delta$ UNG) was PCR amplified (with primers 1 $\times$ FLAG\_F, 3 $\times$ FLAG\_F and 1 $\times$ /3 $\times$ FLAG\_R) and was cloned into the NdeI/XhoI sites of the vector pET-15b (Novagen) FLAG-tagged (1 $\times$  or 3 $\times$ ) yielding the constructs 1 $\times$ FLAG- $\Delta$ UNG and 3 $\times$ FLAG- $\Delta$ UNG respectively. Constructs  $\Delta$ UNG-DsRed and FLAG- $\Delta$ UNG-DsRed were PCR amplified (with primers 1 $\times$ FLAG\_Ds\_F, Ds\_F and Ds\_R) and cloned into the NdeI/XhoI sites of the vector pET-20b. The vector expressing the human codon optimized UGI along with GFP (pLGC-hUgi) (54) was a kind gift of Michael D. Wyatt (South Carolina College of Pharmacy, University of South Carolina). Primers used in this study were synthesized by Eurofins MWG GmbH (Ebersberg, Germany) and are summarized in Supplementary Table S1. All constructs were verified by sequencing at Eurofins MWG GmbH.

### Cell culture and transfection

The MLH1-deficient (a mismatch repair deficient) human colorectal adenocarcinoma cell line, HCT116, was purchased from the European Collection of Cell Cultures (ECACC, Salisbury, UK). Mouse embryonic fibroblast (MEF) cells lacking functional UNG (55) were a generous gift from Dr Hilde Nilsen (University of Oslo). HCT116 cells were cultured in McCoy's 5A medium (Gibco, Life Technologies, Carlsbad, CA, USA) while MEF cells in DMEM/F12 HAM (Sigma); supplemented with 50  $\mu$ g/ml Penicillin-Streptomycin (Gibco) and 10% FBS (Gibco) in a humidified 37°C incubator with 5% CO<sub>2</sub> atmosphere. Schneider S2 cells (derived from *Drosophila melanogaster*)

were purchased from Gibco and were cultured in Schneider Insect Medium (Sigma) supplemented with 10% FBS (Gibco) and 50 µg/ml Penicillin-Streptomycin (Gibco) and kept in a 26°C incubator.

HCT116 cells were transfected with FuGENE® HD (Promega, Madison, WI, USA) in T25 tissue culture flasks according to the manufacturer's recommendation. For immunocytochemistry MEF cells were transfected in a 6-well plate with 4 µg of normal pEGFP-N1 (purified from XL1-Blue [*dut*+, *ung*+] *E. coli* cells) or uracil-rich pEGFP-N1 vector (purified from CJ236 [*dut*−, *ung*−] *E. coli* cells) and 12 µl FuGENE® HD transfection reagent according to the recommendation of the manufacturer. After 16 h cells were thoroughly trypsinized and washed with phosphate buffered saline (PBS) extensively (to wash away extracellular plasmid aggregates attached to the cell surface) and finally splitted onto 24-well plates containing poly-L-lysine coated cover glasses.

### Treatment of cells, DNA isolation and purification

**Plasmid DNA.** pEGFP-N1 plasmid (Clontech) was transformed into XL1-Blue [*dut*+, *ung*+] (Stratagene) or CJ236 [*dut*−, *ung*−] (NEB, Ipswich, MA, USA) *E. coli* strains. Cell cultures were grown for 16 h in Luria broth (LB) media supplemented with kanamycin at 37°C, and the plasmids were purified using PureYield™ Plasmid Midiprep Kit (Promega) according to the instructions of the manufacturer.

**Genomic DNA.** XL1-Blue, BL21(DE3) *ung-151* (56) and CJ236 *E. coli* strains were propagated in LB media at 37°C and were harvested at log-phase ( $A_{600\text{nm}} = 0.5$ ). BL21(DE3) *ung-151* cells were also grown either in the presence of 30.7 µM 5-fluoro-2'-deoxyuridine (5FdUR) or 200 µM 2'-deoxyuridine (dUR) or in the presence of both drugs. Genomic DNA was purified with MasterPure™ DNA Purification Kit (Epicentre, Madison, WI, USA), followed by an additional purification with the Genomic DNA Clean & Concentrator Kit (ZYMO Research, Irvine, CA, USA) using the recommendations of the manufacturer. *Drosophila* Schneider S2 cells were grown either in the absence or presence of 100 µM 5FdUR, 500 µM dUR or 10 µM methotrexate (MTX), 100 nM raltitrexed (RTX), 500 µM dUR for 48 h. Genomic DNA of S2 cells was purified as above. Forty hours before treatment, HCT116 cells were transfected either with the UGI-GFP expressing vector or with an empty vector expressing GFP alone (pEGFP-N1). Both transfected and non-transfected cells were grown for an additional 48 hours in the presence or absence of 20 µM 5FdUR before collecting them for genomic DNA purification described as above.

### Recombinant protein production

All UNG constructs were expressed in the *E. coli* BL21(DE3) *ung-151* strain and purified using Ni-NTA affinity resin (Qiagen, Hilden, Germany). Transformed cells growing in LB medium were induced at  $A_{600\text{nm}} = 0.6$  with 0.6 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) for 24 h at 18°C. Cells were harvested and lysed in

lysis buffer (50 mM TRIS-HCl, pH = 8.0, 300 mM NaCl, 0.5 mM ethylenediaminetetraacetic acid (EDTA), 0.1% Triton X-100, 10 mM β-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, 5 mM benzamidine, 1×cOmplete ULTRA™ EDTA free protease inhibitor cocktail tablet (Roche), 0.1 mg/ml lysozyme, 0.1 mg/ml DNase (Sigma, St. Louis, MO, USA) and 0.01 mg/ml RNase A (Invitrogen, Carlsbad, CA, USA)) assisted with sonication. Cell debris was pelleted by centrifugation at  $20\,000 \times g$  for 30 min. Supernatant was applied onto a Ni-NTA column and washed with a set of washing buffers: low salt buffer (50 mM HEPES, pH = 7.5, 30 mM KCl, 5 mM β-mercaptoethanol), high salt buffer (50 mM HEPES, pH = 7.5, 300 mM KCl, 5 mM β-mercaptoethanol) and very high salt buffer (50 mM HEPES, pH = 7.5, 500 mM NaCl, 40 mM Imidazole, 5 mM β-mercaptoethanol). UNG constructs were finally eluted with elution buffer (50 mM HEPES, pH = 7.5, 30 mM KCl, 300 mM imidazole, 5 mM β-mercaptoethanol) and dialyzed against the following buffer: 30 mM Tris-HCl, pH = 7.4, 140 mM NaCl, 0.01% Tween-20, 1 mM EDTA, 15 mM β-mercaptoethanol.

### Assay for testing UNG activity

A total of 150 ng of SmaI (NEB) linearized pEGFP-N1 or uracil-rich pEGFP-N1 vector was incubated with 0.02 µg of each UNG constructs (hUNG2-DsRed WT, hUNG2-DsRed, 1×-Flag-ΔUNG, 3×-Flag-ΔUNG, Flag-ΔUNG-DsRed, ΔUNG-DsRed) for 1 h at 37°C in a final volume of 11 µl, in Endo IV buffer (50 mM Tris-acetate (pH 7.5), 50 mM KCl, 1 mM EDTA, 0.05% Triton X-100; Fermentas, Waltham, MA, USA). One unit of Endonuclease IV (Fermentas) was added to the reaction mixture and was further incubated for 1 h at 37°C. Endonuclease IV (Endo IV) is apurinic/AP endonuclease that will hydrolyse AP sites in DNA. AP sites are cleaved at the phosphodiester bond that is 5' to the lesion leaving a hydroxyl group at the 3' terminus and a deoxyribose 5'-phosphate at the 5' terminus. UNG and Endo IV treatment leads to nicks in the phosphodiester backbone of the DNA, resulting in extensive fragmentation of uracil-rich DNA. Reaction was stopped by adding 4 µl of inactivation mixture containing 2.5% sodium dodecyl sulfate, 2.5 mg/ml Proteinase K (Sigma) and 1.5× concentrated DNA Loading Dye (Fermentas). Standard agarose gel electrophoresis was performed in a 1% gel.

### Electrophoretic mobility shift assay (EMSA)

A total of 100 ng of SmaI (NEB) linearized pEGFP-N1 or uracil-rich pEGFP-N1 vector was incubated with a series of two-fold dilution of the different UNG constructs (hUNG2-DsRed WT, hUNG2-DsRed, 1×-Flag-ΔUNG, 3×-Flag-ΔUNG, Flag-ΔUNG-DsRed, ΔUNG-DsRed) starting with 1 µg of protein, for 5 min at room temperature in UNG buffer (30 mM TRIS-HCl; 140 mM NaCl; 0.01% Tween-20; 1 mM EDTA; 15 mM β-mercaptoethanol; pH = 7.4). Standard agarose gel electrophoresis was performed in a 0.75% gel.



### Dot-blot based assay for quantification of DNA

Genomic DNA isolated from CJ236 *E. coli* strain [*dut*−, *ung*−] (in log phase) served as a uracil standard. A total of 5 ng of this genomic DNA was diluted into 2 μg of ultrapure salmon sperm DNA as an inert carrier (Invitrogen), which was kept constant during the two-third dilution series of this standard. The two-third serial dilutions for XL1-Blue, BL21(DE3) and BL21(DE3) *ung*-151 *E. coli* samples started with 1 μg of DNA mixed into 1 μg of carrier salmon sperm DNA. In case of samples derived from S2 and HCT116 cells 0.6 μg of their genomic DNA was diluted into 1.4 μg of carrier salmon sperm DNA. In every case the 2 μg total DNA content was kept constant with salmon sperm DNA. DNA samples were spotted onto a prewetted positively charged nylon membrane (Amersham Hybond-Ny+; GE Healthcare, Little Chalfont, UK) using a vacuum-driven microfiltration apparatus (Bio-Dot, Bio-Rad, Hercules, CA, USA). After 10 min of air-drying, immobilization of DNA was performed with 2 h of incubation at 80°C. Membrane was blocked by a 15 min incubation in blocking buffer: ETBS-T (25 mM Tris-HCl, pH = 7.4, 2.7 mM KCl, 137 mM NaCl, 1 mM EDTA, 0.05% Tween-20) containing 100 μg/ml salmon sperm DNA, 5% non-fat milk powder and 10 mM β-mercaptoethanol. Membrane was incubated with the 3×-Flag-ΔUNG construct (18.1 μg/ml) in blocking buffer overnight at 4°C. After several washing steps with ETBS-T, anti-Flag M2 antibody (Sigma) was added for 1 h at room temperature (1:2000 dilution in ETBS-T with 5% non-fat milk powder). After washing the membranes, horseradish peroxidase coupled secondary antibody was applied (Sigma). Immunoreactive bands were visualized by enhanced chemiluminescence reagent (GE Healthcare, Buckinghamshire, UK) and 16 bit images were captured by a BioRad ChemiDoc™ MP Imaging system. Densitometry was done using ImageJ 1.48p software (National Institutes of Health, Bethesda, MD, USA). Normalized intensity values were calculated by adjusting the raw intensity values for the background originating from the carrier salmon sperm DNA.

The average molar mass of a nucleotide in *E. coli* ( $M_{E.coli}$ ) or *D. melanogaster* ( $M_{D.melanogaster}$ ) genomic DNA was calculated with the following form:

$$M\left[\frac{\text{g}}{\text{mol}}\right] = \frac{\text{GC}\%}{2} \times (M_{dGMP} + M_{dCMP} - 2M_{H_2O}) + \frac{1-\text{GC}\%}{2} \times (M_{dAMP} + M_{dTMP} - 2M_{H_2O})$$

where  $M_{dGMP} = 347.2$  g/mol,  $M_{dCMP} = 307.2$  g/mol,  $M_{dAMP} = 331.2$  g/mol,  $M_{dTMP} = 322.2$  g/mol,  $M_{H_2O} = 18.0$  g/mol, are the respective molecular weights of the given compounds; dGMP-dCMP percentage (GC%) of the *E. coli* genome is 50.7% and is 42.1% for *D. melanogaster* (based on NCBI genome database, average of reference genomes). Calculated values were 308.95 g/mol and 308.90 g/mol for *E. coli* ( $M_{E.coli}$ ) and *D. melanogaster* ( $M_{D.melanogaster}$ ), respectively. The very slight difference between  $M_{E.coli}$  and  $M_{D.melanogaster}$  was neglected during further analysis.

In each dot of the standard samples, the mass of non-carrier uracil-containing DNA was known and termed as  $m_{\text{dot, stand}}$ . The number of DNA nucleotides present in each dot of the standard samples were calculated using the fol-

lowing form:  $n_{E.coli} = \frac{m_{\text{dot, stand}}}{M_{E.coli}}$ . The number of deoxyuridine nucleotides were calculated using the previously determined uracil content of genomic DNA isolated from log phase culture of CJ236 [*dut*−, *ung*−] *E. coli*, i.e.  $6580 \pm 174$  deoxyuridine/million nucleotide (45). The following equation gives the amount of uracil in each dot of the standard sample ( $n_{U, \text{standard}}$ ):

$$n_{U, \text{standard}} = 6580/1000000 \times n_{E.coli}$$

Calibration curve from the dilution of the standard was visualized the following way for quantification: the amount of uracil in each dot of the standard samples ( $n_{U, \text{standard}}$ ) were plotted against the corresponding normalized intensity values ( $I_{\text{norm, standard}}$ ). Values were fitted with a polynomial with the least order that provided a fit with  $R^2 \geq 0.99$ . The  $\frac{\text{number of uracil}}{\text{million bases}}$  in the 'unknown' genomic DNAs were determined by interpolating their normalized intensities ( $I_{\text{norm, unknown}}$ ) in the calibration plot based on the amount of DNA applied ( $m_{\text{dot, unknown}}$ ).

### Statistics

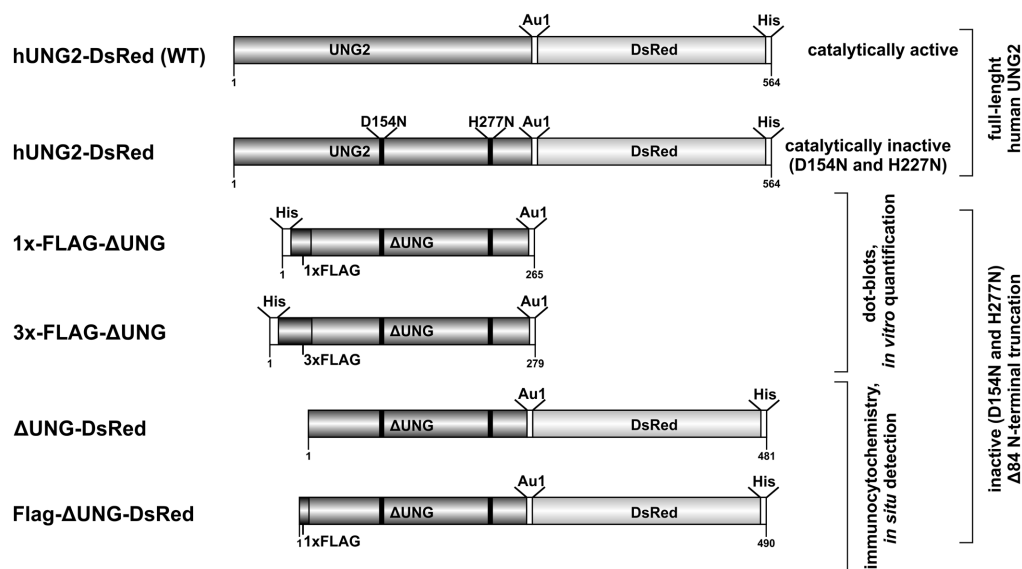
Statistical analysis was carried out by InStat 3.05 software (GraphPad Software, San Diego CA, USA) using the non-parametric Kruskal–Wallis test or one-way ANOVA test with Student-Newman-Keuls multiple comparison post-hoc test when samples passed equal variance (Bartlett's test) and normal distribution tests (Kolmogorov–Smirnov test). Differences were considered statistically significant at  $P < 0.05$ .

### Western blot

Cells were collected, washed with PBS, and resuspended in lysis buffer (50 mM TRIS-HCl pH = 7.4; 140 mM NaCl; 0.4% NP-40; 2 mM dithiothreitol (DTT); 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride; 5 mM benzamidin, 1×Complete ULTRA™ EDTA free protease inhibitor cocktail tablet (Roche)). Cell lysis was assisted with sonication. Insoluble fraction was removed by centrifugation ( $20\,000 \times g \times 15$  min at 4°C). Protein concentration was measured with BioRad Protein Assay to ensure equivalent total protein load per lane. Proteins were resolved under denaturing and reducing conditions on a 12% polyacrylamide gel and transferred to PDVF membrane (Immobilon-P, Merck Millipore, Billerica, MA, USA). Membranes were blocked with 5% non-fat dried milk and were developed against GFP (1:2000, Molecular probes, Life Technologies, Carlsbad, CA, USA) and actin (1:500, Sigma) for loading control. After applying horseradish peroxidase coupled secondary antibodies (Amersham Pharmacia Biotech), immunoreactive bands were visualized by enhanced chemiluminescence reagent (GE Healthcare) and images were captured by a BioRad ChemiDoc™ MP Imaging system.

### Staining uracil residues in *E. coli* genomic DNA

Immunofluorescence staining was done based on the work of (57) with modifications. Briefly, 500 μl of XL1-Blue [*dut*+, *ung*+] and CJ236 [*dut*−, *ung*−] *E. coli* cells were collected in log-phase ( $A_{600\text{nm}} = 0.5$ ), by centrifuging them at



**Figure 1.** Schematics of the used constructs for uracil detection. In our constructs, human UNG2 was used as the uracil sensor core domain. During *in vitro* quantification and for *in situ* detection, a double mutant UNG2 was created (D154N and H277N, mutated sites indicated with black lines within the schematics of the protein domains). This mutant is catalytically inactive but is still capable of binding uracil moieties in DNA (cf text for more details). The N-terminal 84 residues, responsible for the binding to RPA and PCNA, were also removed ( $\Delta$ UNG) to prevent non-specific binding. The  $\Delta$ UNG uracil recognizing core was fused to epitope tags (1 $\times$ /3 $\times$ FLAG, Au1) for immunodetection, DsRed-monomer for direct fluorescent detection and to His-tag for affinity purification.

7000  $\times$  g for 5 min and were washed with PBS. Cells were fixed with Carnoy's fixative (ethanol:acetic acid:chloroform = 6:3:1) for 20 min at 4°C. Rehydration was performed as following: washing with 1:1 ethanol:PBS, 3:7 ethanol:PBS and finally incubating in PBS containing 0.05% Triton X-100 (PBST) for 5 min. Cells were washed once with GTE buffer (50 mM glucose, 20 mM Tris, pH = 7.5 and 10 mM EDTA), and finally resuspended in GTE buffer containing 10 mg/ml lysozyme (Sigma) for 5 min. The suspension was applied onto poly-L-lysine coated cover glasses for an additional 5 minutes, excess fluid was drained and cells were left to air-dry. Cells were washed with PBST for 10 min and were blocked in blocking buffer (5% BSA, in PBST) for 15 min. Uracil residues were visualized by applying 4.64  $\mu$ g/ml of the Flag- $\Delta$ UNG-DsRed construct in blocking buffer, overnight at 4°C. After several washing steps with PBST, anti-FLAG M2 antibody (Sigma) was applied (1:2000 dilution) for 1 h in a blocking buffer. FLAG epitope was visualized by applying Alexa 488 conjugated secondary antibody (1:1000, Molecular Probes). Cells were counterstained with 1  $\mu$ g/ml DAPI (4',6-diamidino-2-phenylindole, Sigma) and embedded in FluorSave<sup>TM</sup> Reagent (Calbiochem, Merck Millipore, Billerica, MA, USA). Images were acquired with a Zeiss LSCM 710 microscope using a 63 $\times$  NA = 1.4 Plan Apo objective.

#### Staining uracil residues of extrachromosomal plasmids in MEF cells

Cells were washed with prewarmed (37°C) phosphate buffered saline (PBS, pH = 7.4) and were fixed with ice cold Carnoy's fixative (ethanol:acetic acid:chloroform = 6:3:1) for 20 minutes at 4°C. Rehydration was performed as following: washing step with 1:1 ethanol:PBS, 3:7

ethanol:PBS, finally with PBS for 5 min. Epitope unmasking was done by applying 1 N HCl, 0.5% Triton X-100 for 15 min, after which 0.1 M Na<sub>2</sub>B<sub>4</sub>O<sub>7</sub> (pH = 8.5) was used for neutralization for 5 min followed by PBS washing. HCl was used to denature DNA, allowing our UNG construct to have better access to genomic uracil. This method is also routinely applied when using antibodies against BrdU, doing cell proliferation assays (see (58)). Blocking was done at room temperature for 1 h in blocking buffer: 200  $\mu$ g/ml salmon sperm DNA, 5% fetal goat serum (FGS), 3% fetal bovine serum albumin (BSA) and 0.05% Triton X-100 in PBS. Uracil residues were visualized by applying 4.64  $\mu$ g/ml of the Flag- $\Delta$ UNG-DsRed construct in blocking buffer, overnight at 4°C. After several washing steps with blocking buffer, anti-FLAG M2 antibody (Sigma) was applied (1:2000 dilution) for 1 h in a blocking buffer not containing salmon sperm DNA. FLAG epitope was visualized by applying Alexa 488 conjugated secondary antibody (1:1000, Molecular Probes). Cells were counterstained with 1  $\mu$ g/ml DAPI (Sigma) and embedded in FluorSave<sup>TM</sup> Reagent (Calbiochem). Images were acquired with a Zeiss LSCM 710 microscope using a 63 $\times$  NA = 1.4 Plan Apo objective.

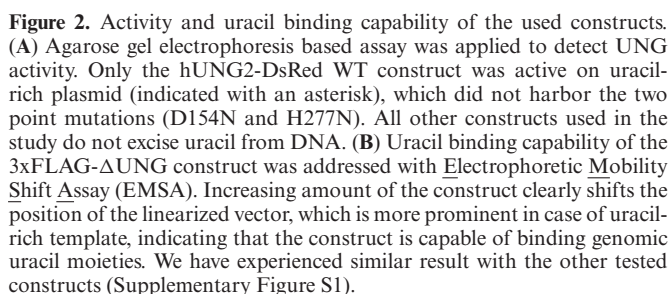
## RESULTS AND DISCUSSION

### Construction and analysis of catalytically inactive uracil-sensor proteins

Wild-type human UNG2 possesses a highly selective substrate binding site for uracil and is specific for excising uracil from DNA, with a negligible activity toward the natural DNA bases cytosine or thymine (59). The catalytically inactive double mutant (D145N, H268N) human UNG2 preserves this highly specific and strong binding interac-

In order to check the required and expected functionality of our constructs, we have carried out enzyme activity assays (as described in ‘Materials and Methods’ section). Figure 2A confirms that the only construct showing catalytic activity on uracil-rich DNA is the wild type UNG, whereas all of our  $\Delta$ UNG constructs lack any detectable excising activity. On the other hand, as shown on Figure 2B and Supplementary Figure S1, all of our constructs show preferential binding to uracil-rich as compared to normal DNA.

In testing our different constructs, we have used either just a single (1×) or a triplicated (3×) Flag-tag form and employed them on our standard, namely the genomic DNA isolated from log phase CJ236 [*dut*−, *ung*−] *E. coli* cells for which uracil content has been previously quantified (43,45). Figure 3 shows that the triplicated Flag-tag containing ΔUNG sensor is more sensitive in the dot-blot assay. This construct was therefore used in all further experiments. Figure 3 also indicates that a linear response could be reached on a series of dilutions in wide dynamic range. For each experiment, a calibration curve was always recorded



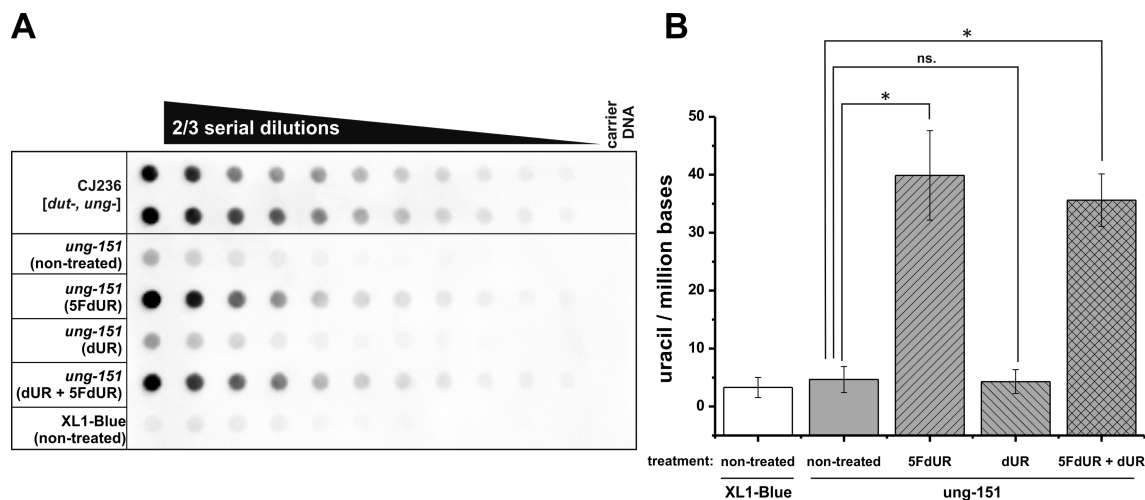
Amount of uracil / dot ( $10^{-15}$ mol)	1x-FLAG (Norm, stand)	3x-FLAG (Norm, stand)
0	0	0
2.5	~5	~8
5	~10	~15
7.5	~15	~22
12.5	~25	~35
21.5	~40	~50
32.5	~55	~65



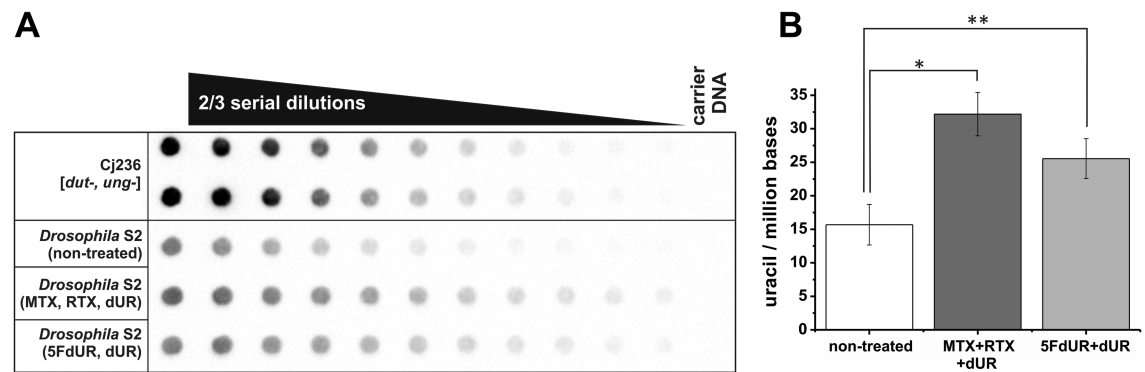
**Figure 3.** Design of a standard curve for *in vitro* quantification of genomic uracil levels. Genomic DNA isolated from log phase growing CJ236 [*dut*–, *ung*–] *Escherichia coli* strain was used as a standard with well-defined uracil-content during quantification. Applying a serial dilution of this standard provides a wide and reproducible range for uracil quantification. The normalized calibration curve is from four independent datasets ( $n = 4$ ), where error bars show standard errors of the mean (SEM). The inset shows that the 3×FLAG-ΔUNG construct is slightly more sensitive under similar conditions, compared to the 1×FLAG-ΔUNG construct (as based on a 4 point two-third serial dilution, starting with 100 ng of standard genomic DNA).







**Figure 5.** Dot-blot assay for measuring genomic uracil levels of 5FdUR, dUR treated *Escherichia coli* cells. (A) CJ236 [*dut*<sup>-</sup>, *ung*<sup>-</sup>] *E. coli* genomic DNA was used as standard for the dot-blot assay. Quantity of genomic uracil of different drug-treated (5FdUR and dUR or both) and non-treated *E. coli* BL21(DE3) *ung*-151 samples were measured along with XL1-Blue [*dut*<sup>+</sup>, *ung*<sup>+</sup>], applied as a negative control. (B) Bar graph shows the uracil moieties/million bases of each sample (mean values  $\pm$  the standard errors of the mean). Significant increase (\*) in uracil-DNA content was only observed using 5FdUR treatment, or the combined 5FdUR and dUR treatment as compared to non-treated cells ( $P < 0.05$ ). Calculations were based on six independent datasets ( $n = 6$ ).



**Figure 6.** Dot-blot assay for measuring genomic uracil levels of *Drosophila* Schneider S2 cells after treatment with *de novo* thymidylate biosynthesis pathway inhibitors. (A) CJ236 [*dut*<sup>-</sup>, *ung*<sup>-</sup>] *Escherichia coli* genomic DNA was used as standard for the dot-blot assay. Genomic uracil content of different drug-treated (5FdUR, dUR or MTX, RTX, dUR) and non-treated *Drosophila* S2 cells were measured. (B) Bar graph shows the uracil moieties/million bases of each sample (mean values  $\pm$  the standard errors of the mean). Both types of treatments led to significantly elevated genomic uracil levels as compared to non-treated cells (\* =  $P < 0.01$ , \*\* =  $P < 0.05$ ). Calculations were based on six independent datasets ( $n = 6$ ).

bases for the 5FdUR treated. Similarly no significant difference was measured in case of cells that were transfected with an empty vector, expressing only GFP as control ( $4.17 \pm 1.19$  for the non-treated and  $7.39 \pm 2.98$  uracil/million bases for the 5FdUR treated cells, respectively). These results are in good agreement with previously reported data (cf (68)) and also highlight the importance of simultaneous UNG inhibition along with drug treatments targeting the *de novo* thymidylate biosynthesis pathway for effective cancer therapy.

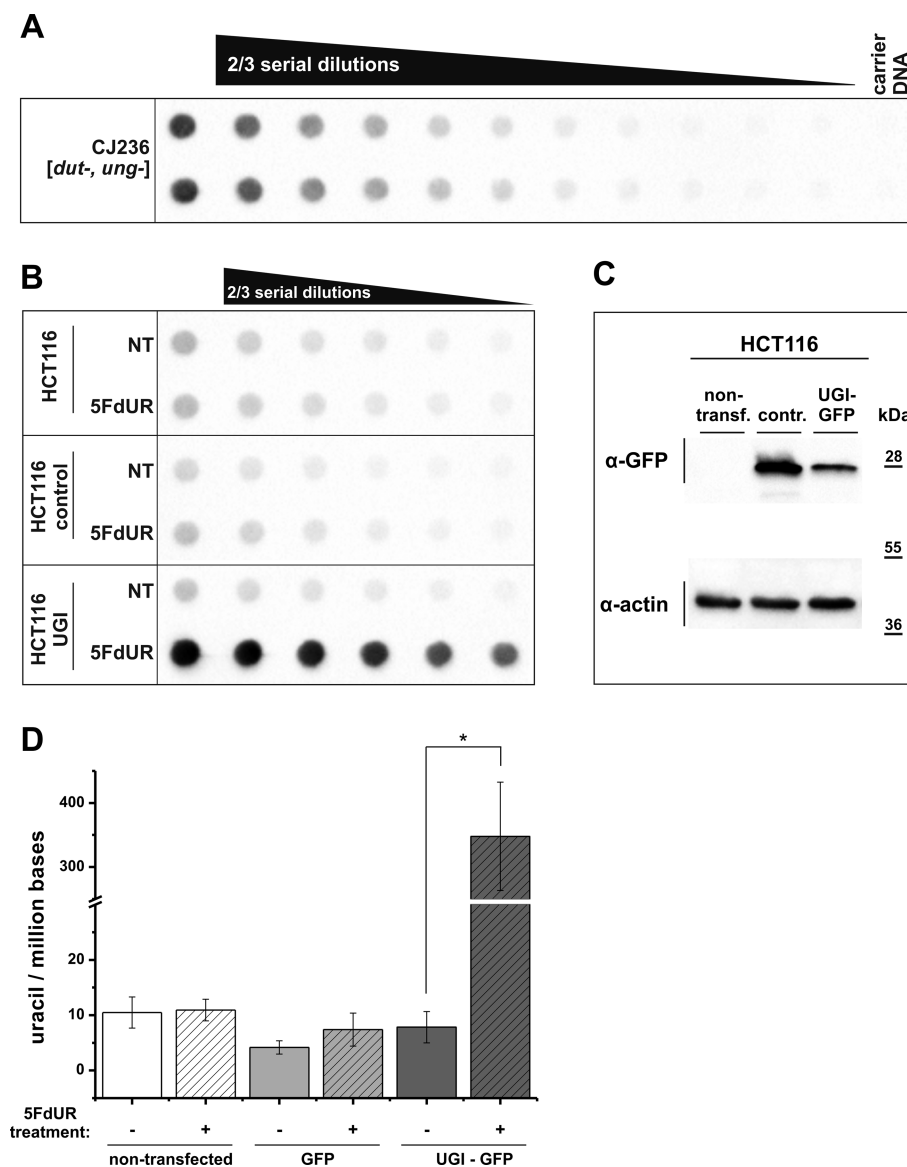
#### Application of the catalytically inactive UNG constructs for detection of uracil in DNA *in situ*

Figure 8 presents that the highly uracil-rich character of the CJ236 [*dut*<sup>-</sup>, *ung*<sup>-</sup>] *E. coli* cells, containing 6580 uracil/million bases, can be readily visualized via immunocytochemistry using our presently developed uracil sensor

construct. Relying on the Flag-tag in our  $\Delta$ UNG constructs, the uracil-DNA staining is easy to detect and colocalizes, as expected, with the DAPI signal for DNA. It is also important to note that the DsRed-tag may also be used for direct visualization. Specificity of the signal is adequately corroborated by lack of the staining in the XL1-Blue [*dut*<sup>+</sup>, *ung*<sup>+</sup>] cells. Application of the UNG inhibitor UGI (obtained from NEB) protein erases the signal, showing again the specific character of our assay (Supplementary Figure S3). The well-described UNG-UGI interaction has been documented to prevent binding of UNG to DNA (74).

We have also attempted to use similar strategy for staining of uracil-DNA in a mammalian cellular background. To this end, we have transfected [*ung*<sup>-</sup>/*-*] MEF cells with plasmid DNA produced by CJ236 [*dut*<sup>-</sup>, *ung*<sup>-</sup>] *E. coli* cells. Such plasmid DNA contains  $\sim 6580$  uracil/million bases (45) and its tolerance in the MEF cells is ensured by lack of



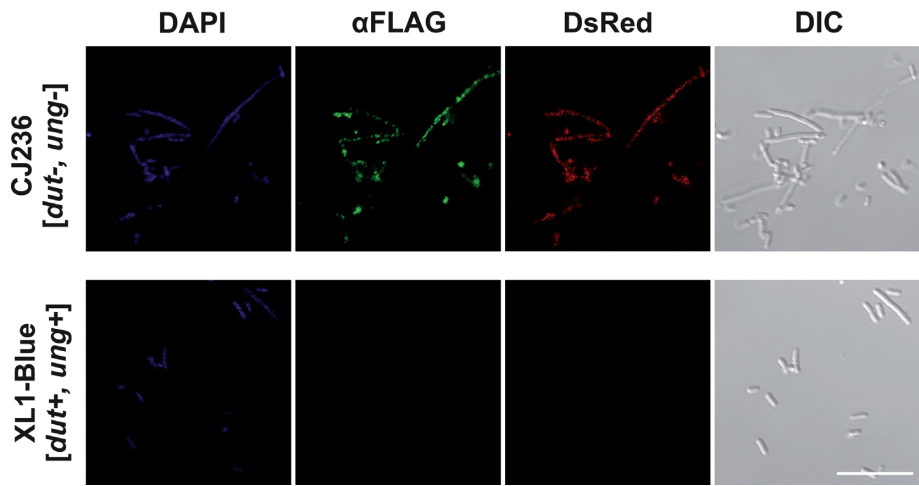


**Figure 7.** Dot-blot assay for measuring genomic uracil levels of HCT116 cells after treatment with *de novo* thymidylate biosynthesis pathway inhibitors and UNG inhibition. (A) CJ236 [*dut-*, *ung-*] *Escherichia coli* genomic DNA was used as standard for the dot-blot assay. (B) Genomic uracil levels of 5FdUR treated and non-treated HCT116 cells were measured in the context of endogenous UNG inhibition with UGI expression. (C) Bar graph shows the uracil moieties/million bases of each sample (mean values  $\pm$  the standard errors of the mean). 5FdUR treatment led to significantly elevated uracil levels only in cells also expressing UGI ( $* = P < 0.05$ ) when compared to non-treated cells. Calculations were based on four independent datasets ( $n = 4$ ). (D) Western blot showing GFP expression of cells transfected by the UGI-GFP vector and the empty vector used as a control (only expressing GFP). The membrane were also developed against actin as a loading control.

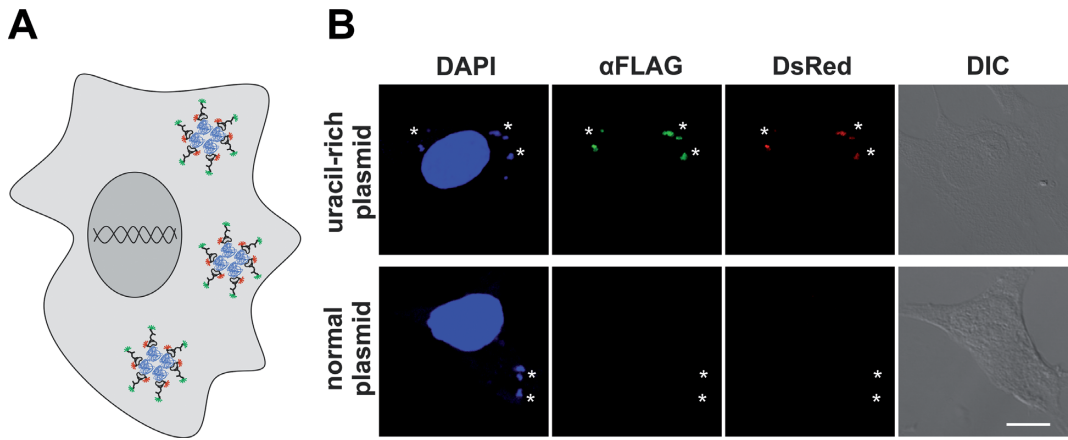
the UNG enzyme in this mouse cell line. In this experiment, we make use of a well-known artifactual effect of plasmid transfection, namely that upon using high amount of plasmid DNA during transfection, plasmid aggregates can occur within the cells (75) (Figure 9A). We therefore worked out our transfection experimental conditions such that to allow the potential accumulation of intracellular plasmid aggregates (cf 'Materials and Methods' section). Figure 9B leftmost panels clearly show DAPI staining of these plasmid aggregates indicated by white asterisks. In case of non-transfected cells (negative control), no plasmid aggregates can be observed (Supplementary Figure S4). On Figure 8, upper panels with cells transfected with uracil-rich plasmid

show positive reaction with our uracil sensor molecules, either via Flag-tag or via direct DsRed detection. Additionally as a negative control, cells transfected by normal plasmid (produced in XL1-Blue cells) show no staining. Further controls are also shown in Supplementary Figure S4 where the lack of uracil-specific staining can be observed when the UNG-inhibitor UGI was applied.

Although these results clearly show the possibility of using our method for *in situ* microscopic detection, the sensitivity of this staining approach needs to be developed further to allow detection of lower uracil levels. Additional enhancement of the signal might be expected from the application of brighter and more stable fluorophores, such as



**Figure 8.** *In situ* genomic uracil detection in *Escherichiacoli* using an immunocytochemistry approach. The genomic uracil content of CJ236 [*dut*–, *ung*–] *E. coli* cells was visualized with the Flag-ΔUNG-DsRed construct. As a negative control, XL1-Blue cells [*dut*+, *ung*+] were also used in the same staining procedure. Only the CJ236 [*dut*–, *ung*–] *E. coli* sample showed staining, either detected directly through the signal of DsRed (red) or through the FLAG epitope tag (green). DAPI was used to counterstain DNA. Scale bar represents 10 μm.



**Figure 9.** Detection of uracil-rich and normal plasmid DNA aggregates in MEF [*ung*–/–] cells. (A) Schematic image of the cytoplasmic plasmid aggregates visualized by the Flag-ΔUNG-DsRed construct. (B) Asterisks (\*) show plasmid aggregates. Only cells transfected with uracil-rich plasmids could be visualized both through the DsRed (red) tag and the FLAG epitope tag (green). The DAPI staining is oversaturated to show the faint DAPI positive plasmid aggregates in the cytoplasm. Scale bar represents 10 μm.

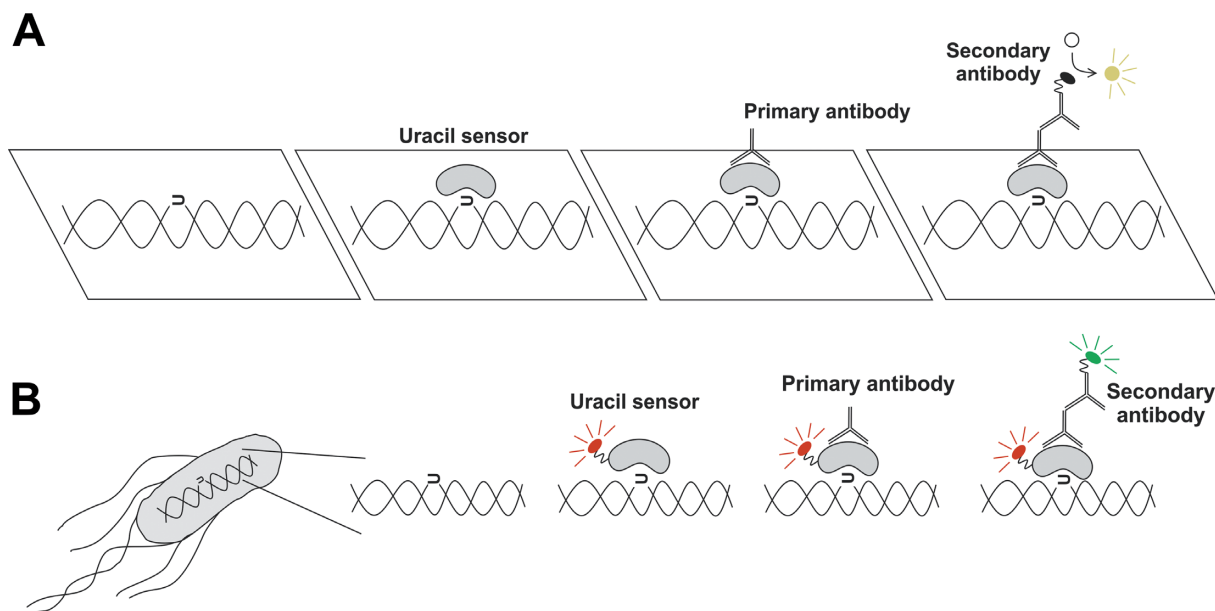
nanocrystals or quantum dots (cf (76)). Our present method forms the basis for such further developments.

CONCLUSION

The workflow of the novel method described in this study is shown schematically on Figure 10, both for the *in vitro* dot-blot based detection (Figure 10A) and the *in situ* immunocytochemical approach used for cellular detection (Figure 10B). Uracil moieties are recognized by a catalytically inactive UNG sensor protein and the readout signal is enhanced by using primary and secondary sets of antibodies. The recently described novel mycobacterial uracil-DNA binding protein may also serve as an alternative sensor framework (77). The method is straightforward, easy to use and can be applied in a high-throughput manner to analyze DNA from any organism. It does not require expensive instruments or complex know-how, facilitating its easy implementation in any basic molecular biology laboratory. Elevated genomic

deoxyuridine content of cells from diverse genetic background and/or treated with different drugs can be demonstrated *in situ*, within the cell. Direct detection is possible through the DsRed-tagged construct, or antibodies may be used for signal enhancement through the different epitope tags.

Direct comparisons between values for genomic uracil content obtained by different methods are far from straightforward (cf e.g. data for genomic DNA from CJ236 [*dut*–, *ung*–] *E. coli* cells range between 3000–18 000/million bases (38,43,45)). Hence, the truly reliable approach is to compare relative differences induced by different cellular stages, environment and/or drug treatments using the same method. Our present method is optimal for detecting such differences due to its ease of application, robustness and amenity to high-throughput studies. With the dot-blot based assay, comparative data between different organisms and different cellular conditions are obtained fast and in a quantita-



**Figure 10.** Schematics summarizing the developed *in vitro* quantification and *in situ* detection method. (A) Schematic image of the applied *in vitro* quantification through a dot-blot approach. (B) Schematic image of the applied immunocytochemical approach for *in situ* uracil detection.

tive manner. Finally, a major further significance and novelty in our presently proposed study is that it provides potential for *in situ* detection of uracil-DNA within cells. To our knowledge at least, such *in situ* detection has not yet been described for uracil in DNA. Considering the numerous physiological events and pathological conditions where uracil levels in DNA are modified, our method presents a biologically relevant tool for assessing the composition of genomic DNA and its putative alterations during different cellular conditions. In this respect, we wish to point out that our results obtained with UGI-expressing HCT116 cells simultaneously treated with 5FdUR indicates that the combined perturbation of base-excision repair and *de novo* thymidylate biosynthesis leads to a synergistic cellular response elevating genomic uracil levels. The significance of our proposed technique is further enhanced by the fact that uracil detection has yet escaped the highly powerful single-molecule real-time sequencing (SMRT) technology, as well.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENT

Human uracil-DNA glycosylase 2 (UNG2) cDNA was a generous gift of Professor Salvatore Caradonna (Department of Molecular Biology, University of Medicine and Dentistry of New Jersey). The vector expressing the human codon optimized UGI along with GFP (pLGC-hUgi) was a kind gift of Michael D. Wyatt (South Carolina College of Pharmacy, University of South Carolina).

## FUNDING

Hungarian Scientific Research Fund OTKA [NK 84008, K109486]; Baross Program of the New Hungary De-

velopment Plan [3DSTRUCT, OMFB-00266/2010 REG-KM-09-1-2009-0050]; Hungarian Academy of Sciences (MedinProt program) [TTK IF-28/2012]; ICGEB Research Grant [CRP/HUN14-01]; New Hungary Development Plan KMR\_12-1-2012-0140. Funding for open access charge: Hungarian Academy of Sciences; Hungarian Scientific Research Fund OTKA.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Visnes, T., Doseth, B., Pettersen, H.S., Hagen, L., Sousa, M.M., Akbari, M., Otterlei, M., Kavli, B., Slupphaug, G. and Krokan, H.E. (2009) Uracil in DNA and its processing by different DNA glycosylases. *Philos. Trans. R Soc. Lond. B Biol. Sci.*, **364**, 563–568.
2. Krokan, H.E., Drablos, F. and Slupphaug, G. (2002) Uracil in DNA—occurrence, consequences and repair. *Oncogene*, **21**, 8935–8948.
3. Wardle, J., Burgers, P.M., Cann, I.K., Darley, K., Heslop, P., Johansson, E., Lin, L.J., McGlynn, P., Sanvoisin, J., Stith, C.M. *et al.* (2008) Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya. *Nucleic Acids Res.*, **36**, 705–711.
4. Vertessy, B.G. and Toth, J. (2009) Keeping uracil out of DNA: physiological role, structure and catalytic mechanism of dUTPases. *Acc. Chem. Res.*, **42**, 97–106.
5. Takacs, E., Grolmusz, V.K. and Vertessy, B.G. (2004) A tradeoff between protein stability and conformational mobility in homotrimeric dUTPases. *FEBS Lett.*, **566**, 48–54.
6. Vertessy, B.G., Zalud, P., Nyman, P.O. and Zeppezauer, M. (1994) Identification of tyrosine as a functional residue in the active site of Escherichia coli dUTPase. *Biochim. Biophys. Acta*, **1205**, 146–150.
7. Nagy, G.N., Leveles, I. and Vertessy, B.G. (2014) Preventive DNA repair by sanitizing the cellular (deoxy)nucleoside triphosphate pool. *FEBS J.*, **281**, 4207–4223.
8. Wallace, S.S. (2014) Base excision repair: a critical player in many games. *DNA Repair (Amst)*, **19**, 14–26.
9. Krokan, H.E. and Bjoras, M. (2013) Base excision repair. *Cold Spring Harb. Perspect. Biol.*, **5**, a012583.
10. Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P.A., Hagen, L., Krokan, H.E. and Slupphaug, G. (2002) 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.*, **277**, 39926–39936.
11. Panayotou, G., Brown, T., Barlow, T., Pearl, L.H. and Savva, R. (1998) Direct measurement of the substrate preference of uracil-DNA glycosylase. *J. Biol. Chem.*, **273**, 45–50.
  12. Slupphaug, G., Eftedal, I., Kavli, B., Bharati, S., Helle, N.M., Haug, T., Levine, D.W. and Krokan, H.E. (1995) Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase. *Biochemistry*, **34**, 128–138.
  13. Lindahl, T., Ljungquist, S., Siebert, W., Nyberg, B. and Sperens, B. (1977) DNA N-glycosidases: properties of uracil-DNA glycosidase from *Escherichia coli*. *J. Biol. Chem.*, **252**, 3286–3294.
  14. Bellamy, S.R. and Baldwin, G.S. (2001) A kinetic analysis of substrate recognition by uracil-DNA glycosylase from herpes simplex virus type 1. *Nucleic Acids Res.*, **29**, 3857–3863.
  15. Eftedal, I., Guddal, P.H., Slupphaug, G., Volden, G. and Krokan, H.E. (1993) Consensus sequences for good and poor removal of uracil from double stranded DNA by uracil-DNA glycosylase. *Nucleic Acids Res.*, **21**, 2095–2101.
  16. Nilsen, H., Yazdankhah, S.P., Eftedal, I. and Krokan, H.E. (1995) Sequence specificity for removal of uracil from U:A pairs and U:G mismatches by uracil-DNA glycosylase from *Escherichia coli*, and correlation with mutational hotspots. *FEBS Lett.*, **362**, 205–209.
  17. Dizdaroğlu, M., Karakaya, A., Jaruga, P., Slupphaug, G. and Krokan, H.E. (1996) Novel activities of human uracil DNA N-glycosylase for cytosine-derived products of oxidative DNA damage. *Nucleic Acids Res.*, **24**, 418–422.
  18. Hatahet, Z., Kow, Y.W., Purnmal, A.A., Cunningham, R.P. and Wallace, S.S. (1994) New substrates for old enzymes. 5-Hydroxy-2'-deoxycytidine and 5-hydroxy-2'-deoxyuridine are substrates for *Escherichia coli* endonuclease III and formamidopyrimidine DNA N-glycosylase, while 5-hydroxy-2'-deoxyuridine is a substrate for uracil DNA N-glycosylase. *J. Biol. Chem.*, **269**, 18814–18820.
  19. Liu, P., Burdzy, A. and Sowers, L.C. (2002) Substrate recognition by a family of uracil-DNA glycosylases: UNG, MUG, and TDG. *Chem. Res. Toxicol.*, **15**, 1001–1009.
  20. Warner, H.R. and Rockstroh, P.A. (1980) Incorporation and excision of 5-fluorouracil from deoxyribonucleic acid in *Escherichia coli*. *J. Bacteriol.*, **141**, 680–686.
  21. Bharti, S.K. and Varshney, U. (2010) Analysis of the impact of a uracil DNA glycosylase attenuated in AP-DNA binding in maintenance of the genomic integrity in *Escherichia coli*. *Nucleic Acids Res.*, **38**, 2291–2301.
  22. Krusong, K., Carpenter, E.P., Bellamy, S.R., Savva, R. and Baldwin, G.S. (2006) A comparative study of uracil-DNA glycosylases from human and herpes simplex virus type 1. *J. Biol. Chem.*, **281**, 4983–4992.
  23. Pettersen, H.S., Sundheim, O., Gilljam, K.M., Slupphaug, G., Krokan, H.E. and Kavli, B. (2007) Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. *Nucleic Acids Res.*, **35**, 3879–3892.
  24. Mol, C.D., Arvai, A.S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H.E. and Tainer, J.A. (1995) Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell*, **80**, 869–878.
  25. Savva, R., McAuley-Hecht, K., Brown, T. and Pearl, L. (1995) The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature*, **373**, 487–493.
  26. Wilson, P.M., Danenberg, P.V., Johnston, P.G., Lenz, H.J. and Ladner, R.D. (2014) Standing the test of time: targeting thymidylate biosynthesis in cancer therapy. *Nat. Rev. Clin. Oncol.*, **11**, 282–298.
  27. Skurnik, M., Hyytiäinen, H., Happonen, L.J., Kiljunen, S., Datta, N., Mattinen, L., Williamson, K., Kristo, P., Szeliga, M., Kalin-Manttari, L. et al. (2012) Characterization of the genome, proteome and structure of Yersiniophage  $\Phi$ R1-37. *J. Virol.*, **86**, 12625–12642.
  28. Kiljunen, S., Hakala, K., Pinta, E., Huttunen, S., Pluta, P., Gador, A., Lonnberg, H. and Skurnik, M. (2005) Yersiniophage  $\phi$ R1-37 is a tailed bacteriophage having a 270 kb DNA genome with thymidine replaced by deoxyuridine. *Microbiology*, **151**, 4093–4102.
  29. Langridge, R. and Marmur, J. (1964) X-ray diffraction study of a DNA which contains uracil. *Science*, **143**, 1450–1451.
  30. Lozeron, H.A. and Szybalski, W. (1967) Incorporation of 5-fluorodeoxyuridine into the DNA of *Bacillus subtilis* phage PBS2 and its radiobiological consequences. *J. Mol. Biol.*, **30**, 277–290.
  31. Pettersen, H.S., Galashevskaya, A., Doseth, B., Sousa, M.M., Sarno, A., Visnes, T., Aas, P.A., Liabakk, N.B., Slupphaug, G., Saetrom, P. et al. (2015) AID expression in B-cell lymphomas causes accumulation of genomic uracil and a distinct AID mutational signature. *DNA Repair (Amst)*, **25**, 60–71.
  32. Maul, R.W. and Gearhart, P.J. (2010) AID and somatic hypermutation. *Adv. Immunol.*, **105**, 159–191.
  33. Liu, M. and Schatz, D.G. (2009) Balancing AID and DNA repair during somatic hypermutation. *Trends Immunol.*, **30**, 173–181.
  34. Yan, N., O'Day, E., Wheeler, L.A., Engelman, A. and Lieberman, J. (2011) HIV DNA is heavily uracilated, which protects it from autointegration. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 9244–9249.
  35. Horvath, A., Bekesi, A., Muha, V., Erdelyi, M. and Vertessy, B.G. (2013) Expanding the DNA alphabet in the fruit fly: uracil enrichment in genomic DNA. *Fly (Austin)*, **7**, 23–27.
  36. Muha, V., Horvath, A., Bekesi, A., Pukancsik, M., Hodoscsek, B., Merenyi, G., Rona, G., Batki, J., Jankovics, F. et al. (2012) Uracil-containing DNA in *Drosophila*: stability, stage-specific accumulation, and developmental involvement. *PLoS Genet.*, **8**, e1002738.
  37. Galashevskaya, A., Sarno, A., Vagbo, C.B., Aas, P.A., Hagen, L., Slupphaug, G. and Krokan, H.E. (2013) A robust, sensitive assay for genomic uracil determination by LC/MS/MS reveals lower levels than previously reported. *DNA Repair (Amst)*, **12**, 699–706.
  38. Atamna, H., Cheung, I. and Ames, B.N. (2000) A method for detecting abasic sites in living cells: age-dependent changes in base excision repair. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 686–691.
  39. Blount, B.C. and Ames, B.N. (1994) Analysis of uracil in DNA by gas chromatography-mass spectrometry. *Analyt. Biochem.*, **219**, 195–200.
  40. Blount, B.C., Mack, M.M., Wehr, C.M., MacGregor, J.T., Hiatt, R.A., Wang, G., Wickramasinghe, S.N., Everson, R.B. and Ames, B.N. (1997) Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 3290–3295.
  41. Chango, A., Abdel Nour, A.M., Niquet, C. and Tessier, F.J. (2009) Simultaneous determination of genomic DNA methylation and uracil misincorporation. *Med. Princ. Pract.*, **18**, 81–84.
  42. Mashiyama, S.T., Courtemanche, C., Elson-Schwab, I., Crott, J., Lee, B.L., Ong, C.N., Fenech, M. and Ames, B.N. (2004) Uracil in DNA, determined by an improved assay, is increased when deoxynucleosides are added to folate-deficient cultured human lymphocytes. *Analyt. Biochem.*, **330**, 58–69.
  43. Lari, S.U., Chen, C.Y., Vertessy, B.G., Morre, J. and Bennett, S.E. (2006) Quantitative determination of uracil residues in *Escherichia coli* DNA: Contribution of ung, dug, and dut genes to uracil avoidance. *DNA Repair (Amst)*, **5**, 1407–1420.
  44. Shalhout, S., Haddad, D., Sosin, A., Holland, T.C., Al-Katib, A., Martin, A. and Bhagwat, A.S. (2014) Genomic uracil homeostasis during normal B cell maturation and loss of this balance during B cell cancer development. *Mol. Cell. Biol.*, **34**, 4019–4032.
  45. Horvath, A. and Vertessy, B.G. (2010) A one-step method for quantitative determination of uracil in DNA by real-time PCR. *Nucleic Acids Res.*, **38**, e196.
  46. Wu, D., Chen, L., Sun, Q., Wu, X., Jia, S. and Meng, A. (2014) Uracil-DNA glycosylase is involved in DNA demethylation and required for embryonic development in the zebrafish embryo. *J. Biol. Chem.*, **289**, 15463–15473.
  47. Ladopoulos, V., Hofemeister, H., Hoogenkamp, M., Riggs, A.D., Stewart, A.F. and Bonifer, C. (2013) The histone methyltransferase KMT2B is required for RNA polymerase II association and protection from DNA methylation at the MagohB CpG island promoter. *Mol. Cell. Biol.*, **33**, 1383–1393.
  48. Cliffe, L.J., Hirsch, G., Wang, J., Ekanayake, D., Bullard, W., Hu, M., Wang, Y. and Sabatini, R. (2012) JBP1 and JBP2 proteins are Fe2+/2-oxoglutarate-dependent dioxygenases regulating hydroxylation of thymidine residues in trypanosome DNA. *J. Biol. Chem.*, **287**, 19886–19895.
  49. Moriel-Carretero, M. and Aguilera, A. (2010) A postincision-deficient TFIIH causes replication fork breakage and uncovers alternative



- Rad51- or Pol32-mediated restart mechanisms. *Mol. Cell*, **37**, 690–701.
50. Soultanakis, R.P., Melamed, R.J., Bepalov, I.A., Wallace, S.S., Beckman, K.B., Ames, B.N., Taatjes, D.J. and Janssen-Heininger, Y.M. (2000) Fluorescence detection of 8-oxoguanine in nuclear and mitochondrial DNA of cultured cells using a recombinant Fab and confocal scanning laser microscopy. *Free Radic. Biol. Med.*, **28**, 987–998.
  51. Terasaki, Y., Akuta, T., Terasaki, M., Sawa, T., Mori, T., Okamoto, T., Ozaki, M., Takeya, M. and Akaike, T. (2006) Guanine nitration in idiopathic pulmonary fibrosis and its implication for carcinogenesis. *Am. J. Respir. Crit. Care Med.*, **174**, 665–673.
  52. Bekesi, A., Zagya, I., Hunyadi-Gulyas, E., Pongracz, V., Kovari, J., Nagy, A.O., Erdei, A., Medzihradsky, K.F. and Vertessy, B.G. (2004) Developmental regulation of dUTPase in *Drosophila melanogaster*. *J. Biol. Chem.*, **279**, 22362–22370.
  53. Rona, G., Borsos, M., Ellis, J.J., Mehdi, A.M., Christie, M., Kornyei, Z., Neubrand, M., Toth, J., Bozoky, Z., Buday, L. *et al.* (2014) Dynamics of re-constitution of the human nuclear proteome after cell division is regulated by NLS-adjacent phosphorylation. *Cell cycle*, **13**, 3551–3564.
  54. Kaiser, S.M. and Emmerman, M. (2006) Uracil DNA glycosylase is dispensable for human immunodeficiency virus type 1 replication and does not contribute to the antiviral effects of the cytidine deaminase Apobec3G. *J. Virol.*, **80**, 875–882.
  55. Andersen, S., Ericsson, M., Dai, H.Y., Pena-Diaz, J., Slupphaug, G., Nilsen, H., Aarset, H. and Krokan, H.E. (2005) Monoclonal B-cell hyperplasia and leukocyte imbalance precede development of B-cell malignancies in uracil-DNA glycosylase deficient mice. *DNA Repair (Amst)*, **4**, 1432–1441.
  56. Bennett, S.E., Chen, C.Y. and Mosbaugh, D.W. (2004) Escherichia coli nucleoside diphosphate kinase does not act as a uracil-processing DNA repair nuclease. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 6391–6396.
  57. Sun, Q., Yu, X.C. and Margolin, W. (1998) Assembly of the FtsZ ring at the central division site in the absence of the chromosome. *Mol. Microbiol.*, **29**, 491–503.
  58. Dolbeare, F. (1995) Bromodeoxyuridine: a diagnostic tool in biology and medicine, Part I: Historical perspectives, histochemical methods and cell kinetics. *Histochem. J.*, **27**, 339–369.
  59. Pearl, L.H. (2000) Structure and function in the uracil-DNA glycosylase superfamily. *Mutat. Res.*, **460**, 165–181.
  60. Drohat, A.C., Jagadeesh, J., Ferguson, E. and Stivers, J.T. (1999) Role of electrophilic and general base catalysis in the mechanism of Escherichia coli uracil DNA glycosylase. *Biochemistry*, **38**, 11866–11875.
  61. Otterlei, M., Warbrick, E., Nagelhus, T.A., Haug, T., Slupphaug, G., Akbari, M., Aas, P.A., Steinsbekk, K., Bakke, O. and Krokan, H.E. (1999) Post-replicative base excision repair in replication foci. *EMBO J.*, **18**, 3834–3844.
  62. 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.* (1997) 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.*, **272**, 6561–6566.
  63. Torseth, K., Doseth, B., Hagen, L., Olaisen, C., Liabakk, N.B., Graesmann, H., Durandy, A., Otterlei, M., Krokan, H.E., Kavli, B. *et al.* (2012) The UNG2 Arg88Cys variant abrogates RPA-mediated recruitment of UNG2 to single-stranded DNA. *DNA Repair (Amst)*, **11**, 559–569.
  64. Scaramozzino, N., Sanz, G., Crance, J.M., Saparbaev, M., Drillien, R., Laval, J., Kavli, B. and Garin, D. (2003) Characterisation of the substrate specificity of homogeneous vaccinia virus uracil-DNA glycosylase. *Nucleic Acids Res.*, **31**, 4950–4957.
  65. Santi, D.V., McHenry, C.S., Raines, R.T. and Ivanetich, K.M. (1987) Kinetics and thermodynamics of the interaction of 5-fluoro-2'-deoxyuridylate with thymidylate synthase. *Biochemistry*, **26**, 8606–8613.
  66. Wyatt, M.D. and Wilson, D.M. 3rd. (2009) Participation of DNA repair in the response to 5-fluorouracil. *Cell. Mol. Life Sci.*, **66**, 788–799.
  67. Andersen, S., Heine, T., Sneve, R., Konig, I., Krokan, H.E., Epe, B. and Nilsen, H. (2005) 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*, **26**, 547–555.
  68. Luo, Y., Walla, M. and Wyatt, M.D. (2008) Uracil incorporation into genomic DNA does not predict toxicity caused by chemotherapeutic inhibition of thymidylate synthase. *DNA Repair (Amst)*, **7**, 162–169.
  69. dos Santos, G., Schroeder, A.J., Goodman, J.L., Strelets, V.B., Crosby, M.A., Thurmond, J., Emmert, D.B., Gelbart, W.M. and FlyBase Consortium. (2015) FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res.*, **43**, D690–D697.
  70. Blackledge, G. (1998) New developments in cancer treatment with the novel thymidylate synthase inhibitor raltitrexed ('Tomudex'). *Br. J. Cancer*, **77**(Suppl. 2), 29–37.
  71. Rajagopalan, P.T., Zhang, Z., McCourt, L., Dwyer, M., Benkovic, S.J. and Hammes, G.G. (2002) Interaction of dihydrofolate reductase with methotrexate: ensemble and single-molecule kinetics. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 13481–13486.
  72. Rancourt, S.L. and Walker, V.K. (1990) Kinetic characterization of dihydrofolate reductase from *Drosophila melanogaster*. *Biochem. Cell Biol.*, **68**, 1075–1082.
  73. Bennett, S.E., Schimerlik, M.I. and Mosbaugh, D.W. (1993) Kinetics of the uracil-DNA glycosylase/inhibitor protein association. Ung interaction with Ugi, nucleic acids, and uracil compounds. *J. Biol. Chem.*, **268**, 26879–26885.
  74. Lundquist, A.J., Beger, R.D., Bennett, S.E., Bolton, P.H. and Mosbaugh, D.W. (1997) Site-directed mutagenesis and characterization of uracil-DNA glycosylase inhibitor protein. Role of specific carboxylic amino acids in complex formation with Escherichia coli uracil-DNA glycosylase. *J. Biol. Chem.*, **272**, 21408–21419.
  75. Zabner, J., Fasbender, A.J., Moninger, T., Poellinger, K.A. and Welsh, M.J. (1995) Cellular and molecular barriers to gene transfer by a cationic lipid. *J. Biol. Chem.*, **270**, 18997–19007.
  76. Beke, D., Szekrényes, Z., Pálfi, D., Róna, G., Balogh, I., Maák, P.A., Katona, G., Czigány, Z., Kamarás, K., Rózsa, B. *et al.* (2013) Silicon carbide quantum dots for bioimaging. *J. Mater. Res.*, **28**, 205–209.
  77. Sang, P.B., Srinath, T., Patil, A.G., Woo, E.J. and Varshney, U. (2015) A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily. *Nucleic Acids Res.*, doi:10.1093/nar/gkv854.
  78. Requena, C.E., Perez-Moreno, G., Ruiz-Perez, L.M., Vidal, A.E. and Gonzalez-Pacanowska, D. (2014) The NTP pyrophosphatase DCTPP1 contributes to the homeostasis and cleansing of the dNTP pool in human cells. *Biochem. J.*, **459**, 171–180.
  79. Zhang, J. and Inouye, M. (2002) MazG, a nucleoside triphosphate pyrophosphohydrolase, interacts with Era, an essential GTPase in Escherichia coli. *J. Bacteriol.*, **184**, 5323–5329.

## Supporting data

### Detection of uracil within DNA using a sensitive labeling method for *in vitro* and cellular applications

Gergely Róna<sup>1,2,\*</sup>, Ildikó Scheer<sup>1,2</sup>, Kinga Nagy<sup>1,2</sup>, Hajnalka L. Pálincás<sup>1,3</sup>, Gergely Tihanyi<sup>1,2</sup>, Máté Borsos<sup>1</sup>, Angéla Békési<sup>1</sup> and Beáta G. Vértessy<sup>1,2,\*</sup>

<sup>1</sup> Institute of Enzymology, RCNS, Hungarian Academy of Sciences, H-1117 Budapest, Hungary;

<sup>2</sup> Department of Applied Biotechnology and Food Sciences, Budapest University of Technology and Economics, H-1111 Budapest, Hungary;

<sup>3</sup> Doctoral School of Multidisciplinary Medical Science, University of Szeged, H-6720 Szeged, Hungary

#### \*Corresponding authors:

Gergely Róna (rona.gergely@ttk.mta.hu) and Beáta G. Vértessy (vertessy.beata@ttk.mta.hu)

#### Author to communicate with the Editorial and Production offices:

Beáta G. Vértessy, phone: +3613826707, fax: +3614665465,

e-mail: vertessy@mail.bme.hu, Address: Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar Tudósok Str. 2, H-1117 Budapest, Hungary and Department of Applied Biotechnology and Food Sciences, Budapest University of Technology and Economics, Szt Gellért Square 4, H-1111 Budapest, Hungary

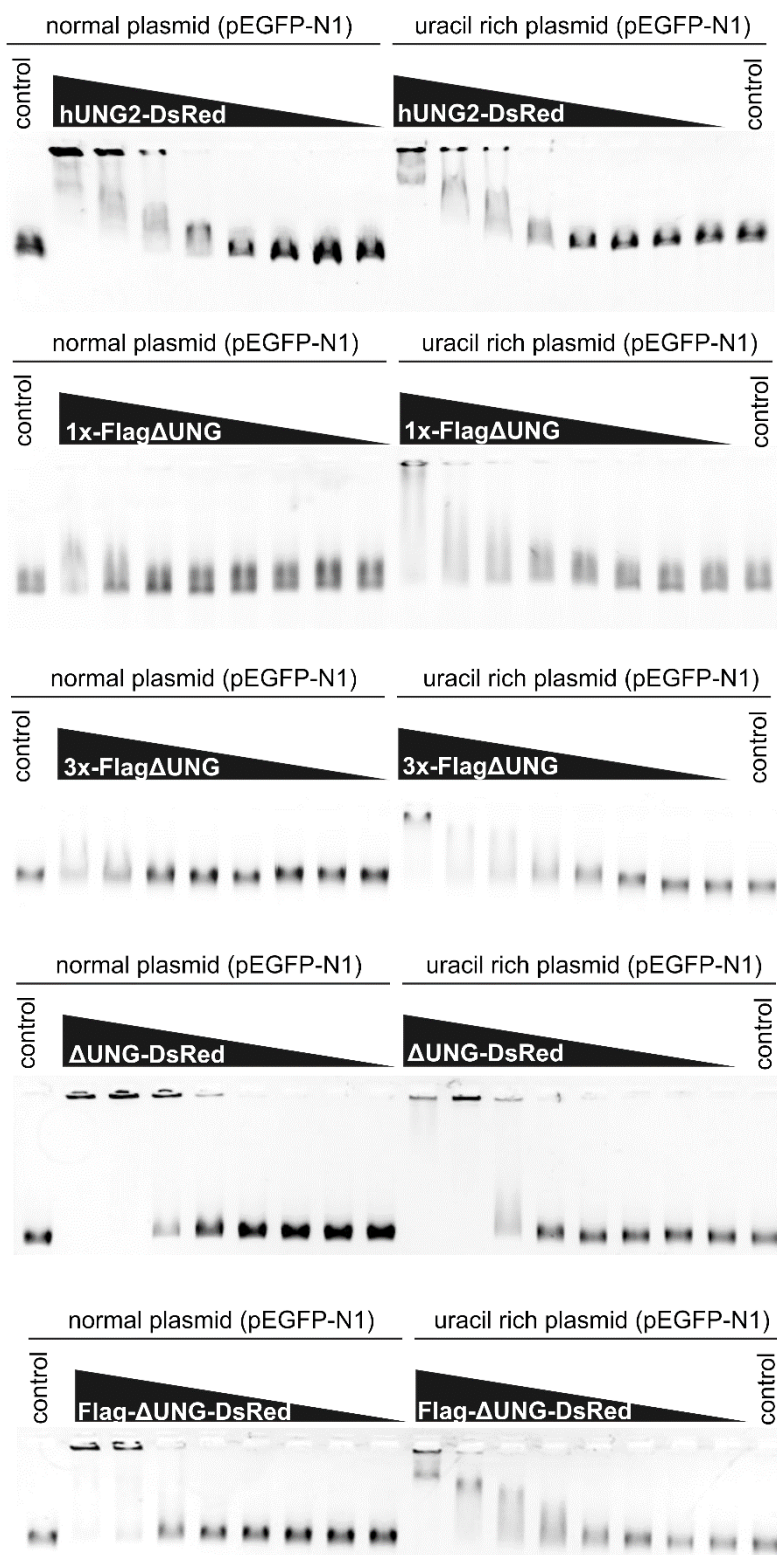
**Keywords:** uracil-DNA, quantification, dot blot, immunocytochemistry

#### Supporting Data include:

Supplemental data	Related figure	Significance
Figure S1	Figure 2	Electrophoretic Mobility Shift Assay (EMSA) of all used constructs in the study.
Figure S2	Figure 5	5FdUR treatment has no effect on [ <i>dut+</i> , <i>ung+</i> ] background.
Figure S3	Figure 7	Further control experiments regarding the genomic uracil staining in <i>E. coli</i>
Figure S4	Figure 8	Further control experiments regarding the uracil staining of plasmids in MEF cells.
Table S1	All figures	List of all the oligonucleotides used in the study.
Table S2	All figures	UNG and dUTPase genetic background of the used organisms in this study.

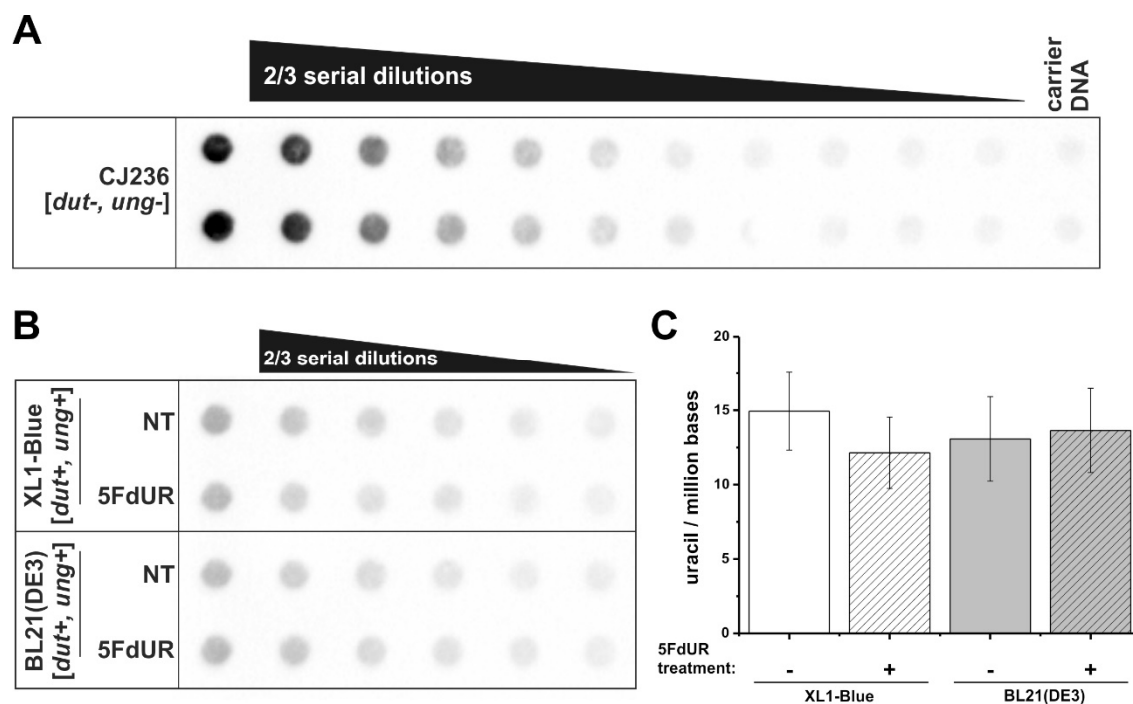
### Figure S1. Analysis of DNA-binding capability of the designed constructs

Uracil binding capability of the constructs were addressed with Electrophoretic Mobility Shift Assay (EMSA). All constructs clearly shift the position of the linearized vector in a concentration-dependent manner. The shift is more prominent in the case of uracil-rich template indicating that our constructs are capable of binding genomic uracil residues.



**Figure S2. Dot blot assay for measuring genomic uracil levels of 5FdUR treated *E. coli* [*dut*<sup>+</sup>, *ung*<sup>+</sup>] cells**

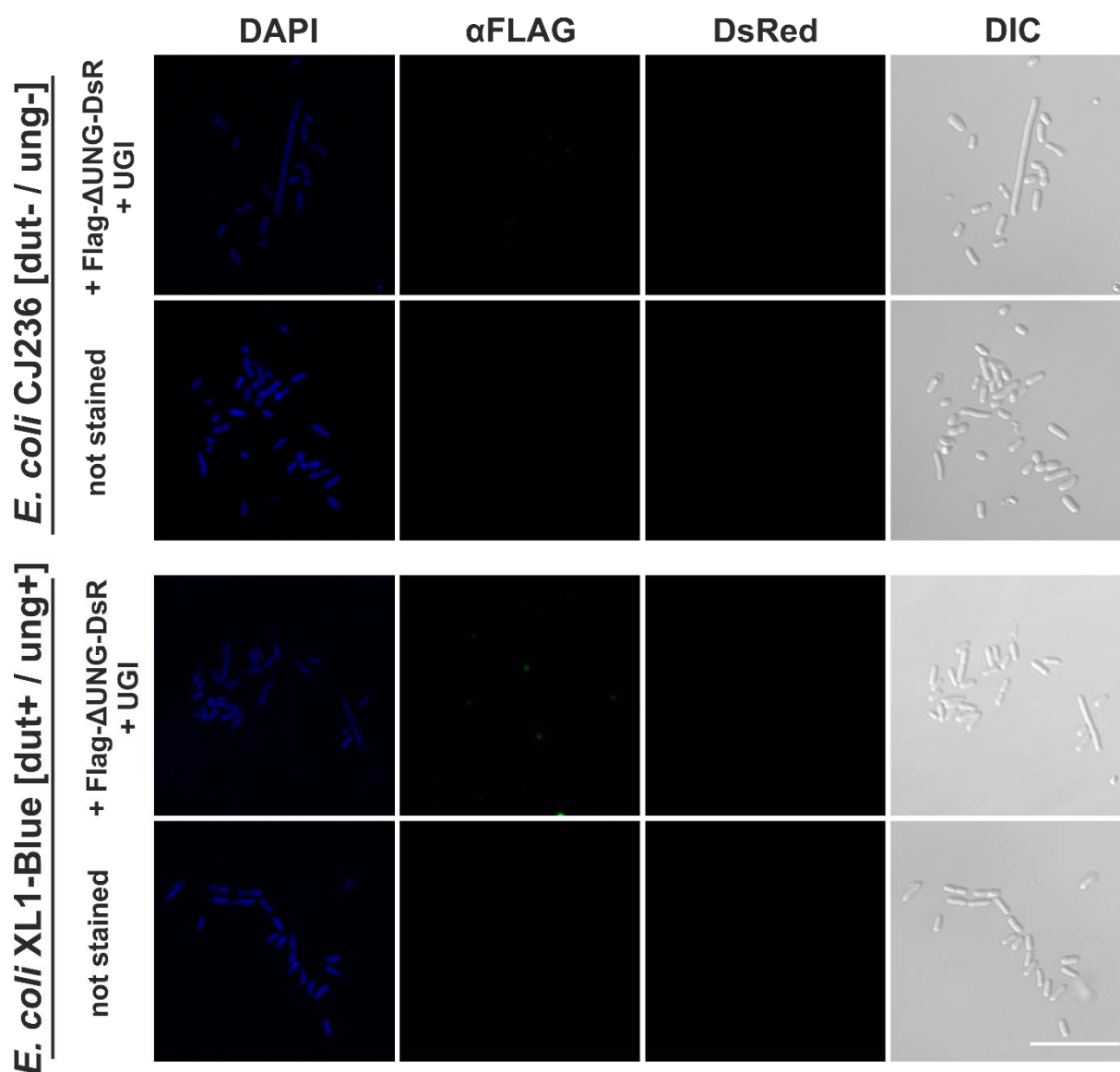
(A) CJ236 [*dut*<sup>-</sup>, *ung*<sup>-</sup>] *E. coli* genomic DNA was used as standard for the dot blot assay. (B) Quantity of genomic uracil of 5FdUR and non-treated XL1-Blue [*dut*<sup>+</sup>, *ung*<sup>+</sup>] (Stratagene) and BL21(DE3) [*dut*<sup>+</sup>, *ung*<sup>+</sup>] (NEB) samples were measured. (C) Bar graph shows the uracil moieties / million bases of each sample (mean values  $\pm$  the standard errors of the mean). There was no significant increase in uracil-DNA content with 5FdUR treatment as compared to non-treated cells. Genomic uracil level of XL1-Blue without 5FdUR were  $14.94 \pm 2.63$  and was  $12.14 \pm 2.39$  uracil / million bases for the 5FdUR treated cells. BL21(DE3) genomic uracil levels were  $13.07 \pm 2.84$  and  $13.64 \pm 2.84$  uracil / million bases for the non-treated and the 5FdUR treated cells respectively. Calculations were based on 3 independent datasets (n=3).





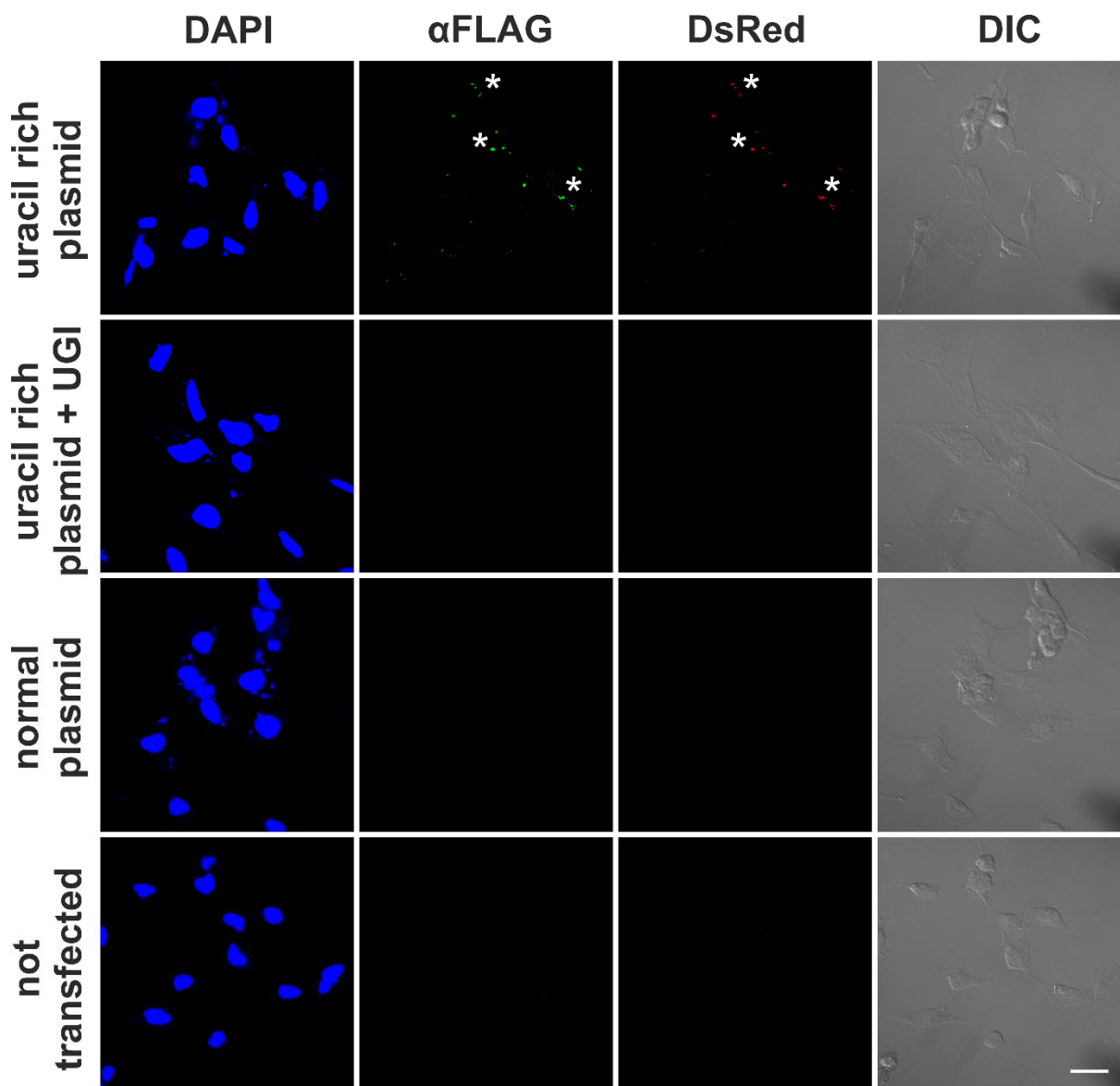
### Figure S3. Further control experiments to accompany Figure 8 in the main text

Further control stainings are shown. The genomic uracil content of CJ236 [*dut*<sup>-</sup>, *ung*<sup>-</sup>] *E. coli* cells could not be visualized with the Flag-ΔUNG-DsRed construct when UGI was applied simultaneously. Also without applying the Flag-ΔUNG-DsRed construct, there is no background signal from either the green channel (where primary (anti-FLAG) and secondary antibodies were applied), or the red channel (where DsRed fluorescence is monitored). As expected, no staining was detected in either case with the XL1-Blue cells [*dut*<sup>+</sup>, *ung*<sup>+</sup>] which were applied as negative controls. DAPI was used to counterstain DNA. Scale bar represents 10 μm.



**Figure S4. Further control experiments to accompany Figure 8 in the main text**

Further control stainings are shown for Figure 8, where cytoplasmic plasmid aggregates are visualized by the Flag- $\Delta$ UNG-DsRed construct. Asterisks (\*) show plasmid aggregates. Only cells transfected with uracil-rich plasmids could be visualized by the construct through both the DsRed (red) tag and the FLAG epitope tag (green). If UGI was applied or if cells were transfected with normal plasmids, there was no observable signal in either channel, although DAPI image clearly show the presence of plasmids aggregated in the cytoplasm. If cells were not transfected (last row), DAPI staining does not show cytoplasmic plasmid aggregates as expected. Scale bar represents 20  $\mu$ m.



**Table S1. Oligonucleotides used in the present study**

Restriction sites are underlined.

Used in	Oligo name	Oligonucleotides 5'-3'
Cloning full length UNG-DsRed into pET-20b	UNG_F	GGAGGCG <u>CATATG</u> ATCGGCCAGAAGACGC
	UNG_R	CCG <u>CTCGAG</u> CTGGGAGCCGGAGTGG
Mutagenesis of UNG	D154N_F	GGTTGTCATCCTGGGACAGAATCCATATCATGGACCTAATC
	D154N_R	GATTAGGTCCATGATATGGATTCTGTCCCAGGATGACAACC
	H277N_F	CCATGTACTACAGACGGCTAATCCCTCCCCTTTGTCAG
	H277N_R	CTGACAAAGGGGAGGGATTAGCCGTCTGTAGTACATGG
Cloning of 1x/3x FLAG-ΔUNG into pET-15b	1xFLAG_F	ATTACATATGGACTACAAAGACGATGACGACAAACCACTCGCGGCCCGCAACG
	3xFLAG_F	ATTACATATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGCCCACTCGCGGCCCGCAACG
	1x/3xFLAG_R	GGTC <u>CTCGAG</u> TTAGATGTACCTGTAGGTGTCCAGCTCCTTCC
Cloning of FLAG-ΔUNG-DsRed and ΔUNG-DsRed	1xFLAG_Ds_F	ATTACATATGGACTACAAAGACGATGACGACAAACCACTCGCGGCCCGCAACG
	Ds_F	ATTACATATGCTCGCGGCCCGCAACG
	Ds_R	TCGACTCGAGCTGGGAGCCGGAGTGG

**Table S2. UNG and dUTPase genetic background of the organisms used in the study**

Organism	Uracil-DNA glycosylase (UNG)	dUTPase (DUT)
<i>E. coli</i> XL1-Blue	+	+
<i>E. coli</i> BL21(DE3) <i>ung-151</i>	-	+
<i>E. coli</i> CJ236**	-	-
<i>D. melanogaster</i> Schneider S2 cells	*	+
Mouse Embryonic Fibroblasts (MEF), UNG KO	-	+
HCT116 cell line	+	+

+ functional gene

- disrupted gene

\* gene not found in the genome




\*\* CJ236 *E. coli* strain contains a point mutation in the *dut* gene generating a Thr24Ile amino acid change (1). This results in 2.2-3.3 % residual dUTPase activity compared to the wild type enzyme (2). This strain also contains a single base deletion in the *ung* gene leading to a frame shift mutation thus the disruption of *ung* gene (3).

1. Lari, S.U., Chen, C.Y., Vertessy, B.G., Morre, J. and Bennett, S.E. (2006) Quantitative determination of uracil residues in Escherichia coli DNA: Contribution of ung, dug, and dut genes to uracil avoidance. *DNA Repair (Amst)*, **5**, 1407-1420.
2. Kouzminova, E.A. and Kuzminov, A. (2004) Chromosomal fragmentation in dUTPase-deficient mutants of Escherichia coli and its recombinational repair. *Mol Microbiol*, **51**, 1279-1295.
3. Varshney, U. and van de Sande, J.H. (1989) Characterization of the ung1 mutation of Escherichia coli. *Nucleic Acids Res*, **17**, 813.

## **Annex II.**

## Article

# CRISPR/Cas9-Mediated Knock-Out of dUTPase in Mice Leads to Early Embryonic Lethality

Hajnalka Laura Pálincás <sup>1,2,3,\*</sup>, Gergely Attila Rácz <sup>1,3</sup>, Zoltán Gál <sup>4</sup>, Orsolya Ivett Hoffmann <sup>4</sup>, Gergely Tihanyi <sup>1,3</sup> , Gergely Róna <sup>3,5,6</sup>, Elen Gócza <sup>4,\*</sup> , László Hiripi <sup>4,\*</sup>  and Beáta G. Vértessy <sup>1,3,\*</sup>

<sup>1</sup> Institute of Enzymology, RCNS, Hungarian Academy of Sciences, H-1117 Budapest, Hungary; racz.gergely@ttk.mta.hu (G.A.R.); tihanyi@stud.uni-heidelberg.de (G.T.)

<sup>2</sup> Doctoral School of Multidisciplinary Medical Science, University of Szeged, H-6720 Szeged, Hungary

<sup>3</sup> Department of Applied Biotechnology and Food Sciences, Budapest University of Technology and Economics, H-1111 Budapest, Hungary; Gergely.Rona@nyulangone.org

<sup>4</sup> Department of Animal Biotechnology, Agricultural Biotechnology Institute, National Agricultural Research and Innovation Centre, H-2100 Gödöllő, Hungary; zoltan.gal89@gmail.com (Z.G.); hoffmannorsi@gmail.com (O.I.H.)

<sup>5</sup> Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY 10016, USA

<sup>6</sup> Perlmutter Cancer Center, New York University School of Medicine, New York, NY 10016, USA

\* Correspondence: palinkas.hajnalka@ttk.mta.hu (H.L.P.); gocza.elen@abc.naik.hu (E.G.); hiripi1@gmail.com (L.H.); vertessy@mail.bme.hu (B.G.V.)

Received: 14 March 2019; Accepted: 2 April 2019; Published: 4 April 2019



**Abstract:** Sanitization of nucleotide pools is essential for genome maintenance. Deoxyuridine 5'-triphosphate nucleotidohydrolase (dUTPase) is a key enzyme in this pathway since it catalyzes the cleavage of 2'-deoxyuridine 5'-triphosphate (dUTP) into 2'-deoxyuridine 5'-monophosphate (dUMP) and inorganic pyrophosphate. Through its action dUTPase efficiently prevents uracil misincorporation into DNA and at the same time provides dUMP, the substrate for de novo thymidylate biosynthesis. Despite its physiological significance, knock-out models of dUTPase have not yet been investigated in mammals, but only in unicellular organisms, such as bacteria and yeast. Here we generate CRISPR/Cas9-mediated dUTPase knock-out in mice. We find that heterozygous *dut* +/− animals are viable while having decreased dUTPase levels. Importantly, we show that dUTPase is essential for embryonic development since early *dut* −/− embryos reach the blastocyst stage, however, they die shortly after implantation. Analysis of pre-implantation embryos indicates perturbed growth of both inner cell mass (ICM) and trophectoderm (TE). We conclude that dUTPase is indispensable for post-implantation development in mice.

**Keywords:** dUTPase; CRISPR/Cas9-mediated knock-out; blastocyst outgrowth; embryonic development

## 1. Introduction

The maintenance of genome integrity and faithful preservation of genomic information are crucial for viability. Toward these goals, various DNA damage and repair pathways along with the regulation of a well-balanced deoxynucleotide (dNTP) pool work hand in hand [1]. Nucleotide pools are maintained by several families of dNTP hydrolyzing enzymes present in most organisms [2–5]. These enzymes sanitize the nucleotide pool by removing nucleotide building blocks (dNTPs) that contain erroneous bases from the polymerase action. The deoxyuridine 5'-triphosphate nucleotidohydrolase (dUTPase) family of enzymes is responsible for the removal of dUTP from the nucleotide pool by hydrolyzing it into dUMP and inorganic pyrophosphate [6–8]. The importance of this enzymatic

action is evident in light of the fact that most DNA polymerases cannot distinguish 2'-deoxyuridine 5'-triphosphate (dUTP) and 2'-deoxythymidine 5'-triphosphate (dTTP) and will readily incorporate the uracil analog if it is available in the cellular dNTP pool [9]. Through their enzymatic action that generates 2'-deoxyuridine 5'-monophosphate (dUMP), dUTPases also feed into the de novo thymidylate biosynthesis pathway by providing dUMP as the substrate for thymidylate synthase. Two major families of dUTPases have evolved that are referred to as trimeric and dimeric dUTPases, reflecting their corresponding quaternary structure [6,10–13].

Trimeric dUTPases are present in almost all free-living organisms with the notable exception of trypanosomes. Subunits of these enzymes contain a beta-sheeted arrangement [6]. The three subunits donate conserved sequence motifs to build the three active site of the dUTPase trimer [14–16]. This family of dUTPases is characteristic for Archaea, Bacteria, and Eucarya. As such, mammalian species also rely on the action of trimeric dUTPases to keep a well-balanced dUTP/dTTP ratio. Herpesviruses encode an intriguing monomeric homolog of this dUTPase enzyme family [17], where the protein sequence contains a species-specific insert to allow the construction of the usual beta-sheeted dUTPase fold in a monomeric enzyme [18,19]. Dimeric dUTPases perform the same catalytic action, however, the protein sequence and the alpha-helical protein fold are drastically different from those observed in the trimeric dUTPase family [10,11].

Due to the highly significant enzymatic character of dUTPases, the essentiality of this enzyme family was addressed in numerous different organisms. Knock-outs have been generated in several bacteria: *Escherichia coli* [20] and *Mycobacterium smegmatis* [7]. Based on these studies it was argued that the bacterial cells lacking dUTPase activity are not viable. However, genomic analysis of Archaea and prokaryotes identified several species that lack *dut*, the dUTPase encoding gene [12]. These findings indicate that the presence of dUTPases may not be a universal requirement in prokaryotes.

The physiological role and importance of dUTPase have also been addressed in eukaryotes. In yeast, dUTPase knock-out was still viable, although this genotype led to a thymine auxotroph phenotype [21]. In *Caenorhabditis elegans*, RNA-silencing studies indicated that dUTPase might be important in embryonic development [22]. Very recently, in planarians, silencing of *dut* caused lethality in adult animals possibly due to genomic DNA fragmentation. Co-administration of the thymidylate synthase inhibitor 5-fluoro-uracil (5-FU) resulted in more DNA breaks and earlier planarian death [23]. In *Drosophila melanogaster*, dUTPase silencing led to early pupal lethality suggesting a specific role of dUTPase and uracil-DNA metabolism in metamorphosing insects [24,25]. It has been shown that dUTPase is also essential in *Arabidopsis thaliana*. In these plants, reduced dUTPase activity caused DNA damage and increased homologous recombination events. Furthermore, these plants were extremely sensitive to 5-FU [26]. In human cell lines, several laboratories published siRNA dUTPase-silencing studies [27–29]. These all proposed that highly efficient silencing with practically no remaining dUTPase does not perturb the cellular phenotype under normal conditions [29]. Still, the dUTPase-silenced cell lines showed increased sensitivity towards inhibitors of the de novo thymidylate biosynthesis. These findings corroborated the clinical significance of dUTPase inhibition in anticancer chemotherapies [30–32]. To our best knowledge, knock-out studies on dUTPases have not yet been published for any mammalian species.

Motivated by the lack of knowledge in the field, we initiated dUTPase knock-out experiments in mice. Here we report successful generation of dUTPase knock-out mice using CRISPR/Cas9-mediated genome editing. We find that absence of dUTPase leads to early embryonic lethality. No homozygous knock-out offspring could be observed, however, homozygous knock-out blastocysts are still viable and can be cultured in vitro, suggesting that lethality of the dUTPase knock-out sets in around or shortly after implantation.

## 2. Materials and Methods

### 2.1. CRISPR Constructs

The T7 single-guide RNA (sgRNA) harboring the protospacer-adjacent motif (PAM) sequence and the Cas9 mRNA were obtained from Sigma-Aldrich (St. Louis, MO, USA). According to our request, the designed sgRNA targeted the first common exon of dUTPase isoforms (exon 2) on mouse chromosome 2 (Figure 1a).

### 2.2. CRISPR/Cas9 Efficiency Test in Mouse Embryonic Fibroblast (MEF) Cells

Target sgRNA and Cas9 nuclease mRNA were transfected into mouse embryonic fibroblast (MEF) cells by Lipofectamine™ 3000 Transfection Reagent (ThermoFisher Scientific, Waltham, MA, USA). According to the manufacturer's recommendation, 2.5 µg Cas9 mRNA and 250 ng target sgRNA were added to the sub-confluent cultures of cells grown in six-well plates. 24 h after transfection, cells were maintained in a fresh medium for 24 h, and then the genomic DNA was extracted with a MasterPure™ DNA Purification Kit (Epicentre, Madison, WI, USA). After DNA amplification with Cel-1-F and Cel-1-R primers, Cel 1 cleavage assay was performed using the Transgenomic® SURVEYOR® Mutation Detection Kit according to the manufacturer's instructions. The primers used in this study were synthesized by Sigma-Aldrich and are listed in Table S1.

### 2.3. Animals

dUTPase wild-type and heterozygous mice used in the experiments were produced and maintained in the Animal Care Facility at the Agricultural Biotechnology Institute, National Agricultural Research and Innovation Centre (NAIK) (FVB/N background, Envigo, UK). Animals were housed in groups of 2–5 with free access to food and water. Animals were kept under a standard light–dark cycle (06.00–18.00 h) at 22 °C. This study was carried out in strict accordance with the recommendations and rules in the Hungarian Code of Practice for the Care and Use of Animals for Scientific Purposes. The protocol was approved by the Animal Care and Ethics Committee of the Agricultural Biotechnology Institute, NAIK and the Pest County's governmental office (permission number: PEI/001/329-4/2013). The method used for euthanasia was cervical dislocation. All efforts were made to minimize suffering.

### 2.4. Micromanipulation and Detection of Gene Targeting

Microinjection was performed as described previously [33]. Briefly, mouse zygotes were collected at 20 h after injection of human chorionic gonadotropin (hCG) from superovulated FVB/N females mated with FVB/N males. Pronuclei were injected using a manual injector with continuous flow. Following visualization of pronuclear swelling, the needle was pulled out through the cytoplasm, injecting a small amount of additional RNA delivery to the cytoplasm. The microinjection mix contained a sgRNA (Sigma-Aldrich, St. Louis, MO, USA) diluted in 10 mM Tris, 0.2 mM EDTA (pH = 7.4) in a final concentration of 10 ng/µL and Cas9 mRNA (Trilink, San Diego, CA, USA) in 10 mM Tris, 0.2 mM EDTA (pH = 7.4) in a final concentration of 150 ng/µL. Microinjections were finished within 2 h after zygote isolation. Injected zygotes were transferred to pseudopregnant CD-1 females (Envigo, Huntingdon, UK). All animals born from embryo transfer were genotyped by polymerase chain reaction (PCR) and T7 assay.

### 2.5. Cloning and Sequencing

Genotyping of the heterozygous founder animals was carried out by amplifying the CRISPR target sites from genomic DNA using primers pBS-F and pBS-R (listed in Table S1), and the fragments were cloned into *SalI*/*EcoRI* sites of vector pBluescript SK (+) (Stratagene). Twenty individual bacterial colonies were purified with NucleoSpin® Plasmid DNA Purification Kit (MACHEREY-NAGEL GmbH



& Co. KG (Düren, Germany), according to the manufacturer's instructions, then DNA samples were subjected to sequencing. Based on the sequencing results, two animals (founder #2 and #4) showed CRISPR events, their offspring were termed (D6, M1) and D47, respectively. All DNA samples in this study were verified by sequencing by Microsynth SeqLab GmbH.

## 2.6. Off-Target Analysis

Off-target effects of CRISPR/Cas9 nucleases were evaluated via the online predictor CCTop—CRISPR/Cas9 target (<https://crispr.cos.uni-heidelberg.de/>) [34]. Twelve candidate loci for the target site with high potential cleavage in the mouse genome were chosen. The selected potential off-target sites were PCR-amplified using genomic DNA from both wild-type mice and the founder animal #4 and evaluated by DNA sequencing. Ten of the twelve candidate off-target sites could be analyzed. In all these cases no difference was observed (see Figure S2). Two of the candidate off-target sites (Off-2 and Off-4) could not be evaluated due to non-specific PCR product from repetitive elements in both the wild-type and the founder #4 samples. The information on the off-target loci identified by the online program are shown in Table 1 and sequencing primer pairs used are listed in Table S1.

## 2.7. Genotyping

The genotypes of mice were determined by PCR of the total genomic DNA extracted from mouse tails. Genotyping of embryos was also performed by PCR either by isolating DNA from full embryos or from outgrowth assays. All isolated samples were dissolved in a DNA lysis buffer (0.1 M Tris-HCl, pH = 7.4, 0.2 M NaCl, 5 mM EDTA, 0.2% SDS) and DNA was extracted with phenol–chloroform. MyTaq polymerase (Bioline) was activated at 95 °C for 5 min, and PCR was performed for 30 cycles at 95 °C for 30 s, 64 °C for 30 s, and 72 °C for 30 s, with a final extension at 72 °C for 10 min using primers Dut-gen-F and Dut-gen-R (Table S1). Genotyping from blastocysts was performed by semi-nested PCR using 1 µL template from 30× diluted primary PCR product with primers Dut-nest-F and Dut-gen-R (Table S1) under the same reaction conditions. DNA fragments were visualized by 1% agarose gel electrophoresis.

## 2.8. Analysis of Dissected Embryos

For the analysis presented in this manuscript, D47 heterozygous males were mated with D47 heterozygous females, and embryos at various stages (3.5–9.5 dpc (days post coitum)) were collected from pregnant D47 heterozygous females. Dissections were performed in ice-cold phosphate buffered saline (PBS). After dissection, embryos were examined and photographed with a LeicaM205FCA-FC fluorescent stereo microscope linked to a DFC7000-T Leica camera. The day of plug formation was defined as embryonic day 0.5.

## 2.9. Analysis of Blastocyst Outgrowth

Embryos were flushed out from the uteri of pregnant mice at 3.5 dpc in M2 medium. Blastocysts were individually cultured on 0.1% gelatin-coated, 12-well tissue culture dishes (Eppendorf), in KO-DMEM ES cell culture medium supplemented with 1000 U/mL LIF and 20% fetal bovine serum (HyClone), in 5% CO<sub>2</sub> at 37 °C for four days. Outgrowths were photographed daily using a LeicaM205FCA-FC fluorescent stereo microscope linked to a DFC7000-T Leica camera. On the fourth day of culture, outgrowths were photographed, subsequently removed, and genotyped by PCR as described above.

### 2.10. Western Blot

Embryos at 10.5 dpc were dissected immediately following euthanasia of pregnant mice, then washed with PBS and resuspended in a lysis buffer (20 mM HEPES, pH = 7.5, 420 mM NaCl, 1 mM EDTA, 2 mM dithiothreitol (DTT), 25% glycerol). Homogenization was assisted with vortex until the tissue was sufficiently disrupted. Samples were centrifuged at  $20,000\times g$  for 15 min at 4 °C to remove the insoluble fraction, then the supernatant samples were boiled with SDS buffer at 95 °C for 5 min. Total proteins were resolved under denaturing and reducing conditions on a 12% polyacrylamide gel and transferred to the PVDF membrane (Immobilon-P, Merck Millipore, Billerica, MA, USA). Membranes were blocked with 5% non-fat dried milk in TBS-T (25 mM Tris-HCl, pH = 7.4, 140 mM NaCl, 3 mM KCl, 0.05% Tween-20) for 1 h at 4 °C and were developed against dUTPase (1:2000, Sigma-Aldrich) and  $\alpha$ -actin (1:1000, Sigma-Aldrich) for loading control. After applying horseradish peroxidase-coupled secondary antibodies (Amersham Pharmacia Biotech, Piscataway, NJ, USA), immunoreactive bands were visualized by an enhanced chemiluminescence reagent (Millipore, Western Chemiluminescent HRP substrate), and images were captured by a Bio-Rad ChemiDoc™ MP Imaging System. Densitometry was done using Bio-Rad Image Lab™ 6.0 (Hercules, CA, USA).

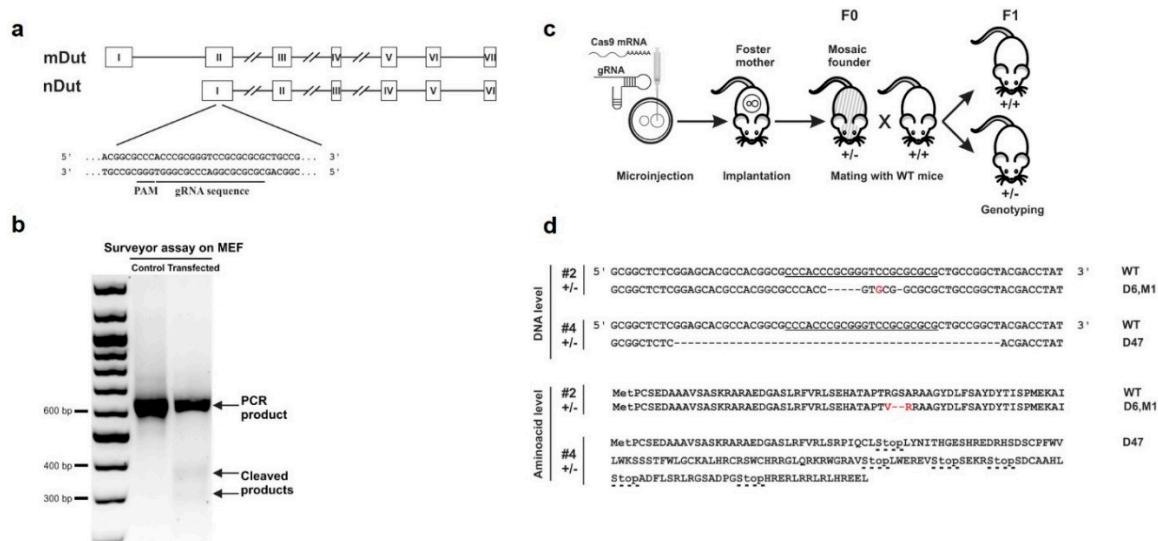
### 2.11. Statistical Analysis

Statistical analysis was done with a two-sample, single-tailed t-test assuming equal variance using Microsoft Excel or was carried out with a two-sided Mann–Whitney U test. The data were considered significant when  $p < 0.05$  (\*).

## 3. Results

### 3.1. Targeted Knock-Out of Mouse dUTPase by CRISPR/Cas9 Gene Editing

To explore the importance of dUTPase in mammalian life, we attempted to establish dUTPase knock-out mice using CRISPR/Cas9-mediated gene editing. Figure 1 shows the outline of the applied knock-out strategy. The respective single-guide RNA was designed to disrupt both nuclear and mitochondrial isoforms of dUTPase (Figure 1a), which arise from a single gene (*dut*) on chromosome 2. Efficiency of the CRISPR-mediated events were first assessed by the surveyor assay in mouse embryonic fibroblast (MEF) cells (Figure 1b). We found evidence of CRISPR/Cas9-induced cleaved products and, based on this result, we started the mouse zygote microinjection. Figure 1c shows the schematics of gene targeting and mice generation. In the microinjection experiments, 107 embryos were flushed and microinjected, then 76 embryos from these 107 embryos were transferred to five foster mothers. Two foster mothers had in sum 15 offspring (resulting in a 20% survival rate). Three out of the 15 offspring were successfully targeted by CRISPR/Cas9 based on T7 assay (20% targeted rate), and two of them (these are numbered as #2 and #4) showed mono-allelic indels verified by sequencing. Sequencing results from mouse #2 showed a six bp deletion and one base substitution leading to altered amino acid sequences. Sequencing results from mouse #4 showed a 47 bp deletion leading to frameshift and early stop codons (Figure 1d). Chromatograms of the sequencing could be found in Figure S1. Heterozygous mouse #2 and mouse #4 were the founder (F0) animals for the downstream mouse strains. We have termed the strain from the founder mouse #2 as (D6, M1) and the strain from the founder mouse #4 as D47. Genotype was determined by PCR throughout the study (see ‘Materials and Methods’ section for details). Founder mouse #2 and chimeric founder mouse #4 were then bred to a wild-type mouse for germline transmission thus generating progeny containing the targeted gene. Homozygous and heterozygous offspring of strain (D6, M1), as well as heterozygous offspring of strain D47 showed no gross visually observable abnormalities and were fertile through multiple generations. Further experiments presented in this study were conducted on F2 or even later generations of the mouse strain D47, thereby excluding the possibility of mosaicism.



**Figure 1.** Generation and assessment of CRISPR knock-out mice. **(a)** Schematic diagram of the *dut* gene encoding the nuclear (nDut) and mitochondrial (mDut) isoforms of deoxyuridine 5'-triphosphate nucleotidohydrolase (dUTPase). Exons are indicated with Roman numerals in rectangles, introns are simplified as lines (for longer introns lines are broken). Guide RNA (gRNA) target site and protospacer-adjacent motif (PAM) sequence in the first common exon of the two isoforms are underlined. **(b)** Surveyor assay performed on mouse embryonic fibroblast (MEF) cells used for the detection of indel events induced by transfection with CRISPR gRNA and Cas9 mRNA. The two lower fragments indicate cleavage of the DNA due to CRISPR activity. These are lacking in the control while they are visible in the transfected sample. **(c)** Schematic diagram showing the generation of CRISPR-targeted knock-out mice. Fertilized oocytes microinjected with gRNA and Cas9 mRNA were implanted into foster mothers. The resulting founders (F0) #2 and #4 were cross-bred with wild-type (WT) mice to generate wild-type (*+/+*) and heterozygous (*+/−*) offspring (F1) containing the targeted locus through germline transmission. **(d)** DNA and predicted amino acid sequence of the two heterozygous founder mice (#2 and #4) showing CRISPR events, compared to the WT. Mouse #2 showed deletion of six nucleotides and a C to G mutation (D6, M1) resulting in the deletion of two amino acids and change of another two. In mouse #4, 47 nucleotides were deleted (D47) which resulted in a frameshift mutation leading to early stop codons indicated with dashed lines. CRISPR target site including PAM sequence is underlined.

In our CRISPR-based knock-out experiments, the target site was situated within the *dut* gene, as shown in Figure 1a. The sequence of this target site was as follows: CGCGCGCGGACCCGCGGGT. To check for potential off-target effects, we had carefully analyzed the twenty most similar sequences to this target sequence within the mouse genome with the online predictor software CCTop [34] (see 'Materials and Methods' section for details) (Table 1). In each similar sequence, at least three or more mismatches occurred reducing the chance of off-target effects [35]. To decide whether any of the first ten sequences may lead to off-target cleavage using the CRISPR/Cas9 method, we sequenced these genomic segments both in the wild-type and in the founder mouse #4. We also sequenced the 16th and 18th most similar genomic segments as they were located on the same chromosome (chr2) as the *dut* gene, thus, an off-target effect in these segments could be inherited together with the *dut* knock-out allele. Sequences from the wild-type and the mouse #4 were identical (see Figure S2 for the sequencing chromatograms). However, two of the candidates' off-target sites (Off-2 and Off-4) could not be evaluated due to non-specific PCR product in both the wild-type and the founder #4 samples. We therefore concluded that no CRISPR-induced event could be observed at these successfully sequenced sites, arguing for lack of off-target effects.

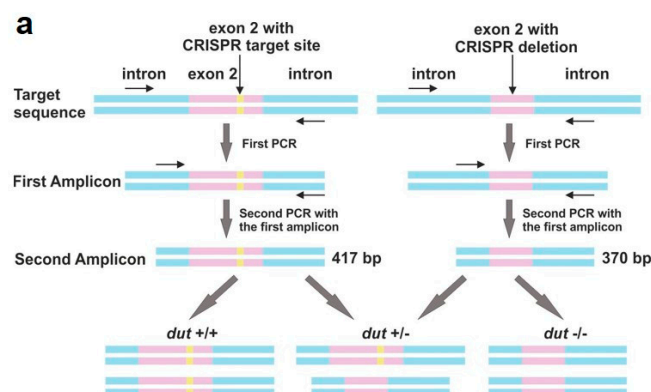
**Table 1.** Potential off-target sites predicted by CCTop—CRISPR/Cas9 target online predictor software.

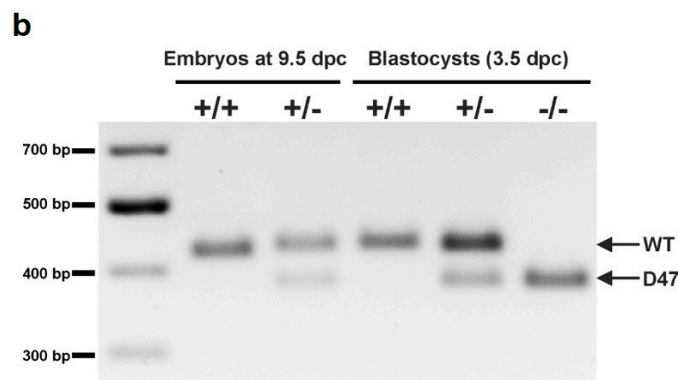
Name	Coordinates	MM	Target Sequence	PAM
<b>Dut</b>	<b>chr2:125247853-125247874</b>	<b>0</b>	<b>CGCGCGCGGAC[CCGCGGGT]</b>	<b>GGG</b>
<b>Off-1</b>	<b>chr13:88680616-88680637</b>	<b>4</b>	<b>TGAGTGGGGAC[CCGCGGGT]</b>	<b>TGG</b>
Off-2	chr4:120746882-120746903	4	CTGGAGCGGCC[CCGCGGGT]	GGG
<b>Off-3</b>	<b>chr15:28025084-28025105</b>	<b>4</b>	<b>CGGCCACGGCC[CCGCGGGT]</b>	<b>AGG</b>
Off-4	chr11:23306808-23306829	4	GGGGCGGGGAG[CCGCGGGT]	GGG
<b>Off-5</b>	<b>chr2:118372992-118373013</b>	<b>4</b>	<b>CGCTGGTGGCC[CCGCGGGT]</b>	<b>TGG</b>
<b>Off-6</b>	<b>chr5:64803646-64803667</b>	<b>4</b>	<b>ACCGCACGGAC[GGCGGGT]</b>	<b>GGG</b>
<b>Off-7</b>	<b>chr18:85179644-85179665</b>	<b>3</b>	<b>CGTGC GCGCAC[GGCGGGT]</b>	<b>GGG</b>
<b>Off-8</b>	<b>chr9:77319229-77319250</b>	<b>4</b>	<b>CGCGCTTACAC[CCGCGGGT]</b>	<b>GGG</b>
<b>Off-9</b>	<b>chr2:104319532-104319553</b>	<b>3</b>	<b>CGCGTGCGCAC[ACGCGGGT]</b>	<b>AGG</b>
<b>Off-10</b>	<b>chr19:36918725-36918746</b>	<b>4</b>	<b>CTCGCTGGGAC[GGCGGGT]</b>	<b>AGG</b>
Off-11	chr5:75044665-75044686	4	TGGGCGCGGGC[GGCGGGT]	GGG
Off-12	chr4:152086570-152086591	4	CGCACCCAGAC[ACGCGGGT]	CGG
Off-13	chr3:41563582-41563603	4	GGCGCGGGGGC[GGCGGGT]	CGG
Off-14	chr8:60640130-60640151	4	GGCGCGTGGGC[ACGCGGGT]	TGG
Off-15	chr9:40192333-40192354	3	CGCGCGGGGCC[CAGCGGGT]	CGG
<b>Off-16</b>	<b>chr2:174438958-174438979</b>	<b>4</b>	<b>AGCGCGTGGGC[CTGCGGGT]</b>	<b>CGG</b>
Off-17	chr17:88792070-88792091	4	CGGGCGGGGGC[CGGCGGGT]	GGG
<b>Off-18</b>	<b>chr2:28641663-28641684</b>	<b>4</b>	<b>GGCACGGGGAC[CCGCGGGT]</b>	<b>GGG</b>
Off-19	chr17:28350853-28350874	4	GGCGGGCGGGC[CCACGGGT]	GGG
Off-20	chr5:107597539-107597560	4	GGCGCGTGGAT[CGGCGGGT]	AGG

The table presents the chromosomal location, the number of mismatches (MM), the target, and the adjacent PAM sequences. The first row depicts the target site of the designed gRNA, further rows list the top 20 candidates for off-target sites. Mismatches are indicated in red. Brackets include core sequences [34]. The genomic segments that were successfully sequenced are shown in bold.

### 3.2. Analysis of Developmental Effects of *dUTPase* Knock-Out

To assess the genotype of the embryos and progeny from *dut* +/− mouse intercross breeding we designed appropriate primers for genotyping PCR reactions. The resulting products were different in length from the wild-type as compared to the knock-out allele in mouse strain D47 (as visualized in Figure 2a). This semi-nested PCR method proved to be efficient for genotyping from low cell number containing samples, like blastocysts. *Dut* +/+ and +/− embryos could be detected at 9.5 dpc (Figure 2b) and were obtained by intercrossing D47 heterozygous mice. However, *dut* −/− embryos could only be observed at early pre-implantation stages. We revealed that all three genotypes (+/+, +/−, −/−) resulted in live 3.5-day-old embryos (Figure 2b).

**Figure 2.** Cont.



**Figure 2.** Genotyping of blastocysts. (a) Schematic representation of the used semi-nested design for genotyping. Introns are shown in blue, exons are shown in pink, and the CRISPR target site is shown in yellow. DNA isolated from blastocysts was subjected to PCR with primers (shown as arrows) adjacent to the CRISPR target site. The resulting amplicon was used in a second round of PCR with the same reverse and a nested inner forward primer to generate a 417 bp length product from the WT allele and a 370 bp product from the D47 allele. (b) Representative image of amplicons from semi-nested PCR visualized on agarose gel. The upper and lower band correspond to WT and D47 allele, respectively. Full-length agarose gel is included in the Supplementary Materials (Figure S7).

Table 2 summarizes the results of the genotype analysis of the animals investigated in this study. The data clearly showed that *dut*  $-/-$  knock-out cells could only be isolated in early embryonic development at the blastocyst stage (3.5 dpc).

**Table 2.** Genotype analysis of offspring from *dut*  $+/-$  intercrosses at different developmental stages.

DNA Source	Genotype			Resorbed	No. Total
	$+/+$	$+/-$	$-/-$		
Postnatal	21	42	0	NA <sup>a</sup>	63
10.5 dpc	3	5	0	3	11
9.5 dpc	5	5	0	0	10
8.5 dpc	10	5	0	5	20
3.5 dpc	11	13	7	NA	31

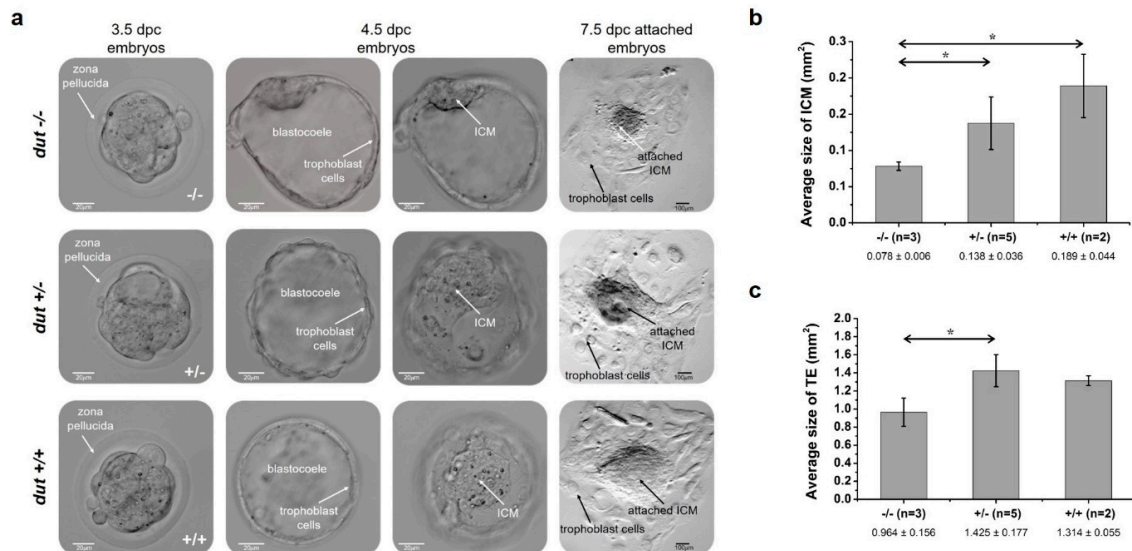
<sup>a</sup> NA, not applicable.

### 3.3. Embryonic Development in the dUTPase Knock-Out as Compared to the Heterozygous and Wild-Type Animals

To further assess early embryonic development of dUTPase knock-out embryos, we cultured *in vitro* for several days of 3.5 dpc blastocysts derived from intercrosses of *dut*  $+/-$  mice. After hatching, these blastocysts were attached to a gelatin-coated surface (Figure 3). Blastocyst development was checked after one day and four days (i.e., at 4.5 and at 7.5 dpc). Attached blastocysts formed an inner cell mass (ICM) outgrowth around which trophoblast giant cells were visible (indicated with arrows in Figure 3a). It is shown in these pictures that the *dut*  $-/-$  ICM was smaller, while we did not observe obvious changes in the heterozygotes as compared to the wild-type. Additional phase contrast images of embryos can be found in Figure S3. Quantitative analysis of ICM and trophectoderm (TE) regions were performed as described in the Supplementary Materials (Figure S4). Further development of the attached embryos revealed that at 7.5 dpc, the ICM clump size was significantly smaller in the *dut*  $-/-$  embryo as compared to the heterozygous ( $p = 0.045$ ) and the wild-type ( $p = 0.022$ ) embryos, while ICM of the heterozygotes did not show significant difference from the wild-type ( $p = 0.14$ ) (Figure 3b). Additionally, the TE size was significantly reduced in the *dut*  $-/-$  embryos as compared to the heterozygous embryos ( $p = 0.015$ ) (Figure 3c). Comparison of



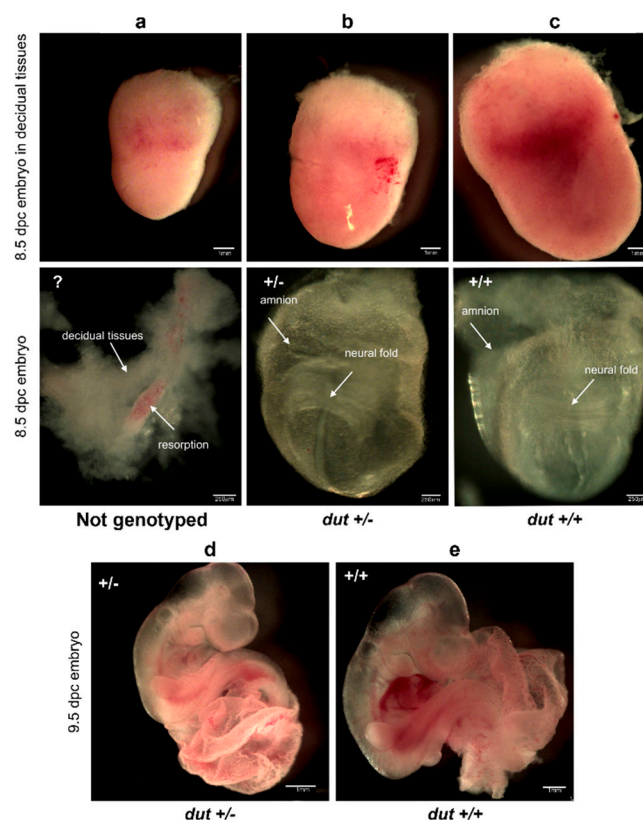
the wild-type and homozygous embryos showed a considerable reduction of the trophectoderm of the homozygous embryos ( $p = 0.060$ ), while the trophectoderm size of wild-type and heterozygous embryos did not differ significantly ( $p = 0.27$ ). These findings indicated that dUTPase deficiency impairs outgrowth of both ICM and trophectoderm cells.



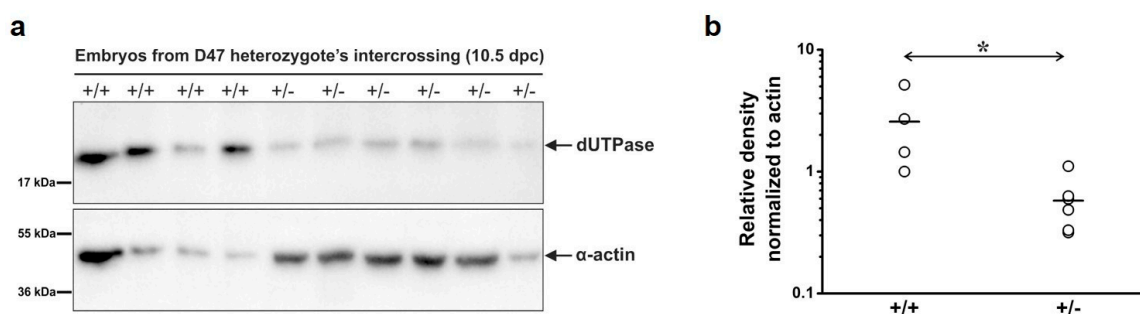
**Figure 3.** Outgrowth assay of pre-implantation embryos obtained by intercrossing D47 heterozygous mice. (a) Phase contrast images of D47 homozygous (−/−), heterozygous (+/−), and wild-type (+/+) blastocysts in in vitro culture. The first column shows embryos at 3.5 dpc after flushing from oviducts. White arrows indicate the zona pellucida surrounding the embryos. The second and third columns show the attached embryos, one day later focusing on the trophoblast cells or the inner cell mass (ICM) in the blastocoel. Scale bar represents 20 μm. The last column presents outgrowths after four days in culture. Scale bar represents 100 μm. Average size of ICM (b) and trophectoderm (TE) (c) was calculated for blastocysts of indicated genotypes. Error bars indicate standard deviation. n = 3 for (−/−), n = 5 for (+/−), and n = 2 for (+/+). Statistical analysis was done with a two-sample, single-tailed t-test assuming equal variance using Microsoft Excel. \*  $p < 0.05$ .

Next, embryos were isolated at 8.5 dpc or at 9.5 dpc for investigations at post-implantation embryonic stages (Figure 4). No obvious changes could be seen in the development of heterozygotes (Figure 4b,d) as compared to the wild-type (Figure 4c,e). We also observed resorbed embryos (Figure 4a, counted in Table 2) where the genotype could not be evaluated unequivocally since the mother's decidual tissue could not be separated from the resorptions. No *dut* −/− embryos were found at these or further stages. More images of wild-type, heterozygote, and the resorbed embryos at 8.5 dpc or at 9.5 dpc are provided in Figures S5 and S6.

To analyze dUTPase protein levels in the wild-type and heterozygous mouse, Western blot analysis was performed on the total protein extract from *dut* +/+ and +/− embryos (Figure 5). dUTPase levels were significantly reduced in the heterozygote as compared to the wild-type ( $p = 0.019$ ). This finding confirmed that the *dut* gene was successfully disrupted using the CRISPR/Cas9 genome editing system. *Dut* −/− homozygous embryo could not be found at this embryonic stage (at 10.5 dpc).



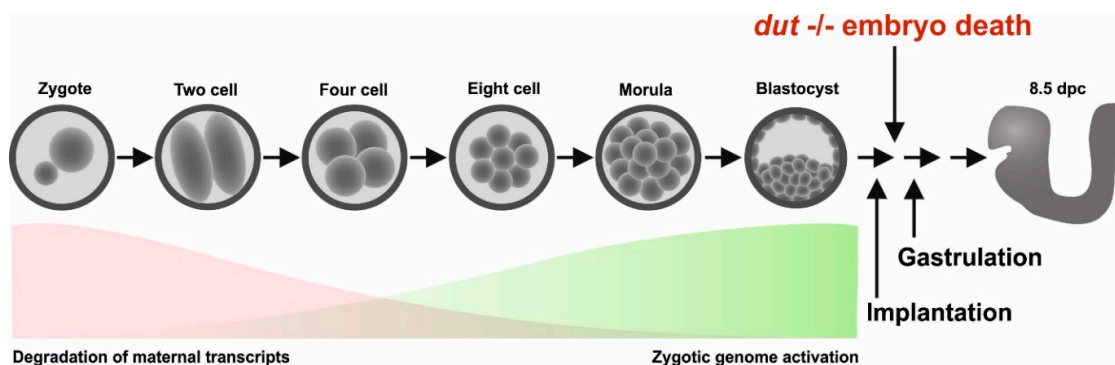
**Figure 4.** Images of embryos at 8.5 and 9.5 dpc obtained by crossing D47 heterozygous mice. (a) Representative image of a resorbed embryo at 8.5 dpc covered by decidua tissues. The resorbed embryo could not be genotyped as indicated with “?”. Heterozygous (+/–) (b) and wild-type (+/+) (c) embryos at 8.5 dpc are shown. Upper panels show embryos in intact decidua tissues. Scale bar represents 1 mm. Lower panels show the embryos dissected from decidua tissues. Arrows indicate the embryonic neural fold and the extra-embryonic amnion. Scale bar represents 250 µm. Heterozygous (+/–) (d) and wild-type (+/+) (e) embryos at 9.5 dpc are also shown. Scale bar represents 1 mm.



**Figure 5.** Protein level of dUTPase in embryos at 10.5 dpc from the intercrossing of D47 heterozygous mice. (a) Western blot displaying dUTPase protein level in wild-type (+/+) and heterozygous (+/–) mice. Membrane was developed against dUTPase (upper part) or α-actin (lower part) as a loading control. Blots are marked with black frame and separated with space. Uncut scans are shown in Figure S8. (b) Densitometric data for dUTPase levels from Western blot normalized for α-actin. Mean values are represented with horizontal lines. Every data point is shown, n = 4 for (+/+) and n = 6 for (+/–). Statistical analysis was carried out with a two-sided Mann–Whitney U test. \*  $p < 0.05$ .

#### 4. Discussion

The schematic pattern of embryonic development in mice is illustrated in Figure 6. In this scheme, the timeline of maternal mRNA and protein degradation and the parallel zygotic genome activation are also indicated [36,37]. Our results showed that the dUTPase knock-out did not affect the first several duplication cycles and viable blastocysts could be isolated. Furthermore, isolated *dut*  $-/-$  blastocysts could grow further in in vitro cultures, although outgrowth of both the inner cell mass and the trophectoderm cells were impaired. However, further development following implantation was prevented in the homozygous knock-out embryo.



**Figure 6.** Schematic diagram of embryonic development in mice. The red and green areas show the timeline of simultaneous degradation of maternal transcripts and activation of zygotic transcription. Arrows illustrate that homozygous knock-out embryos die shortly after implantation.

We hypothesize that the lethality of the dUTPase knock-out in mice is probably related to the lack of its enzymatic function in the *dut*  $-/-$  embryo. It is well known from the literature that the dUTPase enzyme is a key protein in genome integrity, and its deficiency in bacterial, yeast, and *Drosophila* models led to increased DNA damage frequency [7,20,21,24,25]. Elevated levels of ssDNA and double stranded breaks were observed in these models. The underlying mechanism of the increased DNA damage in lack of dUTPase might be attributed to the potential expansion of the cellular dUTP pool that leads to increased incorporation of uracil moieties into DNA. Numerous uracil moieties within the genomic DNA may induce a hyperactive base excision process through the uracil-DNA glycosylase enzymes [6]. Repair synthesis also occurs in the dUTP-enriched milieu therefore leading to further genomic uracil enrichment. These circumstances transform the repair process into a hyperactive futile cycle.

Our results also showed that at the early stages of embryonic development, until the blastocyst stage, embryos with the *dut*  $-/-$  genotype were still viable. Generally, maternal storages are already depleted at the blastocyst stage, however, the exact situation is not yet characterized for maternal dUTPase mRNA, and protein stores. We therefore conclude that viability of the early embryonal stages might be due to various reasons: i) Maternal source of dUTPase may be sufficiently present in these embryonal stages, or ii) the repair mechanism relying on uracil-DNA glycosylase may not yet be effective in these stages. Unfortunately, these aspects are also not yet studied in the literature, necessitating further detailed molecular investigations.

We conclude that mitotic events may proceed in lack of dUTPase in mice, however, the enzyme is indispensable for later differentiation stages. Our model will be useful in later detailed studies to outline the molecular events leading to the observed early embryonic lethal phenotype. In addition, combined multiple knock-outs of *dut* and other relevant genes are also expected to provide important insights into developmental processes.



**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2218-273X/9/4/136/s1>, Figure S1: Sequencing analysis of *dut* gene from founder mouse #2 and #4, Figure S2: Alignment of the sequencing results from wild-type and mouse #4 animals, Figure S3: Phase contrast images of embryos obtained by crossing D47 heterozygous mice, Figure S4: Outgrowth assay of preimplantation embryos obtained by intercrossing D47 heterozygous mice, Figure S5: Images of embryos at 8.5 dpc obtained by crossing D47 heterozygous mice, Figure S6: Images of embryos at 9.5 dpc obtained by crossing D47 heterozygous mice, Figure S7: Full-length agarose gel of Figure 2b, Figure S8: Full-length membrane of Figure 5a, Table S1: Oligonucleotides used in this study.

**Author Contributions:** H.L.P., E.G., L.H., and B.G.V. conceived the experiments, H.L.P., G.A.R., Z.G., O.I.H., G.T., E.G., and L.H. conducted the experiments, H.L.P., G.A.R., Z.G., G.T., G.R., E.G., L.H., and B.G.V. analyzed the results. All authors contributed to manuscript writing.

**Funding:** This research was funded by the National Research, Development and Innovation Office of Hungary (K109486, K119493, NVKP\_16-1-2016-0020, 2017-1.3.1-VKE-2017-00002, 2017-1.3.1-VKE-2017-00013, VEKOP-2.3.2-16-2017-00013, NKP-2018-1.2.1-NKP-2018-00005 to B.G.V.) and the BME-Biotechnology FIKP grant of EMMI (BME FIKP-BIO). H.L.P. was also supported by the UNKP-17-3 New National Excellence Program of the Ministry of Human Capacities. This work was supported by GÉNNET\_21 (VEKOP-2.3.2-16-2016-00012 to E.G.) and OTKA PD 111964 to O.I.H.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

- Mjelle, R.; Hegre, S.A.; Aas, P.A.; Slupphaug, G.; Drabløs, F.; Sætrom, P.; Krokan, H.E. Cell cycle regulation of human DNA repair and chromatin remodeling genes. *DNA Repair (Amst)*. **2015**, *30*, 53–67. [\[CrossRef\]](#)
- Galperin, M.Y.; Moroz, O.V.; Wilson, K.S.; Murzin, A.G. House cleaning, a part of good housekeeping. *Mol. Microbiol.* **2006**, *59*, 5–19. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nakabeppu, Y.; Ohta, E.; Abolhassani, N. MTH1 as a nucleotide pool sanitizing enzyme: Friend or foe? *Free Radic. Biol. Med.* **2017**, *107*, 151–158. [\[CrossRef\]](#) [\[PubMed\]](#)
- Requena, C.E.; Pérez-Moreno, G.; Ruiz-Pérez, L.M.; Vidal, A.E.; González-Pacanowska, D. The NTP pyrophosphatase DCTPP1 contributes to the homeostasis and cleansing of the dNTP pool in human cells. *Biochem. J.* **2014**, *459*, 171–180. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nagy, G.N.; Leveles, I.; Vértessy, B.G. Preventive DNA repair by sanitizing the cellular (deoxy)nucleoside triphosphate pool. *FEBS J.* **2014**, *281*, 4207–4223. [\[CrossRef\]](#) [\[PubMed\]](#)
- 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. [\[CrossRef\]](#)
- Hirmondo, R.; Lopata, A.; Suranyi, E.V.; Vértessy, B.G.; Toth, J. Differential control of dNTP biosynthesis and genome integrity maintenance by the dUTPase superfamily enzymes. *Sci. Rep.* **2017**, *7*, 6043. [\[CrossRef\]](#)
- Hizi, A.; Herzig, E. dUTPase: The frequently overlooked enzyme encoded by many retroviruses. *Retrovirology* **2015**, *12*, 70. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gaidamaviciute, E.; Tauraitė, D.; Gagilas, J.; Lagunavicius, A. Site-directed chemical modification of archaeal *Thermococcus litoralis* Sh1B DNA polymerase: Acquired ability to read through template-strand uracils. *Biochim. Biophys. Acta-Proteins Proteomics* **2010**, *1804*, 1385–1393. [\[CrossRef\]](#)
- Hemsworth, G.R.; González-Pacanowska, D.; Wilson, K.S. On the catalytic mechanism of dimeric dUTPases. *Biochem. J.* **2013**, *456*, 81–88. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hemsworth, G.R.; Moroz, O.V.; Fogg, M.J.; Scott, B.; Bosch-Navarrete, C.; González-Pacanowska, D.; Wilson, K.S. The crystal structure of the *Leishmania* major deoxyuridine triphosphate nucleotidohydrolase in complex with nucleotide analogues, dUMP, and deoxyuridine. *J. Biol. Chem.* **2011**, *286*, 16470–16481. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kerepesi, C.; Szabó, J.E.; Papp-Kádár, V.; Dobay, O.; Szabó, D.; Grolmusz, V.; Vértessy, B.G. Life without dUTPase. *Front. Microbiol.* **2016**, *7*, 1768. [\[CrossRef\]](#) [\[PubMed\]](#)
- Benedek, A.; Horváth, A.; Hirmondó, R.; Ozohanics, O.; Békési, A.; Módos, K.; Révész, Á.; Vékey, K.; Nagy, G.N.; Vértessy, B.G. Potential steps in the evolution of a fused trimeric all- $\beta$  dUTPase involve a catalytically competent fused dimeric intermediate. *FEBS J.* **2016**, *283*, 3268–3286. [\[CrossRef\]](#)

14. Varga, B.; Barabás, O.; Takács, E.; Nagy, N.; Nagy, P.; Vértessy, B.G. Active site of mycobacterial dUTPase: Structural characteristics and a built-in sensor. *Biochem. Biophys. Res. Commun.* **2008**, *373*, 8–13. [[CrossRef](#)] [[PubMed](#)]
15. Chan, S.; Segelke, B.; Lekin, T.; Krupka, H.; Cho, U.S.; Kim, M.; So, M.; Kim, C.-Y.; Naranjo, C.M.; Rogers, Y.C.; et al. Crystal structure of the Mycobacterium tuberculosis dUTPase: Insights into the catalytic mechanism. *J. Mol. Biol.* **2004**, *341*, 503–517. [[CrossRef](#)] [[PubMed](#)]
16. Mol, C.D.; Harris, J.M.; McIntosh, E.M.; Tainer, J.A. Human dUTP pyrophosphatase: Uracil recognition by a  $\beta$  hairpin and active sites formed by three separate subunits. *Structure* **1996**, *4*, 1077–1092. [[CrossRef](#)]
17. Williams, M.V.; Cox, B.; Ariza, M.E. Herpesviruses dUTPases: A New Family of Pathogen-Associated Molecular Pattern (PAMP) Proteins with Implications for Human Disease. *Pathogens* **2017**, *6*, 2. [[CrossRef](#)]
18. Bergman, A.-C.; Nyman, P.O.; Larsson, G. Kinetic properties and stereospecificity of the monomeric dUTPase from herpes simplex virus type 1. *FEBS Lett.* **1998**, *441*, 327–330. [[CrossRef](#)]
19. Tarbouriech, N.; Buisson, M.; Seigneurin, J.-M.; Cusack, S.; Burmeister, W.P. The monomeric dUTPase from Epstein-Barr virus mimics trimeric dUTPases. *Structure* **2005**, *13*, 1299–1310. [[CrossRef](#)]
20. El-Hajj, H.H.; Zhang, H.; Weiss, B. Lethality of a dut (deoxyuridine triphosphatase) mutation in Escherichia coli. *J. Bacteriol.* **1988**, *170*, 1069–1075. [[CrossRef](#)]
21. Gadsden, M.H.; McIntosh, E.M.; Game, J.C.; Wilson, P.J.; Haynes, R.H. dUTP pyrophosphatase is an essential enzyme in Saccharomyces cerevisiae. *EMBO J.* **1993**, *12*, 4425–4431. [[CrossRef](#)] [[PubMed](#)]
22. Dengg, M.; Garcia-Muse, T.; Gill, S.G.; Ashcroft, N.; Boulton, S.J.; Nilsen, H. Abrogation of the CLK-2 checkpoint leads to tolerance to base-excision repair intermediates. *EMBO Rep.* **2006**, *7*, 1046–1051. [[CrossRef](#)] [[PubMed](#)]
23. Alam, M.S.; Moriyama, H.; Matsumoto, M. Inhibition of Dr-dut gene causes DNA damage in planarian. *Mol. Reprod. Dev.* **2018**, *85*, 188–196. [[CrossRef](#)] [[PubMed](#)]
24. 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](#)] [[PubMed](#)]
25. Békési, A.; Zagyva, I.; Hunyadi-Gulyás, É.; Pongrácz, V.; Kovári, J.; Nagy, Á.O.; Erdei, A.; Medzihradsky, K.F.; Vértessy, B.G. Developmental regulation of dUTPase in Drosophila melanogaster. *J. Biol. Chem.* **2004**, *279*, 22362–22370. [[CrossRef](#)] [[PubMed](#)]
26. Dubois, E.; Córdoba-Cañero, D.; Massot, S.; Siaud, N.; Gakière, B.; Domenichini, S.; Guérard, F.; Roldan-Arjona, T.; Doutriaux, M.P. Homologous recombination is stimulated by a decrease in dUTPase in Arabidopsis. *PLoS ONE* **2011**, *6*, e18658. [[CrossRef](#)] [[PubMed](#)]
27. 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*, 620–626.
28. Studebaker, A.W.; Lafuse, W.P.; Kloesel, R.; Williams, M.V. Modulation of human dUTPase using small interfering RNA. *Biochem. Biophys. Res. Commun.* **2005**, *327*, 306–310. [[CrossRef](#)]
29. 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*, 369–390. [[CrossRef](#)]
30. 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](#)]
31. 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*, 1683–1693. [[CrossRef](#)] [[PubMed](#)]
32. Saito, K.; Nagashima, H.; Noguchi, K.; Yoshisue, K.; Yokogawa, T.; Matsushima, E.; Tahara, T.; Takagi, S. First-in-human, phase I dose-escalation study of single and multiple doses of a first-in-class enhancer of fluoropyrimidines, a dUTPase inhibitor (TAS-114) in healthy male volunteers. *Cancer Chemother. Pharmacol.* **2014**, *73*, 577–583. [[CrossRef](#)] [[PubMed](#)]
33. Behringer, R.; Gertsenstein, M.; Nagy, K.V.; Nagy, A. *Manipulating the Mouse Embryo: A Laboratory Manual*, 4th ed.; Cold Spring Harbor Laboratory Press: New York, NY, USA, 2014.
34. Stemmer, M.; Thumberger, T.; Del Sol Keyer, M.; Wittbrodt, J.; Mateo, J.L. CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS One* **2015**, *10*, e0124633. [[CrossRef](#)] [[PubMed](#)]

35. Hsu, P.D.; Scott, D.A.; Weinstein, J.A.; Ran, F.A.; Konermann, S.; Agarwala, V.; Li, Y.; Fine, E.J.; Wu, X.; Shalem, O.; et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **2014**, *31*, 827–832. [[CrossRef](#)] [[PubMed](#)]
36. Li, L.; Zheng, P.; Dean, J. Maternal control of early mouse development. *Development* **2010**, *137*, 859–870. [[CrossRef](#)] [[PubMed](#)]
37. Tadros, W.; Lipshitz, H.D. The maternal-to-zygotic transition: A play in two acts. *Development* **2009**, *136*, 3033–3042. [[CrossRef](#)] [[PubMed](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).

# Supplementary Materials

## CRISPR/Cas9-Mediated Knock-Out of dUTPase in Mice Leads to Early Embryonic Lethality

Hajnalka Laura Pálincás <sup>1,2,3,\*</sup>, Gergely Attila Rácz <sup>1,3</sup>, Zoltán Gál <sup>4</sup>, Orsolya Ivett Hoffmann <sup>4</sup>, Gergely Tihanyi <sup>1,3</sup>, Gergely Róna <sup>3,5,6</sup>, Elen Góczy <sup>4,\*</sup>, László Hiripi <sup>4,\*</sup>, Beáta G. Vértessy <sup>1,3\*</sup>

<sup>1</sup> Institute of Enzymology, RCNS, Hungarian Academy of Sciences, H-1117 Budapest, Hungary; racz.gergely@ttk.mta.hu (G.A.R); tihanyi@stud.uni-heidelberg.de (G.T.)

<sup>2</sup> Doctoral School of Multidisciplinary Medical Science, University of Szeged, H-6720 Szeged, Hungary

<sup>3</sup> Department of Applied Biotechnology and Food Sciences, Budapest University of Technology and Economics, H-1111 Budapest, Hungary ; Gergely.Rona@nyulangone.org (G.R.)

<sup>4</sup> Department of Animal Biotechnology, Agricultural Biotechnology Institute, National Agricultural Research and Innovation Centre, H-2100 Gödöllő, Hungary ; zoltan.gal89@gmail.com (Z.G.); hoffmannorsi@gmail.com (O.I.H.)

<sup>5</sup> Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY10016, United States

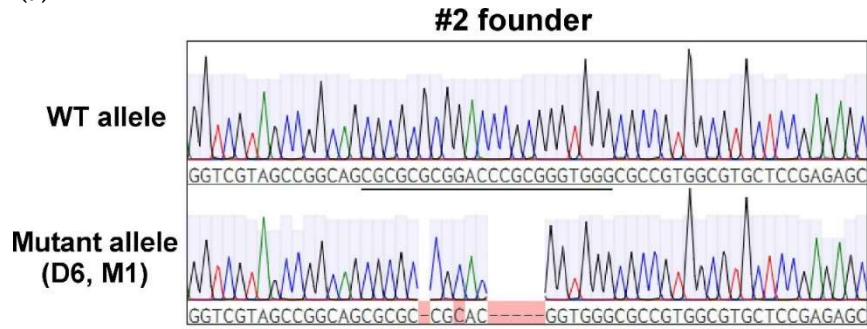
<sup>6</sup> Perlmutter Cancer Center, New York University School of Medicine, New York, NY10016, United States

\* Correspondence: vertessy@mail.bme.hu (B.G.V); palinkas.hajnalka@ttk.mta.hu (H.L.P.); gocza.elen@abc.naik.hu (E.G.); hiripi1@gmail.com (L.H.)

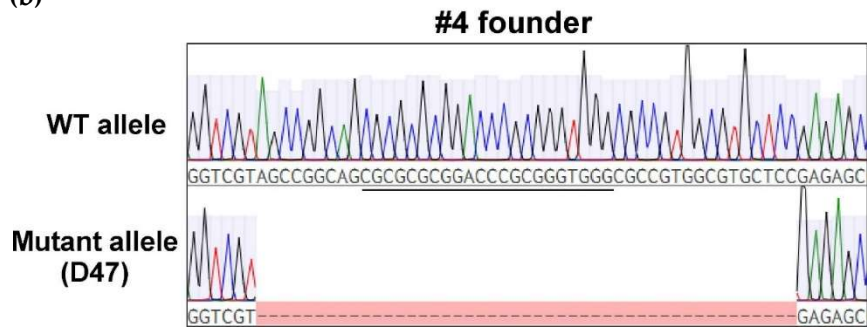
Received: date; Accepted: date; Published: date

**Keywords:** dUTPase; CRISPR/Cas9-mediated knock-out; blastocyst outgrowth; embryonic development

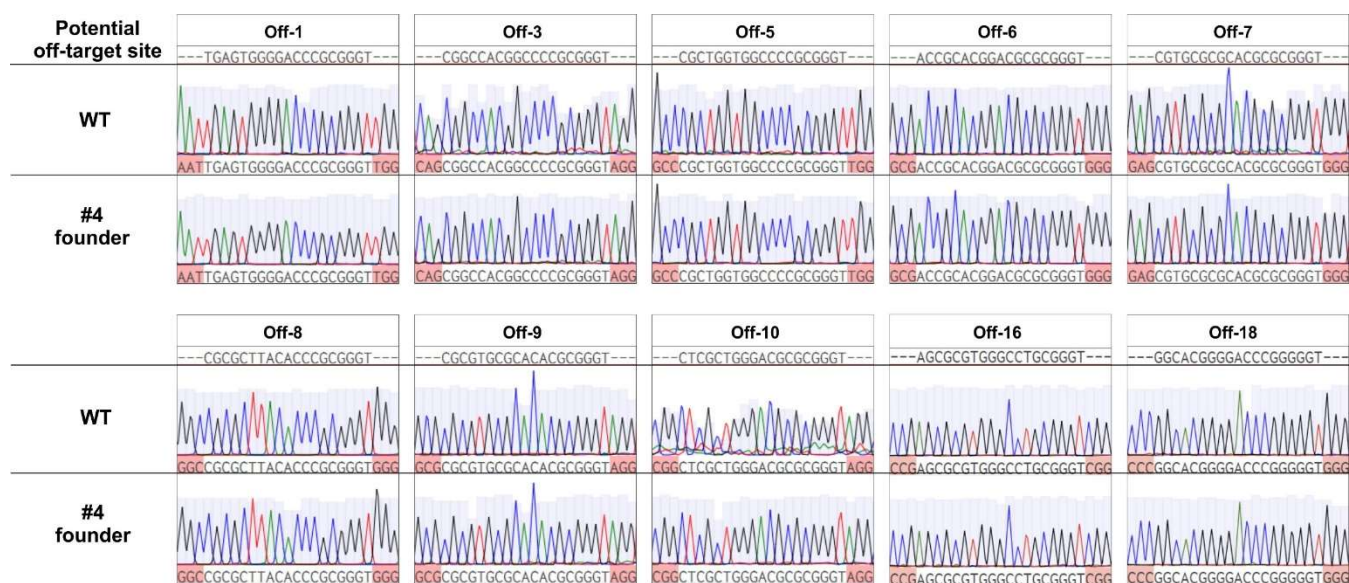
(a)



(b)

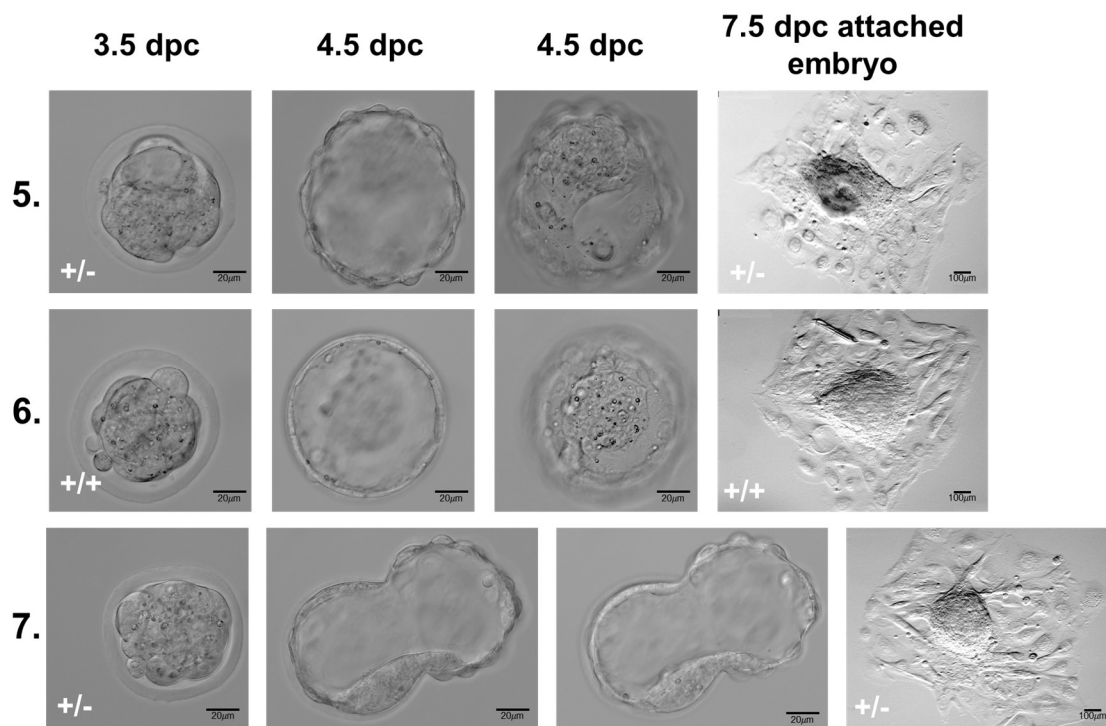
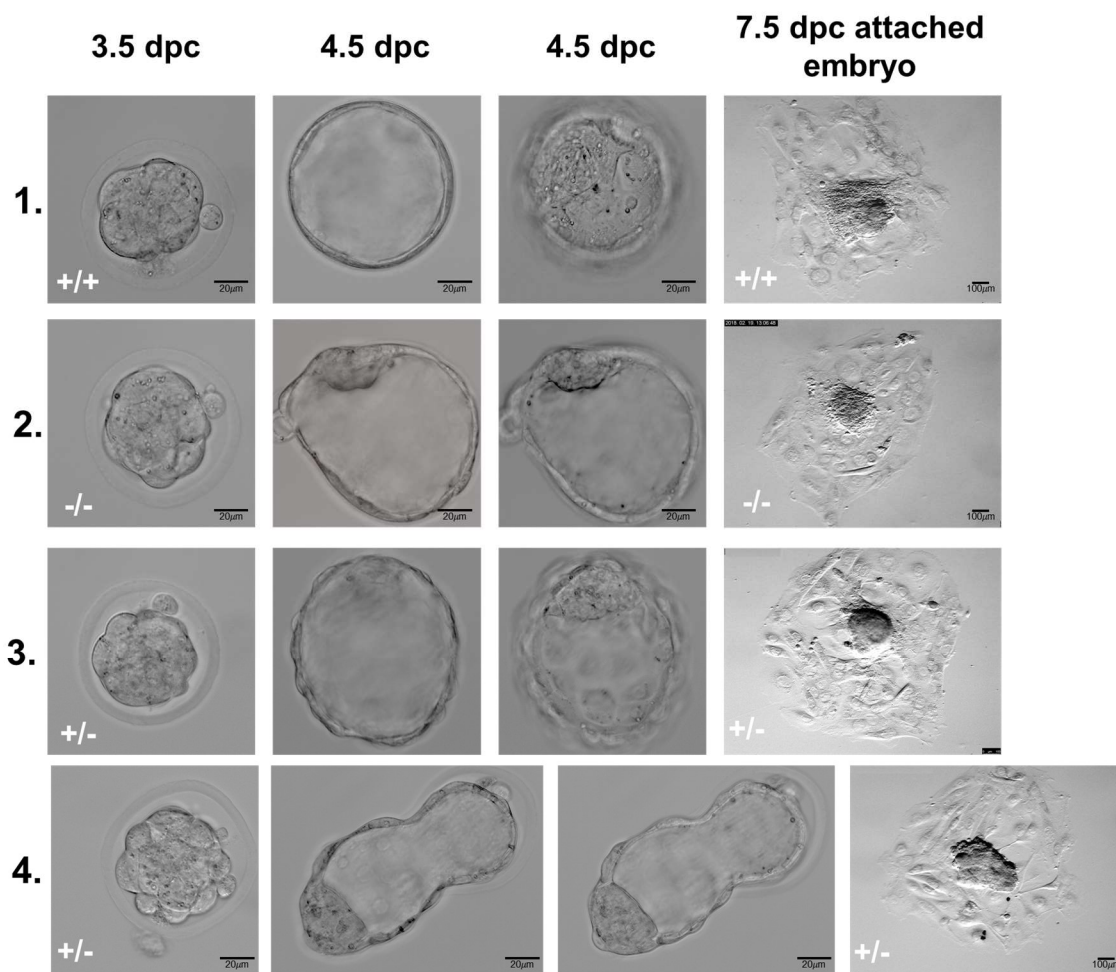


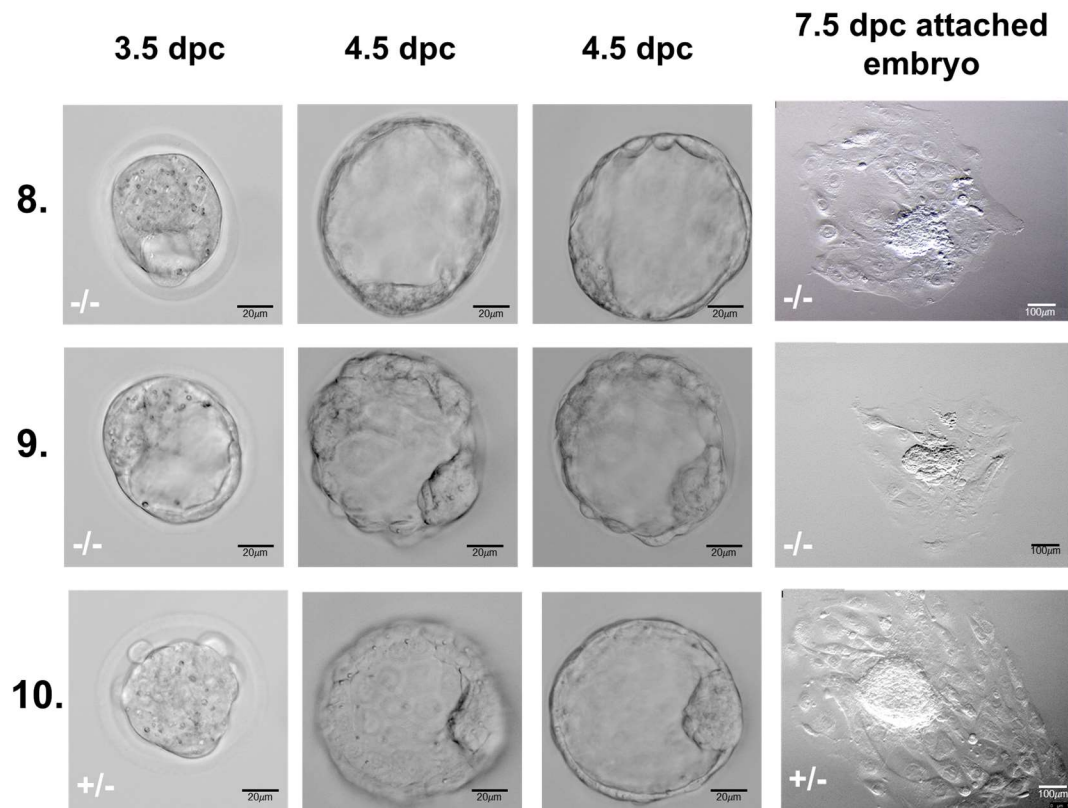
**Figure S1. Sequencing analysis of *dut* gene from founder mouse #2 and #4.** Sequencing results of *dut* gene from the mouse #2 (a) show 6 bp deletion and 1 bp substitution (D6, M1), while from the mouse #4 (b) show 47 bp deletion (D47). CRISPR target site including PAM sequence is underlined. CRISPR induced modifications are marked with red highlight.



**Figure S2. Alignment of the sequencing results from wild type and mouse #4 animals.** Sequences are fully identical at these potential off-target sites identified by the CRISPR/Cas9 target online predictor software CCTop.



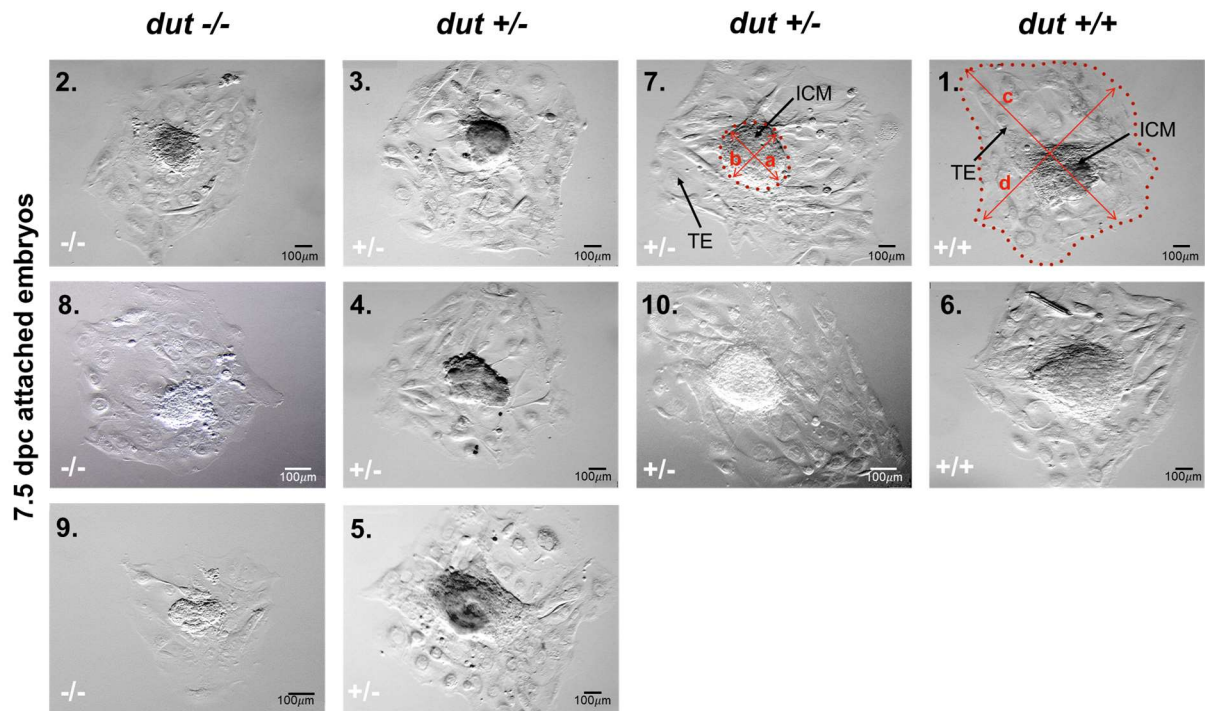




**Figure S3. Phase contrast images of embryos obtained by crossing D47 heterozygous mice.** Homozygous (*dut*<sup>-/-</sup>: 2, 8, 9), heterozygous embryos (*dut*<sup>+/-</sup>: 3, 4, 5, 7, 10), and wild type blastocysts (*dut*<sup>+/+</sup>: 1, 6) in *in vitro* culture. The first column shows embryos at 3.5 dpc after flushing from oviducts. The second and third columns show the embryos one day later focusing on the trophoblast cells or the inner cell mass (ICM). Scale bar, 20 μm. The last columns present outgrowths after 4 days in culture (7.5 dpc). Scale bar, 100 μm.



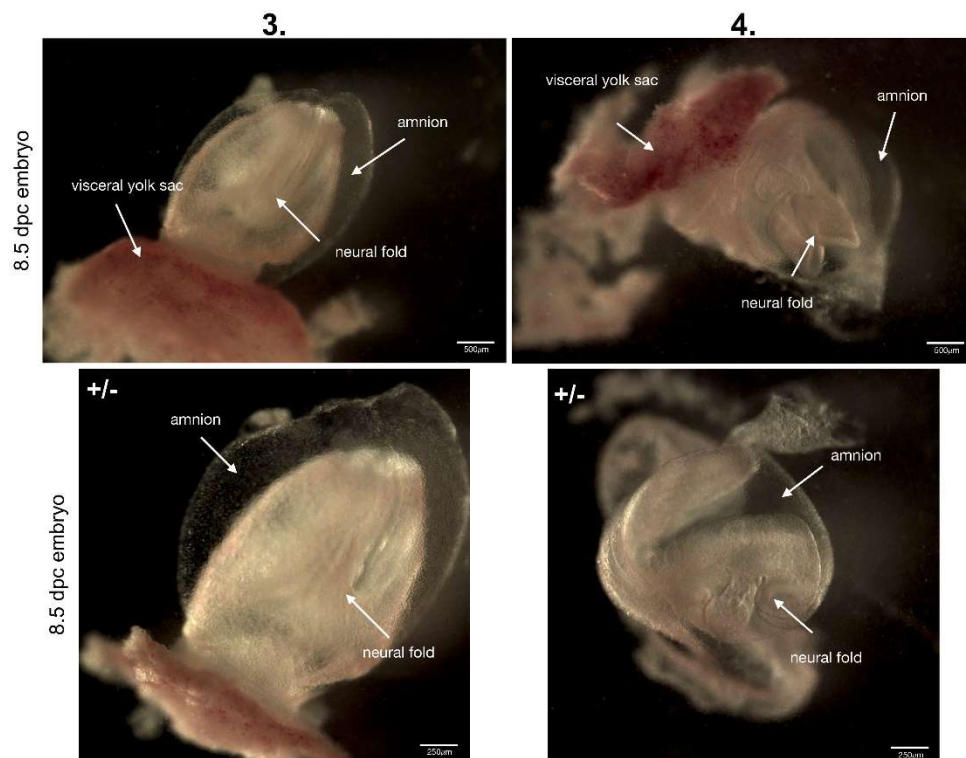
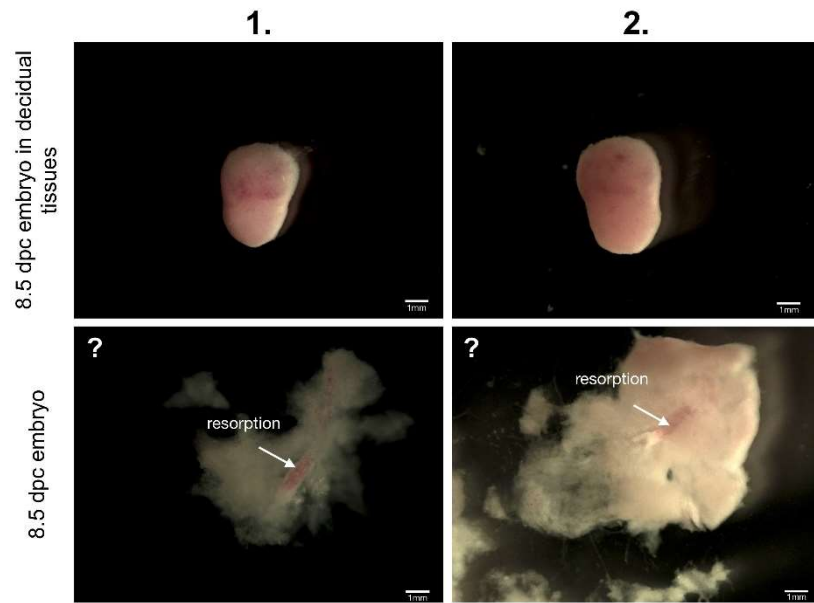
(a)



(b)

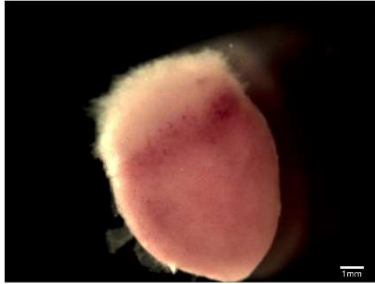
genotype	number of blastocysts	ICM size (a*b) mm <sup>2</sup>		TE size (c*d) mm <sup>2</sup>	
		average	SD( $\pm$ )	average	SD( $\pm$ )
<i>dut -/-</i>	3	0.078	0.006	0.964	0.156
<i>dut +/-</i>	5	0.138	0.036	1.425	0.177
<i>dut +/+</i>	2	0.189	0.044	1.314	0.055

**Figure S4. Outgrowth assay of preimplantation embryos obtained by intercrossing D47 heterozygous mice.** (a) The images represent homozygous (*dut -/-*: 2, 8, 9); heterozygous (*dut +/-*: 3, 4, 5, 7, 10); wild type (*dut +/+*: 1, 6) blastocysts attached to the surface of tissue culture dishes and their outgrowth in *in vitro* cultures for 4 days (7.5 dpc). Scale bar, 100  $\mu$ m. Red arrows (a, b, c, d, respectively) show the parameters used for outgrowth size measurements. (b) Table contains the result of the blastocysts outgrowth size measurements: average size of inner cell mass (ICM) and trophoblastic ectoderm (TE). SD indicates the standard deviation. n = 3 for (*dut -/-*), n = 5 for (*dut +/-*), and n = 2 for (*dut +/+*).

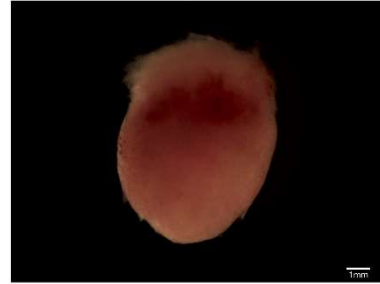


5.

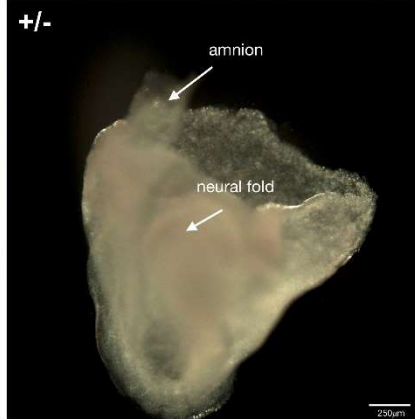
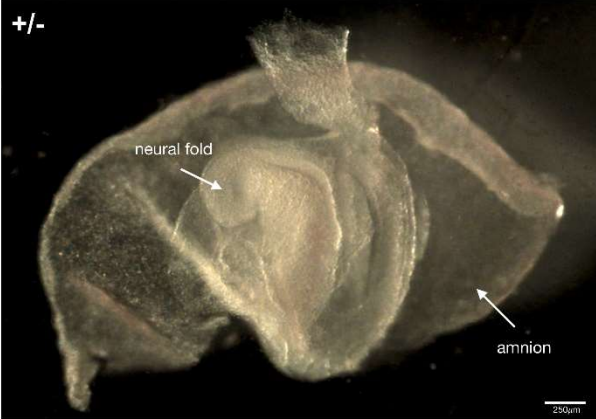
8.5 dpc embryo in decidual tissues



6.

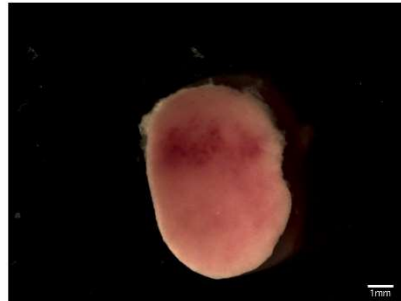


8.5 dpc embryo

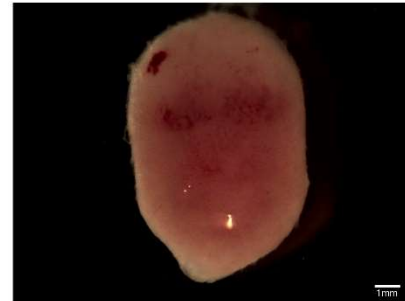


7.

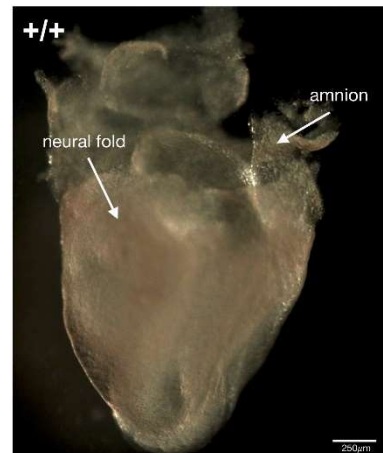
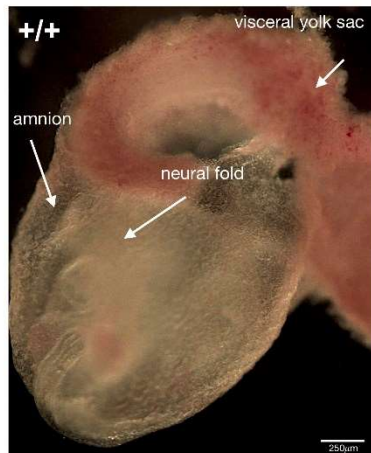
8.5 dpc embryo in decidual tissues

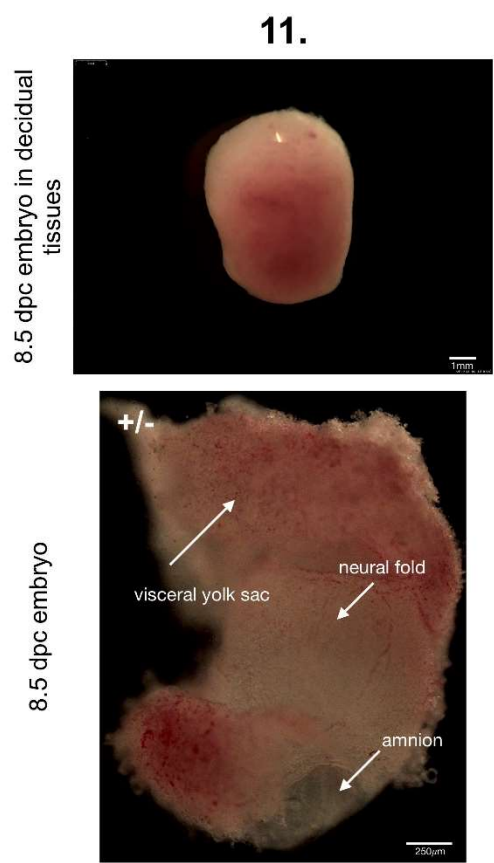
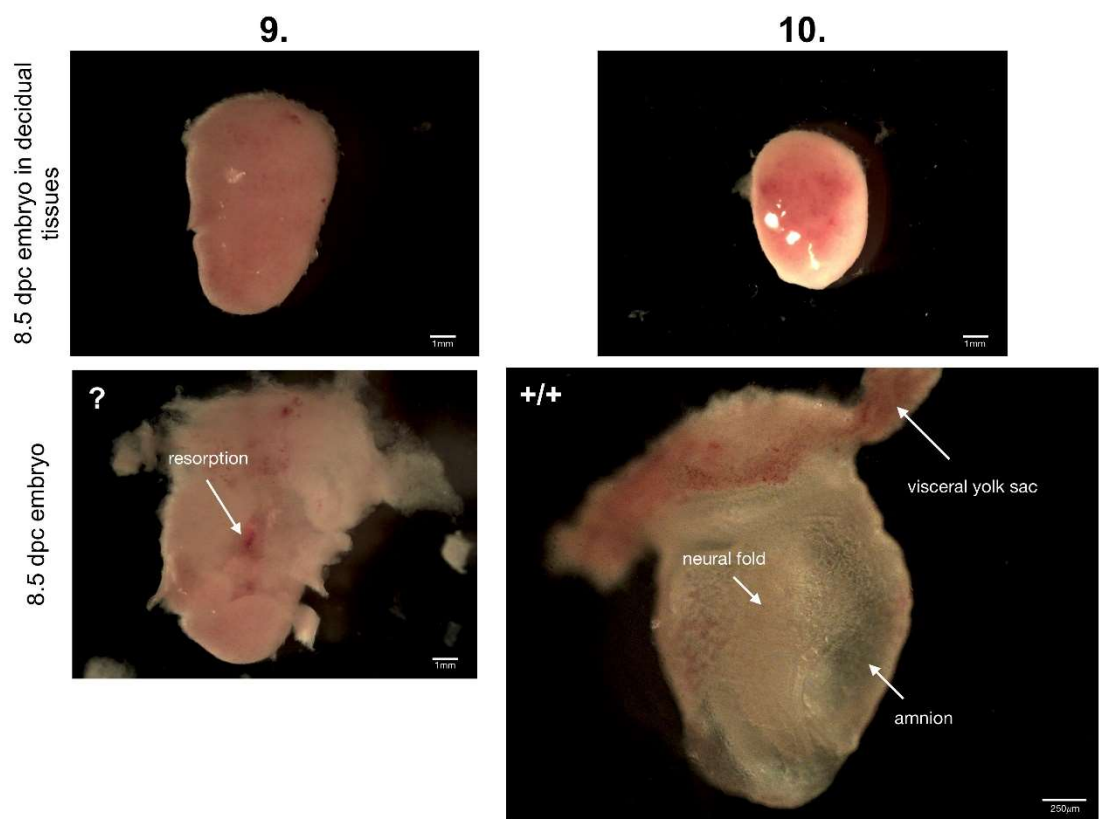


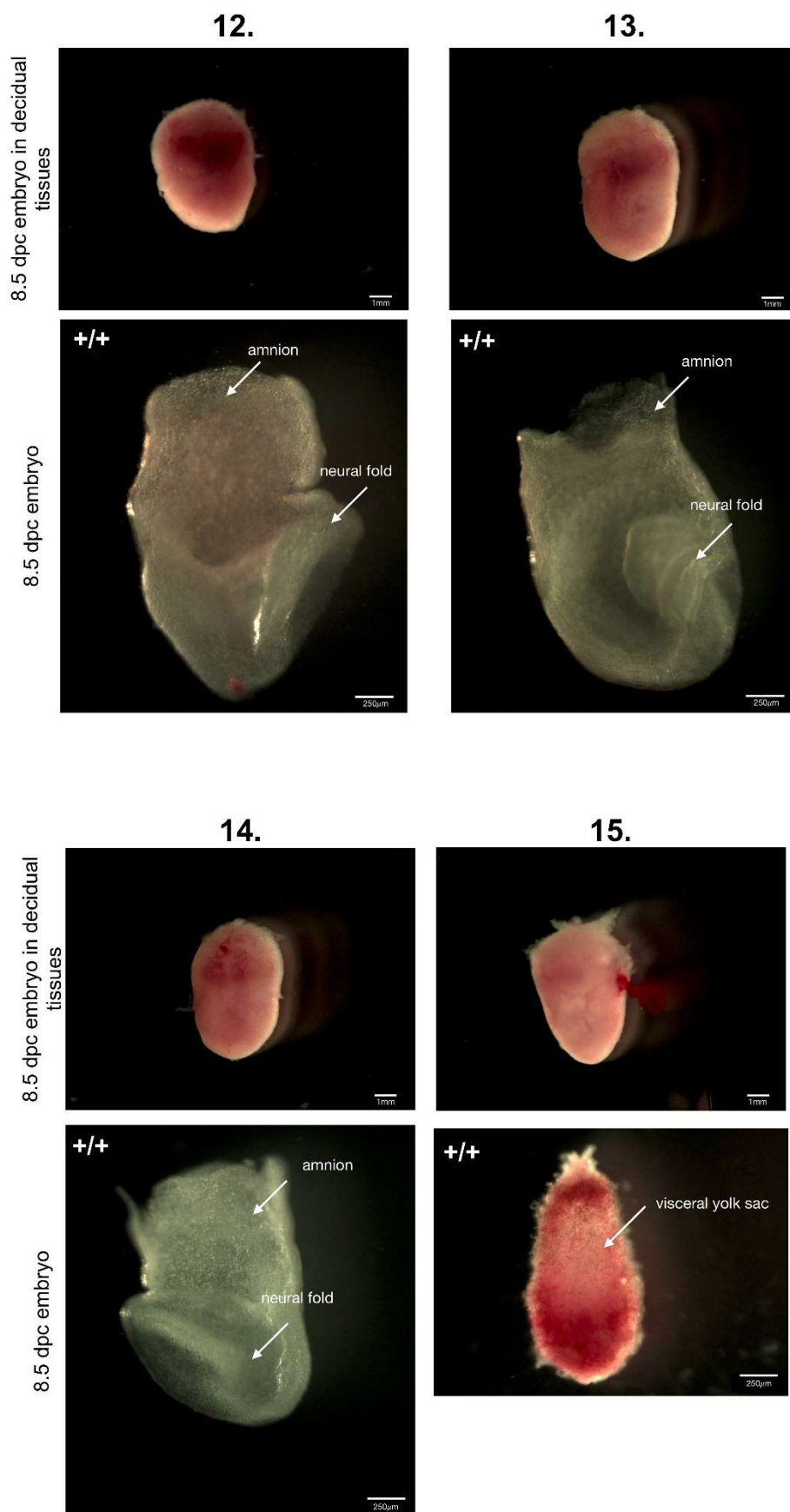
8.



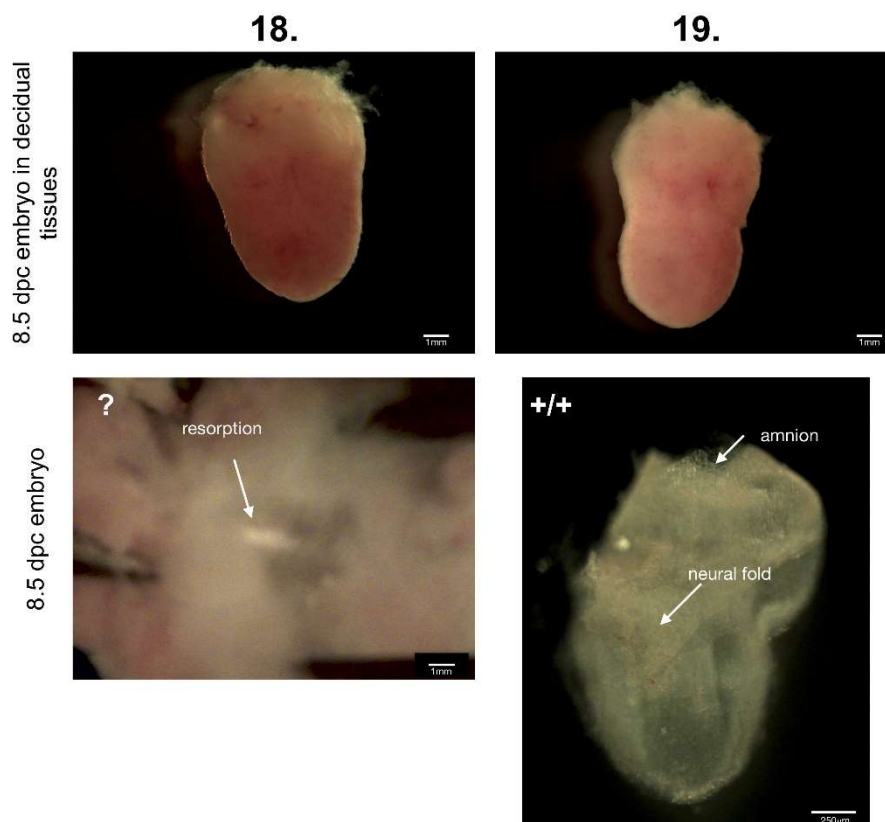
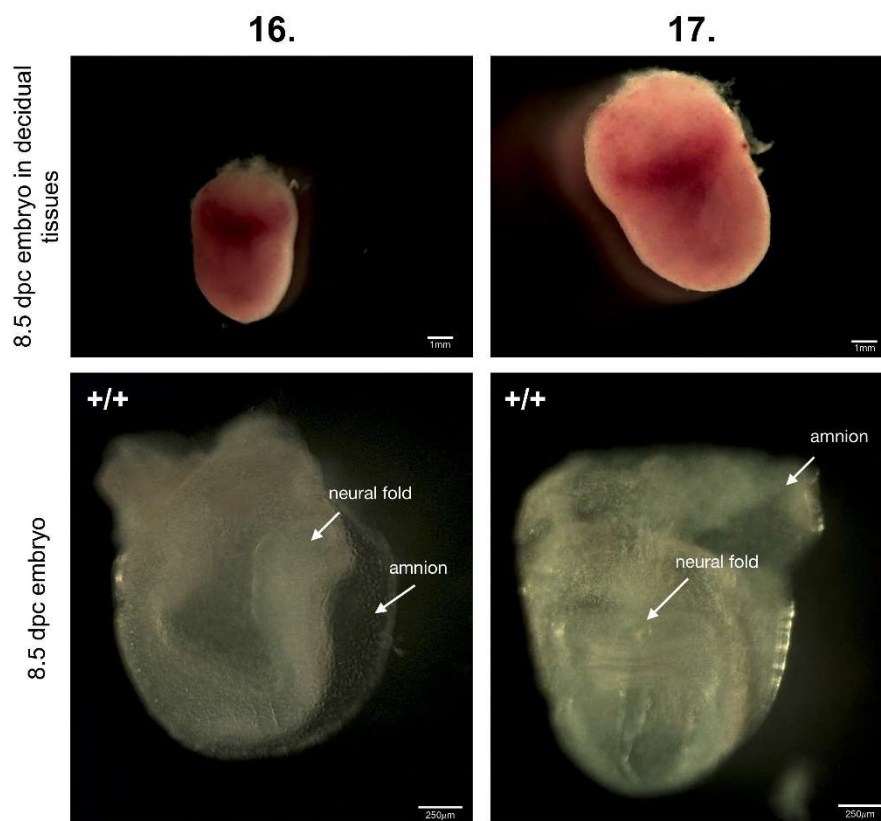
8.5 dpc embryo

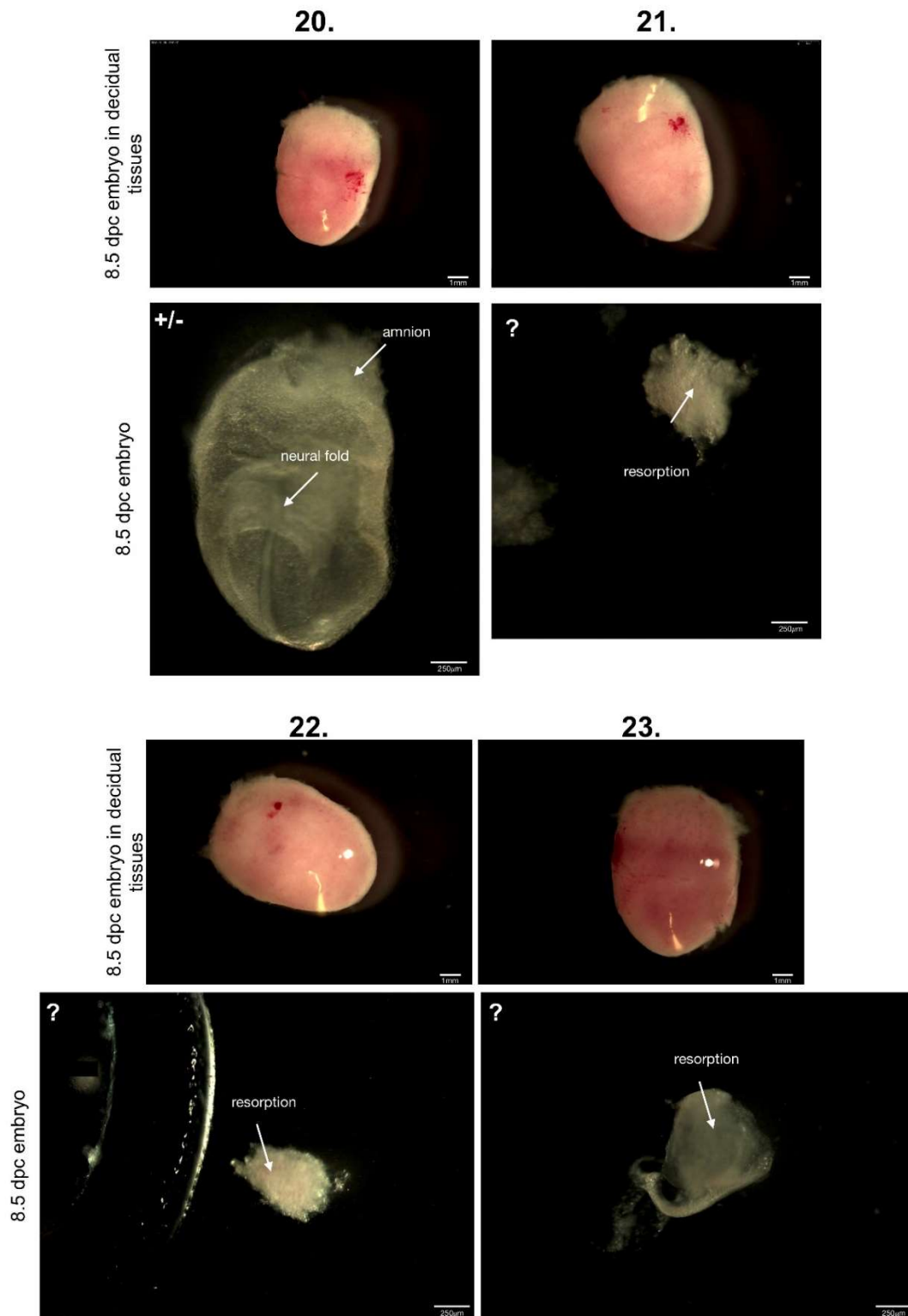




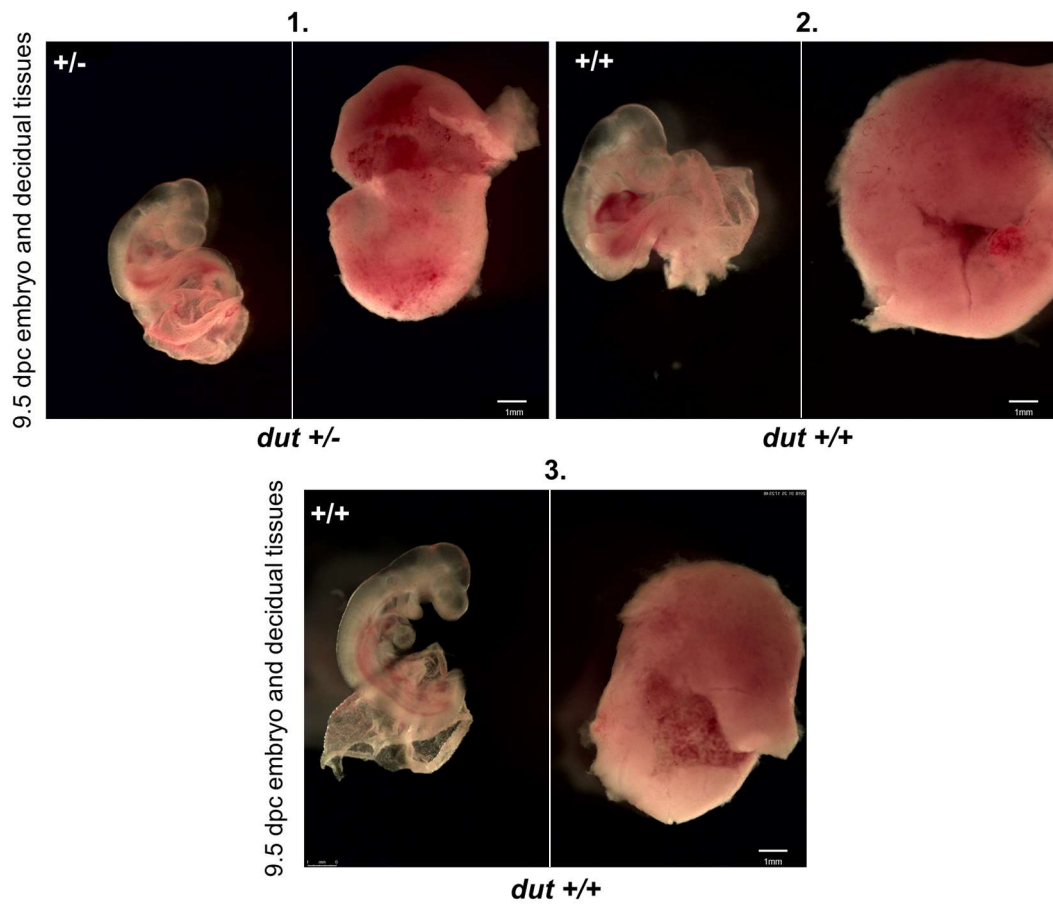








**Figure S5. Images of embryos at 8.5 dpc obtained by crossing D47 heterozygous mice.** Heterozygous embryos (*dut* +/-: 3, 4, 5, 6, 11, 20) and wild type embryos (*dut* +/+: 7, 8, 10, 12, 13, 14, 15, 16, 17, 19). The resorbed embryos (1, 2, 9, 18, 21, 22, 23) could not be genotyped (indicated with „?“). Upper panels show embryos in the intact decidual tissues. Scale bar, 1 mm. Lower panels present the embryos dissected from decidual tissues. Arrows represent the embryonic regions (neural fold) and extra-embryonic tissues (amnion, yolk sac) of embryos. Scale bar, 250 µm.



**Figure S6. Images of embryos at 9.5 dpc obtained by crossing D47 heterozygous mice.** Heterozygous embryos (*dut +/-*: 1) and wild type embryos (*dut +/+*: 2, 3). Right panels show embryos in decidua tissues. Left panels present the embryos dissected from decidua tissues. Scale bar, 1 mm.



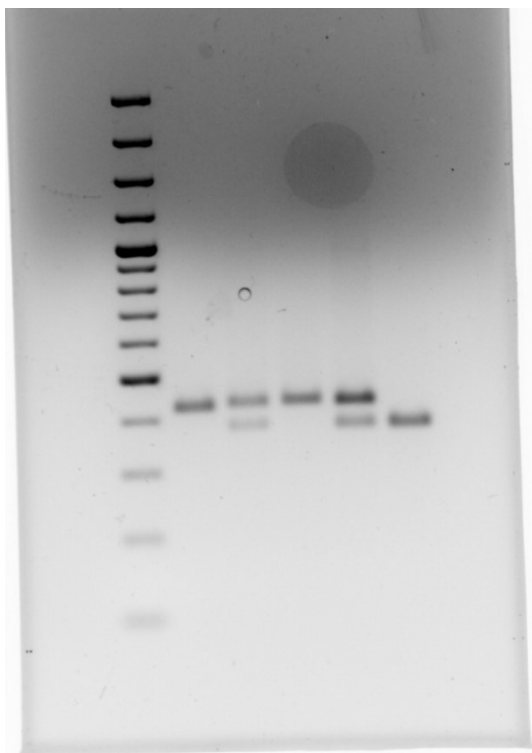


Figure S7. Full-length agarose gel of Figure 2b.

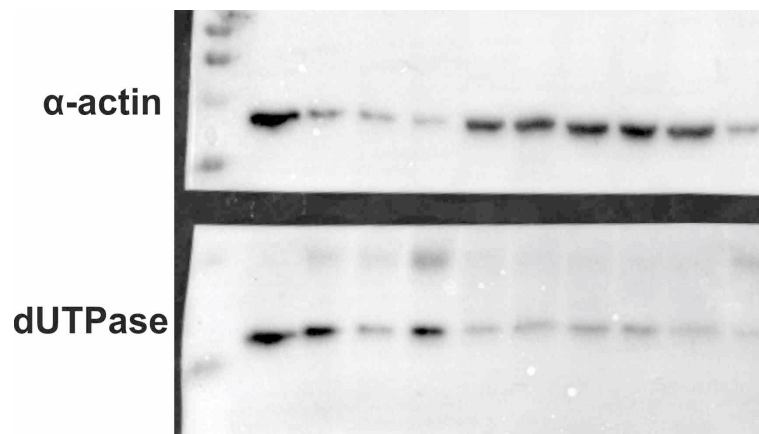


Figure S8. Full-length membrane of Figure 5a.

**Table S1. Oligonucleotides used in this study.** Restriction sites are underlined.

Sequence (5' - 3')	Name	Experiment
CTGGTCTGAGCAGGTAACGG	Cel-1-F	CEL 1 assay
TAACGTGCACACTGCAGACA	Cel-1-R	
GGTCGGTGCCTCCTCTAG	Dut-gen-F	Genotyping
AATAAGCCTTGACATCCGG	Dut-gen-R	
GGAGATTTTCGGCGGGTAGG	Dut-nest-F	
TCCGGTCGACCTGGTCTGAGCAGGTAACGG	pBS-F	Cloning
CTAGGAATTCCTAACGTGCACACTGCAGACA	pBS-R	
AGCTTTTACTCATTGCTACTCTCT	Off-1-F	Off-target analysis
ATTGTGAAGTGTTGGCTTATGGG	Off-1-R	
CCCCGTTTCACAGAACCGTA	Off-3-F	
CCAAGGACCTAGCCGACATC	Off-3-R	
AACAAGAGCTAGAGGCAGCG	Off-5-F	
CAACAGCAGCTGGGTGTCA	Off-5-R	
GGTGCGTGTTCGCCAGTG	Off-6-F	
GAGAGAAGCAAGCAGATGCG	Off-6-R	
GTCACTATGCAAGAAGTTTCCTTAT	Off-7-F	
CCTACCACGTGCCCTGTAAG	Off-7-R	
CACCTCAGCAGTACCAAACAAAG	Off-8-F	
TTGATGGTAAATTCTCCTCTCTTCA	Off-8-R	
GCAGAATCGAAGGTGGTGGT	Off-9-F	
CAGGGCATAGGATACCTGCT	Off-9-R	
CAAAGGGAATGCCACTTCTCG	Off-10-F	
CGCGCGCACTACTGGAG	Off-10-R	
CGAATGGGATGGTAGCAAGT	Off-16-F	
TTTTGCTGTCAGTCCGTGTT	Off-16-R	
CTTCCTCGCCCTCATCGC	Off-18-F	
GAATGAGCGTTCAGCAGTGG	Off-18-R	