Antonio Sarno

Uracil in DNA - necessary intermediate for antibody maturation and mutagenic lesion

Thesis for the degree of Philosophiae Doctor

Trondheim, June 2015

Norwegian University of Science and Technology
Faculty of Medicine
Department of Cancer Research and Molecular Medicine (IKM)

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Medicine
Department of Cancer Research and Molecular Medicine (IKM)

978-82-326-0962-8 (print)
978-82-326-0963-5 (digital)
1503-8181

Doctoral theses at NTNU, 2015:153

Printed by NTNU Grafisk senter

NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET

DET MEDISINSKE FAKULTET

Uracil in DNA - et viktig mellomprodukt for adaptiv immunitet og mutagen skade

Uracil er en av fire nukleobaser i RNA, men små mengder kan også befinne seg i DNA. Det er to kilder til genomisk uracil: misinkorporering av deoksyuridin trifosfat (dUTP) i steden for deoksytymidin trifosfat (dTTP) under DNA replikasjon, og deaminering av cytosin. dUTP misinkorporering danner uracil:adenin (U:A) par uten konsekvenser for genetisk informasjon. Cytosin deaminering resulterer derimot i uracil:guanin (U:G) par med mutagent potensiale. U:G par har derfor lenge blitt ansett som DNA skade, men i begynnelsen av 2000 tallet ble det etterhvert også klart at activation induced deaminase (AID) drevet cytosin deamineringer startpunktet for antistoffsmodningsprosessene somatisk hypermutasjon (SHM) og klasseskift rekombinajon (CSR). Til tross for sin rolle i det antistoffmodning har sekvensringsekperimenter av kreftgenomer vist at andre cytosin deaminaser (APOBECer) er ansvarlige for mutasjonssignaturer i mange ulike krefttyper.

Vårt arbeid har fokusert på uracils tilknytning til kreft og adaptiv immunitet. Mange uracil-relaterte observasjoner kommer fra musemodeller. I artikkel I undersøkte vi derfor aktivitetsnivå og mengde av tre uracilreparerende enzymer UNG2, SMUG1 og TDG i ulike cellelinjer fra menneske og mus. Vi fant at den totale uracilfjerningskapasiteten fra U:A og U:G substrater var større i humane celler enn for museceller, hovedsakelig fordi UNG-nivået var høyere. I museceller stod SMUG1 for 50 % av uracil-fjerningskapasitet fra U:G substrat, men bare 1 % i humane celler. Slike forskjeller er viktige å ta i betraktning når man bruker musemodeller i forskning rettet mot genomisk uracil. Vi stimulerte videre primære B-celler fra mus til å gjennomgå CSR. UNG og SMUG1 aktivitet i de stimulerte B-cellene var på samme nivå som i de øvrige musecellelinjene, hvilket indikerer at både UNG og SMUG1 er uttrykt og kan fjerne U:G fra deaminert cytosiner i aktiverte B-celler.

Flere forskningsgrupper har forsøkt å måle genomiske uracilnivåer, men resultatene deres varierte med nesten tre størrelsesordener. De store forskjellene i målte uracilnivåer kan forklares med biologiske ulikheter i prøvene til de ulike gruppene, men vi foreslår at det stammer fra tekniske svakheter i målemetodene. I artikkel II identifiserte vi og løste problemer ved DNA preparasjon som bidro til overvurderingen av uracilnivåer, hovedsakelig fra med-isolert dUMP og in vitro-generert uracil fra cytosindeaminering. Vi presenter en metode for absolutt kvantifisering av genomisk uracil hvor DNA hydrolyseres med enzymer til nucleosider, deoksycytidin fjernes ved væskekromatografi-fraksjonering, og deoxyuridin måles gjennom væskekromatografi-massespektrometri (LC/MS/MS). Vi testet metoden i en relevant biologiske sammenheng ved måling av genomisk uracilnivåer i mus embryonale fibroblaster (MEFer) og humane lymfoblastoide cellelinjer som ikke utrykker UNG. Vi fant at de UNG-manglene cellene inneholdte fem- til elve ganger mer genomisk uracil. Videre observerte vi at UNG-kompetente celler inneholdt 400 til 600 uraciler per genom. Våre uracilmålinger er lavere enn tidligere rapporterte genomisk uracil verdier. Dette indikerer at det basale uracil nivået har blitt overvurdert og at vår metode gir den mest nøyaktige uracil kvantifiseringen.

UNG2 er ansett som den viktigste glycosylasen i reparasjon av genomisk uracil, mens SMUG1 fungerer som støtte. I artikkel III studerte vi rollene til de to glycosylasene ved bruk av *Ung*$^{-/-}$*Smug1*$^{-/-}$ mus. Vi observerte at SMUG1i hovedsak står for genomisk 5-hydroxymethyluracil (5-hmU) fjerningsaktivitet i musevev og MEFer, og at *Smug1*$^{-/-}$ mus akkumulerte 5-hmU. I motsetning til dette hadde SMUG1-status ingen effekt på genomiske uracil nivåer og nesten ingen effekt på uracil fjerningsaktivitet. Fjerningsaktivteten ble kun funnet lavere i *Smug1*$^{+/-}$ *og Smug1*$^{-/-}$ hjerneprøver. UNG-mangel reduserte uracil fjerningsaktiviteten i alle vevsprøver bortsett fra hjerne, og *Ung*$^{-/-}$ mus hadde to- til tre-gang

høyere genomiske uracilnivåer enn villtype mus. Vi fant imidlertid en mye større økning, 3- til 20-ganger, av genomiske uracilnivåer i *Ung*$^{-/-}$*Smug1*$^{-/-}$ mus, i tillegg til en komplett ablasjon av uracil fjerningsaktiviteten. Vi forutsetter derfor at UNG er i hovedsak ansvarlig for uracil-fjerningsaktivitet i alle testete vev bortsett fra hjerne med SMUG1 som støtte, og at enten UNG eller SMUG1 er tilstrekkelig for å opprettholde basale nivåer av genomisk uracil.

Selv om det er mange bevis for at AID og APOBEC mutasjonssignaturer i kreft stammer fra enzymatisk cytosin deaminering har uracilnivåer ikke ble testet i sammenheng med AID og APOBEC uttrykk. I artikkel IV målte og korrelerte vi derfor genomisk uracil- og AID-nivåer i en rekke ulike B-celle lymfomcellelinjer. Vi fant tydelig korrelasjon mellom genomisk uracil og AID uttrykk i lymfomcellelinjer, og ingen korrelasjon i ikke-lymfomcellelinjer. I tillegg økte eller reduserte vi AID nivåer ved AID-overuttrykk, B-celle stimulering og shRNA mediert AID ekspresjonshemming. Vi observerte at AID økning førte til mer genomisk uracil mens AID senking reduserte mengden genomisk uracil. Vi korrelerte genomisk uracil med uracil-DNA glykosylasenivåer, men fant kun svakere korrelasjoner enn med AID uttrykk og kun i ikke-lymfomcellelinjer. Ved repetisjon av sekvenseringsanalysene rettet mot APOBEC mutasjonssignaturer fant vi en AID-spesifikke mutasjonssignatur i B-celle lymfomer og kronisk lymfatisk leukemi. Vi har dermed påvist at høye AID uttrykk følges av genomisk uracil akkumulering og at dette er assosiert med kreft in vivo.

I sum har dette arbeidet bidratt til bedre forståelse av de biologiske egenskapene til genomisk uracil og de proteinene som kontrollerer dets nivåer. Dette har gitt ny innsikt i hvordan uracil prosesseres og hva konsekvensene er ved feilregulering.

Kandidat: Antonio Sarno
Veiledere: Prof. Dr. Med. Hans E. Krokan og Dr. Per Arne Aas
Institutt: Institutt for Kreftforskning og Molekylærmedisin
Finansieringskilde: NTNU (lønn til PhD-kandidat)

Ovennevnte avhandling er funnet verdig til å forsvares offentlig for graden
PhD i Molekylærmedisin.
Disputas finner sted i
Auditoriet MTA, Medisinsk-teknisk forskningssenter,
Fredag 12. Juni, kl 12:15.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The work presented in this thesis was carried out at the Department of Cancer Research and Molecular Medicine, Medical Faculty, Norwegian University of Science and Technology (NTNU). I am grateful for the financial support from NTNU. I also thank the following people:

My main supervisor, Hans E. Krokan, for giving me the opportunity to work in his group, trusting me with the freedom to make my own mistakes, and always being there when I needed help. As I wrote my thesis, I thought I knew a thing or two about the field of base excision repair and uracil. Then, I would meet with Hans and realize I was on a one-way train to Wrongsville. Hans knows better and he remembers the relevant paper. I honestly can't imagine myself reaching that point, Hans, but it's my goal. Thank you for providing that.

I have never experienced a shortage of guidance during my studies. My co-supervisor, Per Arne Aas, has been a constant help in the lab. Thanks for that and for always checking in on me. Geir Slupphaug has always been like an unofficial supervisor to me. Thanks for your knowledge and support. We began several projects with Hilde Nilsen and her group (Institute of Clinical Medicine, Department of Clinical Molecular Biology, University of Oslo) during a time when my research had lost direction. Thank you, Hilde, for working with us, for great discussions, and for prioritizing my PhD delivery schedule. I also thank the members of the Nilsen group for their contributions, particularly Lene Alsøe and Sergio Carracedo for their work on the SMUG1 manuscript. Finally, Cathrine Broberg Vågbø, Lars Hagen, and Berit Doseth were all instrumental in my training. Takk, Cathrine. Bongiorno, Lorenzo. Nei, Berit.

Much of my work was done with Anastasia Galashevskaya, who showed me that sometimes it's time to stop trying to tweak the method and start collecting data. Thanks for that, for all your help, and for tolerating me for so long. Also, I reprimand Audun Hanssen-Bauer, Rebekka Müller, and Mirta Sousa for both wasting my time and bothering poor, innocent Anastasia while doing so. Audun, thank you for the "science" talks, the games, and the proofreading. Müller, I'll be eating avocadi long after you forget me, but I'll stay away from the berliner (also proofreading). Mirta, you are distracting me literally as I write this. Thank you all for your much-needed distractions. There are quite a few more people who have worked at the department and made this a great place to work. They are too numerous to mention, but I thank you all. A special thanks to some people I've worked with: Anders Wallenius, Henrik Sahlin Pettersen, Ida Ericsson, Kamila Zub, and Nina-Beate Liabakk.

Finally, thank you mom, for the sacrifices you've made to get me here and the love you've provided. Mari, I thank you for your unconditional love and support and for accepting a half husband for these past months. Thank you, William, for washing away my bitterness after a long day's fruitless work with nothing but your laugh. I love you all.

Trondheim, Norway

December, 2014 Antonio Sarno

LIST OF ARTICLES

Article I

Uracil-DNA glycosylase in base excision repair and adaptive immunity: species differences between man and mouse.

Doseth B, Visnes T, Wallenius A, Ericsson I, Sarno A, Pettersen HS, Flatberg A, Catterall T, Slupphaug G, Krokan HE, Kavli B.

J Biol Chem. 2011 May 13;286(19):16669-80

Article II

A robust, sensitive assay for genomic uracil determination by LC/MS/MS reveals lower levels than previously reported.

Galashevskaya A*, Sarno A*, Vågbø CB, Aas PA, Hagen L, Slupphaug G, Krokan HE.

DNA Repair (Amst). 2013 Sep;12(9):699-706. (*shared first authorship)

Article III

UNG and SMUG1 efficiently complement each other in removing genomic uracil from mouse organs.

Sarno A*, Alsøe L*, Galashevskaya A, Tekin NB, Jobert L, SenGupta T, Carracedo S, Krokan HE, Nilsen H.

Manuscript. (*shared first authorship)

Article IV

AID expression in B-cell lymphomas causes accumulation of genomic uracil and a distinct AID mutational signature.

Pettersen HS, Galashevskaya A, Doseth B, Sousa MML, Sarno A, Visnes T, Aas PA, Liabakk NB, Slupphaug G, Sætrom P, Kavli B, Krokan HE.

DNA Repair (Amst). 2014 Nov 24;25C:60-71.

Articles or manuscripts not included in the thesis:

Modulation of cell metabolic pathways and oxidative stress signaling contribute to acquired melphalan resistance in multiple myeloma cells.

Zub KA, Sousa MML, Sarno A, Sharma S, Demirovic A, Rao S, Young C, Aas PA, Ericsson I, Sundan A, Jensen ON, Slupphaug G.

PLoS One. 2015 Mar 13;10(3):e0119857.

Expression and recruitment of uracil-DNA glycosylase are regulated by E2A during antibody diversification.

Wallenius A, Hauser J, Aas PA, Sarno A, Kavli B, Krokan HE, Grundström T.

Mol Immunol. 2014 Jul;60(1):23-31.

An inverse switch in DNA base excision and strand break repair contributes to melphalan resistance in multiple myeloma cells.

Sousa MM, Zub KA, Aas PA, Hanssen-Bauer A, Demirovic A, Sarno A, Tian E, Liabakk NB, Slupphaug G.

PLoS One. 2013;8(2):e55493.

ABBREVIATIONS

5-BU	5-bromouracil
5-caC	5-carboxylcytosine
5-caU	5-carboxyluracil
5-CU	5-chlorouracil
5-fC	5-formylcytosine
5-fU	5-formyluracil
5-FU	5-fluorouracil
5-hU	5-hydroxyuracil
5-hmC	5-hydroxymethylcytosine
5-hmU	5-hydroxymethyluracil
5-mC	5-methylcytosine
A, Ade	adenine
Ado	adenosine
AID	activation induced deaminase
ALL	acute lymphoblastic leukemia
AP	abasic or apurinic/apyrimidinic
APE	apurinic/apyrimidinic endonuclease
APOB	apolipoprotein B
APOBEC, A1/2/3/4	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 1/2/3/4
APC	adenomatous polyposis coli
BAFF	B-cell activating factor
BER	base excision repair
BCL	B-cell lymphoma protein
BCR	B-cell receptor
C	constant region of the immunoglobulin gene
C, Cyt	cytosine
CD40L	CD40 ligand
CLL	chronic lymphocytic leukemia
CSR	class switch recombination
CUL5	E3 ubiquitin ligase Cullin-5
Cyd	cytidine
CYP	cytochrome P450

dA, dAdo	deoxyadenosine
dAMP, dADP, dATP	deoxyadenosine mono/di/triphosphate
dC, dCyd	deoxycytidine
dCMP, dCDP, dCTP	deoxycytidine mono/di/triphosphate
dG, dGuo	deoxyguanosine
dGMP, dGDP, dGTP	deoxyguanosine mono/di/triphosphate,
D	diversity region of the immunoglobulin gene
DKC1	pseudouridine synthetase dyskerin
DNA-PK(cs)	DNA protein kinase (catalytic subunit)
DNMT	DNA (cytosine-5)-methyltransferase
dR	deoxyribose
dRP	deoxyribose phosphate
DSB	DNA double-stranded break
dsDNA	double-stranded DNA
dT, dThd	deoxythymidine
dTMP, dTDP, dTTP	deoxythymidine mono/di/triphosphate
dU, dUrd	deoxyuridine
dUMP, dUDP, dUTP	deoxyuridine mono/di/triphosphate
dUTPase	5′-triphosphate nucleotidohydrolase
DTYMK	dTMP by thymidylate kinase
εC	$3,N^4$-ethenocytosine
ESCC	esophageal squamous cell carcinoma
EXO1	exonuclease 1
FEN1	flap endonuclease 1
G, Gua	guanine
GADD45	DNA damage-inducible protein GADD45
Gag	group specific antigen
GC	germinal center
Guo	guanosine
HBV	hepatitis B virus
HCV	hepatitis C virus
HHV	human herpesvirus
HIGM	hyper-immunoglobulin M
HIV	human immunodeficiency virus

HTLV	human T-cell lymphotropic virus
IDL	insertion/deletion loop
IFN	interferon
Ig	immunoglobulin
IL	interleukin
J	J region of the immunoglobulin gene
LIG	DNA ligase
LP-BER	long patch-base excision repair
LPS	lipopolysaccharide
MBD4	methyl-CpG-binding domain protein 4
MCM7	DNA replication licensing factor MCM7
mESC	mouse embryonic stem cells
MEF	mouse embryonic fibroblast
MLH	MutL homolog
MMR	mismatch repair
MSH	MutS homolog
MTHF	N5,N10-methylene-tetrahydrofolate
MYC	Myc proto-oncogene protein
N_2O_2	nitrous anhydride
NDK	nucleoside diphosphate kinase
NES	nuclear export signal
NHEJ	non-homologous end-joining
NO•	nitric oxide
PCNA	proliferating cell nuclear antigen
PKC	protein kinase C
PMS2	mismatch repair endonuclease PMS2
PNKP	polynucleotide kinase 3′-phosphate
POL	DNA polymerase
PTM	protein post-translational modification
RAG	recombination activating gene
RBBP8	DNA endonuclease RBBP8
REV1	DNA protein REV1
RFC	replication factor C
RPA	replication protein A

rRNA	ribosomal RNA
S	switch region of the immunoglobulin gene
SAM	S-adenosylmethionine
SHM	somatic hypermutation
shRNA	small hairpin RNA
SMUG1	single-strand selective monofunctional uracil DNA glycosylase 1
SP-BER	short patch-base excision repair
SSB	DNA single-stranded break
ssDNA	single-stranded DNA
T, Thy	thymine
TDG	G/T mismatch-specific thymine DNA glycosylase
TET	ten eleven translocation
Thd	thymidine
TLR	toll-like receptor
TLS	translesion synthesis
TP53	cellular tumor antigen p53
TS	thymidylate synthase
U, Ura	uracil
Ugi	uracil-DNA glycosylase inhibitor
UNG	uracil-DNA glycosylase
Urd	uridine
V	variable region of the immunoglobulin gene
Vif	viral infectivity factor
XLF	XRCC4-like factor
XRCC	X-ray repair cross-complementing protein

1. INTRODUCTION

Genomic uracil is both a lesion and a necessary intermediate for adaptive immunity in mammalian cells. In this sense, both an abundance and a shortage of DNA uracil through either genetic abnormalities or exogenous perturbations can lead to pathological phenotypes. In this thesis, I will first provide an overview of the possible sources of genomic uracil (section 2). Then, I will describe the main and alternative pathways for its removal from DNA (sections 3 and 4). Next, I will delve more deeply into endogenous DNA uracilation and explain how genomic uracil is used as an intermediate in antibody maturation and as a tool to restrict the replication of exogenous viral genomes (sections 5 and 6). Finally, I will provide an overview of how both an abundance and a shortage of genomic uracil can lead to human disease (section 7).

Before continuing onwards, I encourage the reader to familiarize him- or herself with the abbreviations section on the previous page. I will generally list an unabbreviated name before using the abbreviation, but a section or more may sometimes stand between the original unabbreviated name of a molecule, idea, or process and its subsequent abbreviated use, leading to confusion. Furthermore, I will not expand on the abbreviations between the five canonical DNA and RNA bases, nucleosides, and nucleotides. The general nomenclature in this case is as follows, using adenine as an example: the free base is A or Ade, the RNA nucleoside is rA or rAdo, the DNA nucleoside is dA or dAdo, and the nucleotides are (d)AMP, ADP, and ATP for mono-, di, and triphosphorylated forms, respectively.

2. Sources of Genomic Uracil

Uracil is a canonical RNA base that can be found in DNA in very low quantities. In this section, I will describe the three mechanisms by which uracil is introduced into DNA: dUTP misincorporation during DNA replication instead of dTTP (section 2.1), non-enzymatic, spontaneous or chemical-induced cytosine deamination to uracil (section 2.2), and enzymatic cytosine deamination to uracil by DNA cytidine deaminases (section 2.3). There is no overt deleterious effect of dUTP misincorporation into DNA instead of dTTP because the resulting U:A pairs identically to the original T:A pair, although the uracil may be recognized and improperly processed to yield a mutation. Cytosine deamination to uracil is mutagenic and yields a C:G to T:A transition mutation after replication.

2.1 dUTP misincorporation during DNA replication

Mammalian DNA polymerases cannot differentiate between dUTP and dTTP [1], so the extent of dUTP misincorporation depends on the dUTP/dTTP ratio in the vicinity of the polymerase during DNA synthesis (figure 1) [2,3]. Thus, a cell must maintain a low

dUTP/dTTP ratio to avoid excessive DNA uracilation. The normal intracellular ratio is estimated at well below 1% [4].

dUTP is an intermediate in normal pyrimidine metabolism (reviewed in [3]). Low dUTP levels are maintained by deoxyuridine 5′-triphosphate nucleotidohydrolase (dUTPase), which dephosphorylates dUTP to dUMP [5]. The dUMP is converted to dTMP by thymidylate synthase (TS) using N5,N10-methylene-tetrahydrofolate (MTHF) as a methyl donor. dTTP is eventually generated from the dTMP by thymidylate kinase (DTYMK) and nucleoside diphosphate kinase (NDK) and used in replication [3,6–8]. The main factors controlling the dUTP/dTTP pool are dUTPase and TS activities, and it has been estimated that one dUTP per 10^4 dTTPs is incorporated per cell per day [2,9]. This amounts to ~80,000 misincorporated uracils per genome (~27 dUrd per 10^6 dN), assuming a 46.1% GC content and that G = C and A = T (Chargaff's rule) [10,11].

A significant reduction of either dUTPase or TS activity leads to a phenomenon called "thymineless death," which is thought to be primarily a result of dTTP pool depletion instead of dUTP misincorporation [12,13]. It has been proposed that thymineless cells accumulate single-stranded DNA (ssDNA) gaps that are converted into double-stranded DNA (dsDNA) breaks (DSBs) behind the replication fork. The released dTMP can be converted to dTTP and used for new polymerase initiations, but further disintegration of small replication bubbles causes replication origin destruction [14]. Thymineless death seems to be independent of dUTP incorporation [15,16]. More subtle attenuation of genomic uracil misincorporation by TS inhibition will be discussed in section 7.2.4.

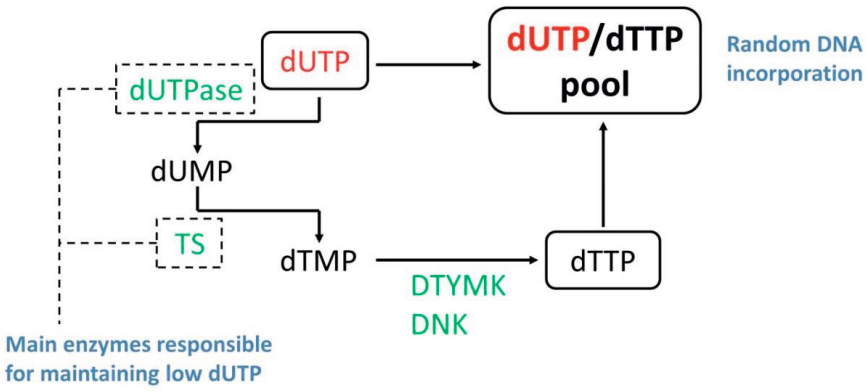


Figure 1: dUTPase and TS activities regulate the amount of dUTP in the cell. dUTP is degraded to dUMP by dUTPase and converted to dTMP by TS. The dTMP is subsequently converted to dTTP by DTYMK and DNK. The dUTP misincorporation rate is a factor of the dUTP/dTTP ratio.

2.2 Non-enzymatic cytosine deamination to uracil

Cytosine is spontaneously deaminated to uracil in DNA. The spontaneous deamination rate of ssDNA is $2x10^{-10}$ deaminations per second and the rate of dsDNA is 200 to 300 times slower [17,18]. The deamination rate was measured *in vitro* by heat-inducing deamination at 95 °C and extrapolating to 37 °C. The actual *in cellulo* deamination rate may be attuned by *e.g.* different solution conditions, protein binding, and DNA supercoiling. The spontaneous deamination rate has been estimated to be 100 to 500 cytosines per cell per day (0.033 to 0.17 dUrd per 10^6 dN) if 0.1% of the genome is single-stranded [19].

Nitrous anhydride (N_2O_2) can also induce cytosine deamination (called nitrosative deamination) [20,21]. N_2O_2 is an intermediate in nitric oxide (NO•) production. NO• is produced as a physiological response to infection, a regulator of immune functions, or a chemical messenger for neurons (reviewed in [22]). Inflammation and cancer have been linked to the generation of high NO• levels, and an estimated 20% of cancers are caused by inflammation and chronic infection [22,23]; however, deamination only occurs at toxic N_2O_2 levels [24,25]. Thus, nitrosative deamination is not likely to be a significant contributor to the genomic uracil load.

2.3 Enzymatic cytosine deamination to uracil

Enzymatic cytosine deamination is primarily performed by the apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) enzyme family. APOBECs have one or two zinc (II)-binding catalytic domains that catalyze cytosine deamination in DNA or RNA (reviewed in [26–28]). There are eleven known APOBECs in humans: activation induced deaminase (AID), APOBEC1, APOBEC2, APOBEC3A through APOBEC3G, and APOBEC4 (A1, A2, A3A through A3G, and A4, respectively). A1, A2, and A4 have not been shown to have verifiable DNA cytosine deamination activity or contribution to the genomic uracil load, so I will describe them only briefly and focus on AID and A3.

A1 is primarily an mRNA editing deaminase present in the gastrointestinal tract that participates in lipid metabolism [29]. Briefly, A1 deaminates apolipoprotein B (APOB) pre-mRNA to change a glutamine codon to a stop codon, shortening the subsequent APOB protein [30,31]. APOB is the primary lipoprotein in chylomicrons. The shortened protein has an accelerated plasma turnover relative to the full length protein, so the editing activity of APOBEC1 is crucial for maintaining physiological concentrations of APOB-containing lipoproteins in plasma [28]. A2 has no known physiological substrate, is expressed only in heart and skeletal muscles, and is thought to promote muscle fiber differentiation [32–34]. Its

function remains largely unelucidated, but it has been shown to deaminate methylated cytosines, which may indicate a role in epigenetic regulation [28,35]. Finally, A4 also has no known substrate. It is found in the testis and has been suggested as a possible mRNA editing enzyme in spermatogenesis, though this has not been researched and no detectable mutagenic activity in bacterial or yeast assays has been detected [36,37].

A3 enzymes have been shown to deaminate ssDNA and their primary role seems to be acting against exogenous viruses and endogenous retroelements [38–41]. Section 6 contains a brief overview of the role of A3 enzymes in innate immunity. A3D, A3F, and A3G are cytoplasmic, A3B is nuclear, and A3A, A3C, and A3H are both cytoplasmic and nuclear [39,42–44]. Moreover, A3 enzymes deaminate at specific target sequences [45–48]. In general, A3 enzymes are found in T-cell, B-cells and phagocytic cells, although they are not confined to immune cells and can be expressed in *e.g.* the ovaries and testes and embryonic stem cells [49,50]. There is no clear agreement as to relative abundance of A3 in different tissues because of technical limitations in their measurement, but the general consensus suggests a broad and constitutive expression profile in humans [28]. Given their prevalence, nuclear localization, and mutagenicity, it is unsurprising that recent evidence indicates that A3 enzymes play a large role in the development of many cancers. The pathological consequences of A3 mutagenicity will be discussed in section 7.2.5.

AID also deaminates ssDNA and is essential for immunoglobulin (Ig) gene diversification in activated mature B-cells [51–54]. Normally, AID expression is restricted to germinal center (GC) B-cells, although there is evidence of AID expression in other cell types as a result of hepatitis C or *H. pylori* infection as well as pro-inflammatory cytokines [55–57]. AID targets the sequence WRCY (W=A/T, R=A/G, Y=T/A). The role of AID in antibody maturation is discussed in sections 5.3 and 5.4 and the consequences of AID deamination outside the *Ig* loci is discussed in section 7.2.5.

Cytosines can also be deaminated by methyltransferases. Methylation of cytosines at CpG islands by methyltransferases is fundamental to epigenetic gene silencing. DNA (cytosine-5)-methyltransferases (DNMTs) convert cytosines to 5-methylcytosines (5-mC) using S-adenosylmethionine (SAM) as a methyl donor [58]. This reaction includes a dihydropurine intermediate that can spontaneously deaminate to uracil prior to methyl transfer from SAM. Thus, either dysfunctional DNMTs or low SAM levels result in DNMT1 cytosine deamination [59,60]. Several prokaryotic methyltransferases as well as the catalytic domain of DNMT3a have been shown to deaminate cytosine [59,61,62]. Not all DNMTs exhibit this activity and their contribution to the genomic uracil load remains unstudied [63].

3. Repair of Genomic Uracil by the Base Excision Repair Pathway

Repair of genomic uracil is usually performed by the base excision repair (BER) pathway, in which uracil-DNA glycosylases initiate the repair process by recognizing and excising uracil (figure 2). BER also repairs nucleobase oxidations, alkylations, and deaminations, as well as misincorporated nucleotides, abasic (AP) sites, and ssDNA breaks (SSBs), the specificity of which depends on the initiating DNA glycosylase (reviewed in [64,65]). BER can proceed via short- or long-patch (SP or LP, respectively) repair pathways or DNA polymerase β (POLβ) dependent two nucleotide insertion [66–68]. In this section, I will provide a step-by-step overview of SP- and LP-BER pathways and then provide a more detailed examination of the known uracil-DNA glycosylases.

3.1 General base excision repair pathway

The exact mechanism by which glycosylases scan DNA for their respective lesions is not fully elucidated. For non-recruited DNA scanning, *i.e.* if the glycosylase is not recruited to the vicinity of the replication fork for scanning during replication, a glycosylase requires a balance of specialized thermodynamics and kinetics for its interaction with both specific and non-specific DNA sequences [69]. Should the enzyme bind too tightly to non-specific sequences, it may not have the opportunity to scan the entire genome before further replication. Conversely, if the enzyme binds too weakly, some lesions may be overlooked [70,71]. The general consensus in the field is that most DNA glycosylases scan DNA in two ways (reviewed in [72]). The first scanning mode is referred to as "DNA sliding" and involves tracking along the DNA using a loosely-associated enzyme state, and the second mode is referred to as "DNA hopping" and involves intermittent dissociation and re-association with the DNA [73–75]. Due to their complexity, studies on glycosylase DNA scanning have been performed with simplified *in vitro* systems, so the effects of glycosylase recruitment to *e.g.* the replication fork, post-translational modifications (PTMs), and glycosylase-protein binding have not been studied.

The common initiation step for all glycosylases upon lesion recognition is the "flipping" of the aberrant base out of the DNA helix and into the enzyme's substrate recognition pocket in which the *N*-glycosidic bond between the C1 on the deoxyribose (dR) and the nucleobase is cleaved, leaving an AP site (reviewed in [76]). Most organisms have several DNA glycosylases for the removal of various damages, which has traditionally been thought to indicate a level of redundancy by the BER glycosylases [76]. After base removal, apurinic/apyrimidinic endonuclease 1 (APE1) nicks the phosphodiester bond 5′ to the AP site, leaving a 3′-OH and a 5′-deoxyribose phosphate (dRP) moiety [76,77]. Some

glycosylases (called bifunctional glycosylases) also have AP-lyase activity that allows them to nick the phosphodiester bond 3′ of the AP site by β-elimination or both 3′ and 5′ of the AP site by β/δ-elimination. The AP-lyase-derived nick leaves a fragmented sugar derivative 3′ of the strand break and is therefore blocks DNA polymerases, so APE1 or polynucleotide kinase 3′-phosphate (PNKP) must process the nick before subsequent DNA synthesis [78].

DNA synthesis can then proceed via SP- or LP-BER, which insert one or two to twelve nucleotides, respectively. The choice of SP/LP-BER is likely determined by relative protein concentrations and specific interactions between repair and scaffolding proteins at the lesion, as well as lesion type, stress response, cell type, and cell cycle [66–68,79–84]. Most glycosylases responsible for oxidatively damaged base excision are bifunctional and tend to follow SP-BER while monofunctional glycosylases follow either SP- or LP-BER [85]. It has therefore been proposed that since oxidatively damaged bases often occur in clusters, closely-spaced LP-BER could lead to strand breakage, so SP-BER is a more sensible choice for oxidatively damaged bases [85]. Despite these insights, the exact mechanisms determining SP- versus LP-BER remain unknown.

In SP-BER, one nucleotide is usually inserted by POLβ or (less frequently) POLλ and the nick is ligated by DNA ligase I (LIG1) or LIG3α in complex with scaffolding protein X-ray repair cross-complementing protein 1 (XRCC1) [65,86–88]. Several other polymerases (POLγ, POLθ, POLι, and POLλ) have also been proposed to be involved in BER [89–92]. LIG3 is stabilized by but can function independently of XRCC1 and is essential in mitochondrial BER, but apparently dispensable in nuclear BER [93–95]. In LP-BER, two to twelve nucleotides are inserted by POLβ or POLδ and POLε in complex with proliferating cell nuclear antigen (PCNA) and replication factor C (RFC). The nucleotide insertion displaces the nucleotides 3′ of the AP site, leaving a ssDNA “flap.” The flap is removed by flap endonuclease 1 (FEN1) and ligation is performed by LIG1 [96].

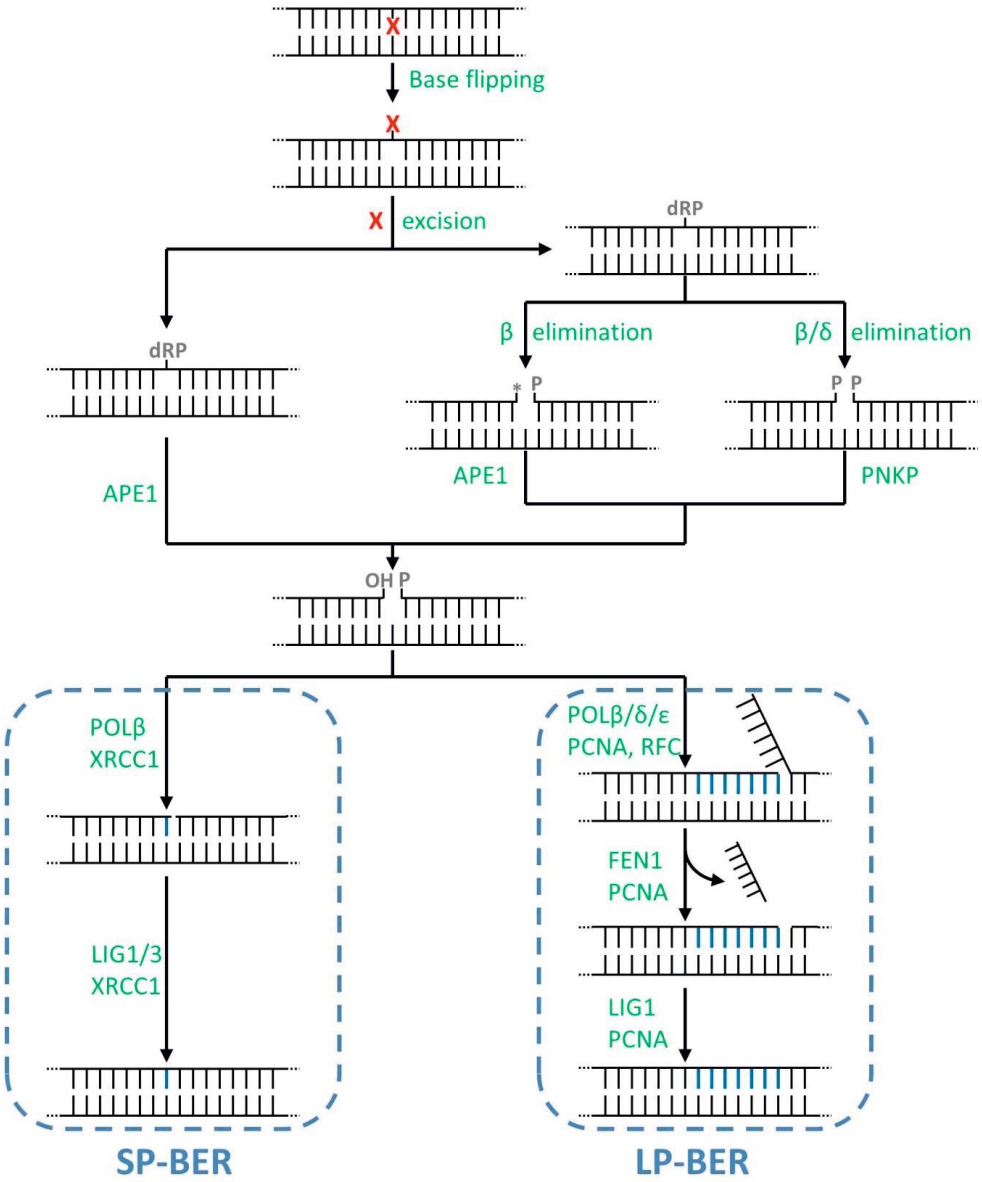


Figure 2: The SP- and LP-BER pathways. The base is recognized and excised by a monofunctional glycosylase that levels an AP-site or bifunctional glycosylase also nick the strand. The AP-site is nicked by APE1 and the nicked strand processed by APE1 or PNKP, resulting in a SSB with a 3′-OH and 5′-phosphate. In SP-BER, POLβ fills in the nucleotide and LIG1 or LIG3 ligate the nick. In LP-BER, POLβ, POLδ, or POLε fill in two to twelve nucleotides, FEN1 removes the flap, and LIG1 ligates the nick. Modified from [97].

3.2 Uracil recognition and excision by uracil-DNA glycosylases

There are five known uracil-excising DNA glycosylases: uracil-DNA glycosylase 1 and 2 (UNG1/2), single-strand selective monofunctional uracil DNA glycosylase 1 (SMUG1), G/T mismatch-specific thymine DNA glycosylase (TDG), and methyl-CpG-binding domain protein 4 (MBD4) (table 1). Their activities depend on sequence context, DNA strandedness (*i.e.* ssDNA or dsDNA), the base opposite the uracil (U:A or U:G), cell cycle, PTMs, interaction partners, and their location in the genome [97,98]. Several of these parameters were studied in Articles I and III of this thesis.

3.2.1 UNG1 and UNG2

UNG is the major uracil glycosylase in mammalian cells [99,100]. The *UNG* gene encodes two isoforms with identical catalytic domains and unique N-termini that determine protein localization [101]. UNG1 expression is controlled by the P_B promoter in the *UNG* gene and localizes to mitochondria, and UNG2 is controlled by the P_A promoter and localizes to the nucleus [101]. UNG1 mRNA is expressed in all tissues, but UNG2 mRNA is mainly found in proliferative tissue like the testis, colon mucosa, small intestine, and the thymus [102]. UNG mRNA and protein levels are cell cycle-dependent: UNG2 is up-regulated in late G_1- and early S-phase and UNG1 is constitutive expressed with a less pronounced up-regulation in early S-phase [102]. UNG2 also preferentially excises uracil from ssDNA contexts than from dsDNA, and U:G over U:A [100,103]. Furthermore, sequence context specificity (*i.e.* the flanking nucleotides) trumps U:G over U:A preference [98,104,105]. In addition to uracil, UNG2 has been shown to less efficiently excise uracil analogues with minor 5′ modifications like 5-fluorouracil (5-FU), 5-hydroxyuracil (5-hU), alloxan, and isodialuric acid [77,100,105,106].

It makes biological sense that the UNG2 is the main uracil-excising enzyme coupled to DNA replication because many more uracils probably arise from dUTP minsincorporation during replication than from spontaneous cytosine deamination. UNG2 has a much higher catalytic turnover rate compared to other uracil-DNA glycosylases, so it is well suited for fast uracil excision in the replication fork where dUTP misincorporation takes place and the spontaneous deamination rate increased [17,100]. Furthermore, UNG2 is phosphorylated at Ser23 in late G_1- and early S-phase, which increases its association with the DNA binding protein replication protein A (RPA) catalytic turnover rate on ssDNA [107]. Conversely, UNG2 Ser64 and Thr60 phosphorylations throughout the S-phase reduce RPA binding and facilitate ubiquitinylation and proteosomal degradation in G_2 [107]. Co-immunoprecipitation experiments have also shown that BER complexes contain UNG as the only uracil-DNA

glycosylase, APE1, POLα, POLβ, POLδ, POLε, DNA protein kinase (DNA-PK), XRCC1, PCNA, LIG1, DNA replication licensing factor MCM7 (MCM7, also referred to as mini-chromosome maintenance protein 7), and cyclin A, supporting replication-associated repair [108,109]. Finally, UNG2 co-localizes with RPA and PCNA in replication foci [110,111].

UNG2's role in the mutagenic processing of genomic uracil in antibody maturation will be discussed in sections 5.3 and 5.4, and the consequences of UNG2 deficiency will be discussed in sections 7.2.1 and 7.2.2.

3.2.2 SMUG1

Like UNG2, SMUG1 excises uracil, 5-hU, 5-FU, alloxan, and it also excises 5-hydroxymethyluracil (5-hmU), 5-carboxyluracil (5-caU) and 5-formyluracil (5-fU) [100,112–114]. Unlike UNG2, SMUG1 is constitutively expressed throughout the cell cycle and is thought to be the primary backup for UNG2 in the repair of U:A and U:G, although it cannot compensate for UNG2 deficiency in human B-cells [100,115]. While UNG2 is responsible for rapid repair of both U:A and U:G during replication, SMUG1 is thought to be more important for U:G repair in non-replicative chromatin [116].

SMUG1 has been suggested as a reader of epigenetic markers like 5-hmU in addition to its role as a repair glycosylase [117]. Up-regulation in ten eleven translocation (TET) enzymes, which are linked to DNA demethylation, induced 5-hmU accumulation [118,119]. Furthermore, 5-hmU was recently shown to be introduced by enzymatic oxidation of thymine by TET enzymes in mouse embryonic stem cells (mESC) independently of 5-hydroxymethylcytosine (5-hmC) or ROS-mediated thymine oxidation [120]. *Smug1*$^{-/-}$ mouse embryonic fibroblasts lack 5-hmU excision activity and resist normally toxic concentrations of 5-hmdUrd treatment [121]. It had been speculated that a glycosylase excising 5-hmU in DNA could be involved in the demethylation of 5-mC [122]. The theory has gained credibility in light of evidence showing that 5-hmC is a natural DNA component [123–125]. 5-mC could thus be demethylated by 5-mC hydroxylation to 5-hmC, deamination to 5-hmU by AID/APOBECs, excision by SMUG1, and BER insertion of C [118,119,126]. This mechanism has been called into question since the discovery that AID/APOBECS do not deaminate 5-hmC and deaminate 5-mC ten-fold slower than the unmodified base [127,128]; however, both 5-hmU and SMUG1 may yet have undiscovered epigenetic roles.

There is also evidence for a role of SMUG1 in ribosomal RNA (rRNA) regulation. SMUG1 has broad nuclear localization and is slightly enriched in the nucleoli, where rRNA synthesis and metabolism takes place [100,129]. SMUG1 depletion in mice leads to 5-hmU accumulation in RNA and SMUG1 interacts with pseudouridine synthetase dyskerin (DKC1),

indicating a role for SMUG1 in rRNA quality control in that 5-hmU-modified RNA may be degraded by the exosome in its absence [130].

3.2.3 TDG

TDG excises U:G (not U:A) and T:G that presumably originate from C or 5-mC deamination, respectively, with a preference towards U:G [131,132]. TDG has a very low turnover compared to UNG2 because it tightly binds to the AP site generated by uracil excision [133,134]. C-terminal TDG SUMOylation at Lys350 reduces AP site binding affinity and increases U:G turnover while decreasing T:G activity [134,135]. So although TDG does not seem like a viable backup for UNG2 and SMUG1, its SUMOylated form plays a role in uracil repair. In addition to uracil and thymine, TDG can excise 5-hU, 5-hmC, 5-FU, the lipid peroxidation product 3,N^4-ethenocytosine (εC), and the 5-mC oxidation products 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) [136–139].

TDG is important in epigenetic regulation during embryonic development. Knockout or catalytic inactivation of TDG leads to embryonic lethality in mice [126,126]. The TDG-deficient embryos showed a decrease in developmental transcription factors (*e.g.* Hox) by perturbed methylation at their regulatory domains. In addition, TDG knockdown led to 5-caC accumulation in embryonic stem cells and simultaneous TET and TDG overexpression led to 5-caC and 5-fC depletion in HEK293 cells [127,138]. The fact that TDG has been shown to excise TET-oxidized 5-fC and 5-caC combined with their accumulation in TDG knockdown cells is a clear indicator of TDGs role in the demethylation of 5-mC after TET oxidation [140]. This represents the first validation of a pathway for active DNA demethylation.

3.2.4 MBD4

MBD4 has similar substrate preference to TDG, but preferentially binds to 5-mCG:TG and excises neither 5-fC nor 5-caC [141–143]. In addition, MBD4 has been shown to excise the peroxidase-mediated inflammation products 5-chlorouracil (5-CU) and 5-bromouracil (5-BU), the chemotherapy products 5-FU and 5-iodouracil, and εC [144–148]. There is evidence for direct DNA demethylation by MBD4 5-mC excision using growth arrest and DNA damage-inducible protein GADD45 (GADD45) as a scaffold [35,149]. Protein kinase C (PKC) phosphorylation of MBD4 has also been shown to potentiate its 5-mC glycosylase activity following parathyroid hormone stimulation, leading to demethylation within the cytochrome P450 27B1 (CYP27B1) promoter [150] . This indicates that MBD4 may have a role in the epigenetic de-repression of hormone regulated genes [143].

Table 1: Uracil-DNA glycosylases (modified from [65]).

Enzyme	Substrates	Mouse knockout	Human disease
UNG2	U, 5-FU alloxan, 5-hU, isodialuric acid	Viable and fertile, 95% reduced CSR, skewed SHM, B-cell lymphomas	HIGM syndrome, lymphoid hyperplasia
SMUG1	5-hmU U:G > U:A > ssU 5-FU, εC	Viable and fertile, increased mutations	Unknown
TDG	U:G > T:G 5-FU, εC, 5-hU 5-hmU, 5-fC, 5-caC	Embryonic lethal, epigenetic role in development	Unknown
MBD4	U:G and T:G 5-hmU, 5-mC εC, 5-FU, 5-CU, 5-BU	Viable and fertile, increased mutations	Mutated in carcinomas with microsatellite instability

4. Repair of Genomic Uracil by the Mismatch Repair Pathway

The primary role of the mismatch repair (MMR) pathway is the repair of base-base mismatches and insertion/deletion loops (IDLs) (reviewed in [151]). Although MMR has not been shown to be employed in the removal of genomic uracil under normal circumstances, it is crucial for mutagenic uracil processing in antibody maturation. In this section, I will give a general overview of the MMR pathway. I will also briefly mention uracil repair by MMR to provide a framework for better understanding the section on antibody maturation (section 5).

4.1 General mismatch repair pathway

The key proteins in MMR are the MutS and MutL protein families (reviewed in [152]). The MutSα dimer is composed of the MSH (MutS homolog) 2 and MSH6 heterodimer and the MutSβ dimer is composed of the MSH2 and MSH3 heterodimer. MutSα recognizes base-base mismatches and IDLs and MutSβ recognizes longer IDLs and is thought to be unable to repair base-base mismatches [153,154]. The MutLα dimer is composed of MutL homolog (MLH) 1 and mismatch repair endonuclease PMS2 (PMS2) and interacts with MutS dimers to recruit downstream repair proteins by signaling the mismatch recognition [155]. This can alternatively be performed by MutLβ (MLH1 and PMS1) or MutLγ (MLH1 and MLH3), which may be involved in repairing IDLs and in meiotic recombination [156,157].

Exonuclease 1 (EXO1) is also recruited to coordinate downstream proteins to finalize repair. EXO1 starts 5′-directed mismatch excision in the presence of MutSα or MutSβ and RPA [158,159]. Alternatively, EXO1 can introduce a 3′ nick directed excision with MutLα, which is activated by PCNA and RFC [160,161]. MutLα endonuclease also incises 5′ to the mismatch after recognition of the 3′ nick and mismatch. Then, EXO1 excises 5′ to 3′ from the MutLα incision site through and beyond the reach of the mismatch [160]. Although sterile because of meiotic effects, EXO1-deficient mice do not accumulate many mutations, so it is possible that other exonucleases are involved [162]. Moreover, EXO1-deficiency in mice increases lymphoma susceptibility and decreases survival compared to wild-type animals, but they have a higher survival than MSH2-deficient mice [162]. After the error is removed, POLδ synthesizes a new strand and LIG1 ligates the nick [152,163].

There are two moving models for the events following mismatch recognition. In the "stationary" model, MMR proteins induce DNA bending or looping that brings two distant sites together [164,165]. MutSα and MutSβ remain bound at the mismatch, and MutSα ATPase activity acts in a proofreading role to verify mismatch binding before proceeding with downstream excision. In the "moving" model, MutSα/β-MutLα complexes load at a mismatch site and then search for the strand break where exonucleases can be recruited to initiate excision [166–170].

To preserve genome integrity, MMR should only occur on the newly-synthesized strand with the mispaired nucleotide. Strand discrimination is strand-specific, but the source of the nicking activity is not fully understood [171]. Bound PCNA determines the orientation of the MutLα incision on the leading strand and enhances its endonuclease activity [172]. In the lagging strand, the 5′ ends of the Okazaki fragments are used for strand discrimination [173].

There is very little evidence of uracil repair by the MMR pathway outside of antibody maturation. Human MutSα has been shown to bind to U:G substrate *in vitro* [174]. MutSα also preferentially recognizes U:G or UU:GG relative to UU:AA homoduplexes, as well as uracil photoproducts [175–177]. Finally, U:G pairs have been shown to activate MutSα ATPase activity [178].

5. A Necessary Intermediate in Antibody Maturation: Genomic Uracil in Somatic Hypermutation and Class Switch Recombination

Uracil is enzymatically introduced into DNA for the antibody maturation processes somatic hypermutation (SHM) and class switch recombination (CSR). In this section, I will

first provide a brief overview on innate and acquired immunity and general B-cell maturation. I will then explain SHM and CSR, focusing on the role of genomic uracil.

5.1 Innate versus acquired immunity

The body is constantly exposed to many dangers from pathogenic microorganisms. To this end, the innate immune system works as a first-line defense to stop pathogens, identify them, and mount short and long term responses (reviewed in [179]). The innate immune system is composed of anatomical and physiological barriers (*e.g.* the skin, placenta, and respiratory, urinary, and gastrointestinal tracts), anti-microbial toxins or molecules (*e.g.* β-defensins, lysozymes, and APOBECs), and phagocytosis and the inflammatory response (by *e.g.* phagocytes, the complement system, and the activation of the adaptive immune system). The inflammatory response is initiated by the activation of non-self-recognizing receptors *e.g.* toll-like receptors (TLRs). TLRs recognize pathogen-associated molecular patterns like bacterial lipopolysaccharide (LPS, an endotoxin in bacterial cell walls), bacterial flagellin, and dsRNA and unmethylated CpG islands from viruses.

The acquired (or adaptive) immune system is activated by the innate immune system and mounts a targeted, long-term response with memory. The acquired immune system's response to infection can be broadly split into the cell-mediated or humoral responses, for which T-cells or B-cells are mainly responsible, respectively. The cell-mediated response is composed of T-helper cells, which help coordinate the activity of the immune system through macrophage antigen presentation and subsequent cytokine secretion and activation of B-cells and cytotoxic T-cells. The cell-mediated response also involves cytotoxic T-cells that can recognize antigens and kill other cells, and memory T-cells that retain antigen recognition over a long time frame. The humoral response is mediated by activated B-cells, which undergo affinity maturation and mature into antibody-secreting plasma cells or memory B-cells. Here, I will only focus on antibody production in B-cells.

5.2 B-cell maturation

B-cells develop from hematopoietic stem cells in the bone marrow (reviewed in [179]. In the bone marrow, stromal cells secrete growth factors for hematopoietic stem cell differentiation and antigen-independent V(D)J recombination takes place in immature B-cells. V(D)J recombination is the initial antigen-independent rearrangement of immunoglobulin genes necessary to produce a high number of unique antibodies from limited genetic material (reviewed in [180]). Briefly, the variable (V), diversity (D), and joining (J) segments of immunoglobulin genes are rearranged through the introduction of site-specific DNA DSBs. The recombination activating proteins recombination activating gene (RAG) 1

and RAG2 assemble a pair of dissimilar recombinational signal sequences (RSSs) into a synaptic complex and cleave the DNA strands. The resulting strands are reorganized and repaired by the non-homologous end-joining (NHEJ) pathway, and further genetic diversity may be achieved by random nucleotide insertions and template-independent DNA fill-in synthesis by POLμ and POLλ. Mature, naïve B-cells that express membrane bound IgM are then released from the bone marrow.

Upon infection, B-cells bind to antigens and are activated by T-cells in secondary lymphoid organs (spleen, lymph nodes, tonsils, and Peyer's patches) and are selected for immunoglobulin antigen binding affinity, which requires B-cell receptor- (BCR, membrane bound Ig), CD40 ligand- (CD40L) and growth-factor (*e.g.* B-cell activating factor, BAFF) stimulation [181–183]. A fraction of B-cells differentiate into centroblasts in the primary follicles of the lymphoid organs to form GCs [184,185]. Further B-cell differentiation occurs in GCs through SHM and CSR. SHM introduces mutations of the V regions of the immunoglobulin heavy (*IgH*) and constant (*IgC*) genes, thereby introducing another level of diversity to immunoglobulins (reviewed in [186,187]).

B-cells are selected for maturation into plasma cell and memory B-cell based on the specificity of their immunoglobulin binding to antigen. This is achieved through pro-proliferative and pro-apoptotic properties of GCs. Centroblasts lack expression of anti-apoptotic factors, allowing rapid apoptosis by default or in response to exogenous signals [188–190]. Indeed, isolated GC B-cells quickly undergo apoptosis *in vitro* unless rescued by anti-apoptotic cytokines [191–193]. Clonal selection of BCR-expressing cells then leads to the selection of B-cells expressing high-affinity antibodies for differentiation into antibody-secreting plasma cells or memory B-cells [194–196]. Together, SHM and V(D)J recombination are estimated to produce more than 10^9 different antibodies [197]. Some animals (*e.g.* cattle, pigs, sheep, and chickens) also employ a process called gene conversion, in which parts of "pseudo" V regions are transferred to rearranged V regions, but this will not be discussed in the thesis [198,199].

The constant (C) region of the IgG chain can be recombined by CSR from IgM or IgD to IgA, IgE, or IgG to have differing effector functions (reviewed in [200,201]). CSR occurs in switch (S) regions of the *IgH* gene, which are located upstream of each C region and recombine DNA to replace Cμ or Cδ regions with a downstream C region, thereby bringing the functional V(D)J region close to the downstream C region (reviewed in [202,203]). The altered C regions produce antibodies of different isotypes with different effector functions without altering antibody specificity [204,205].

B-cell differentiation into antigen-secreting plasma cells or memory B-cells occurs after successful response to antigen in the GC (reviewed in [206]). Plasma cells secrete soluble antibodies and memory B-cells are long-lived and can proliferate into antigen-secreting cells upon re-encountering an antigen. B-cell differentiation is determined by BCR signaling upon antigen stimulation and will not be further explained in the thesis [207–209]. The resulting antibodies from B-cells have three main functions: neutralizing pathogens by forming a coat around them, opsonizing pathogens (*i.e.* labeling them for phagocytocis), and activating the complement system.

5.3 Somatic hypermutation

In SHM, mutations are introduced into the V region of the *Ig* gene, leading to protein-level mutations and subsequent selection for proliferation based on Ig antigen binding affinity (figure 3). AID initiates SHM by deaminating cytosines, creating U:G mismatches [210–212]. Normal BER can be employed and a CTP re-inserted, leaving no mutation. C:G to T:A transition mutations can be generated by normal replication across U by POLδ or POLε [213,214]. *Ung*$^{-/-}$ mice exclusively develop these transitions in *Ig* V regions [215]. Alternatively, the U can be excised by UNG2, creating an AP site that can be processed in several ways. The AP site can be replicated over by the error-prone translesion synthesis (TLS) polymerase DNA protein REV1 (REV1) [216–220]. REV1 has been demonstrated to bypass polymerase-stalling AP sites and induce C:G to G:C transversion mutations [217,218,221]. In the absence of REV1, POLη can generate C:G to G:C transversions as well [222–224]. Finally, C:G to A:T transversions also require UNG2-induced AP sites, but the responsible TLS polymerase or polymerases are unknown [215,225]. APE2 also likely plays a large role in SHM. APE1 is down-regulated and APE2 up-regulated in GC B-cells [226]. APE2 has a functional PCNA-interacting domain and monoubiquitinylated PCNA at Lys164 has been shown to recruit POLη and REV1 [227,228].

Alternatively, U:G mismatches can be recognized by MutSα, activating EXO1 to create a ssDNA gap. The gaps are filled in by error-prone TLS polymerases POLη or POLζ [229–233]. These mutations are associated with A:T pairs 5′ of the U:G mismatch and POLη, MSH2/6, and EXO1-deficient cells have been shown to lack 90 % of A:T mutations found in WT cells [234–237].

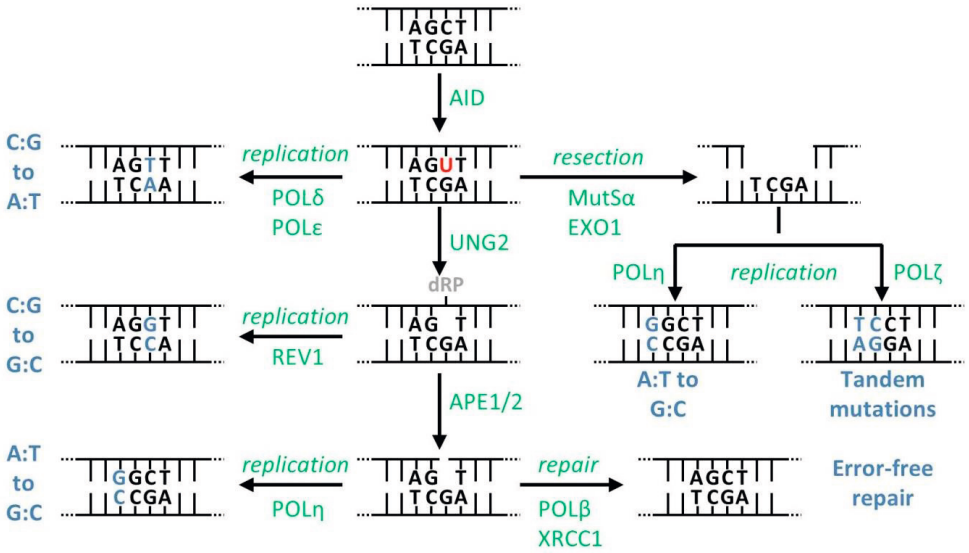


Figure 3: Somatic hypermutation. AID deaminates C to U. Normal replication yields a C:G to A:T mutation. U can be recognized by the MMR pathway, resulting in end-resection and fill-in by TLS POLη or POLζ and A:T to G:C or tandem mutations. UNG2 can excise U, leaving an AP-site that can be replicated over by TLS REV1 to yield a C:G to G:C mutation. The AP-site can also be faithfully repaired or replicated by TLS POLη to yield a A:T to G:C mutation 5′ of the U. Modified from [238].

5.4 Class switch recombination

CSR is the intrachromosomal rearrangement between S regions of *IgH* genes leading to the replacement of the Cμ locus with Cγ, Cε, or Cα loci and thus antibody isotype switching (figure 4) (reviewed in [203,239]). S regions are located upstream of all C genes except for Cδ and are 1-10 kbp in length [240]. Recombination occurs between DSBs introduced in the donor μ S region (Sμ) and a downstream acceptor region, though recombination can also occur with regions farther downstream [240,241]. S regions are G-rich and have a frequency of the preferred AID target motif WGCW (W=A/T) [242–244]. Like with SHM, CSR is induced by AID-mediated DNA deamination of cytosines in the S regions of *Ig* gene V regions [52,53,212,245–247]. CSR occurs in G1 phase during cell division and is linked to RNA transcription [248,249]: AID is regulated by cell division and deaminates ssDNA, and interaction with RNA polymerase II-associated exosome complex enhances AID recruitment to DNA [245,250,251].

AID-generated uracils are excised by UNG2. CSR is reduced by 95% in UNG-deficient mouse splenocytes and in humans with dysfunctional UNG, indicating that it is the

major glycosylase in CSR U excision [212,215,238,252–254]. SMUG1 can function as a weak backup and it has been shown that SMUG1 overexpression can partially rescue CSR in UNG-deficient mouse splenocytes and UNG/SMUG1 deficient splenocytes are devoid of any residual CSR activity [253,255]. Interestingly, SMUG1 overexpression in the presence of WT UNG decreases CSR, suggesting either a role for UNG in the recruitment of error-prone repair components or SMUG1 in the recruitment of error-free BER [255].

After uracil excision, APE1 and/or APE2 nick the DNA at the AP site. If the two nicks are in close enough proximity, a DSB is produced, which requires APE1 and APE2 [256]. S region DSBs are greatly reduced in *Apex1*$^{+/-}$*Apex2*$^{-/-}$ mouse splenocytes but only slightly reduced in *Apex1*$^{+/-}$ or *Ape2*$^{-/-}$, suggesting redundant roles for APE1 and APE2 in CSR [256]. Contrarily, another study showed that APE1-deficiency in a mouse B-cell lymphoma cell line (CH12F3) caused an 80% reduction in CSR while APE2-deficiency did not affect CSR [257,258]; however, CH12F3 cells have a demonstrably abnormal uracil profile in response to CSR stimulation compared to *ex vivo*-stimulated B-cells, so a fundamental difference in the biology of the cell line may explain this discrepancy [259]. Indeed, an interaction between APE2 and PCNA may promote error-prone repair and APE2's 3′ to 5′ exonuclease processivity is enhanced by PCNA [226,260].

If no DSBs are initially produced, POLβ replaces the nucleotide and the nick is ligated by LIGIII in complex with XRCC1. POLβ inhibits S region DSBs and CSR, suggesting that it may compete with DSB formation but is overcome by the amount of AID-induced SSBs [226,261]. Alternatively, the MMR pathway can convert a SSB into a DSB. MutSα binds to U:G mismatches in dsDNA and recruits MutLα and EXO1, which then initiates resection of the SSB 5′ to the mismatch [262,263]. PMS2 (part of MutLα dimer) also exhibits exonuclease activity and can create additional SSBs to provide more entry sites for EXO1 [160,264]. Combined with the SSB from UNG2 excision of U and APE1/2 nicking of the DNA strand, MMR therefore creates a DSB with a long ssDNA tail that results in a blunt DSB upon polymerization of the ssDNA with a DNA polymerase [226,239].

The blunt DSBs are joined by the NHEJ or alternative end-joining (A-EJ) pathways. NHEJ is initiated by the XRCC5/6 dimer (also known as Ku80/Ku70, respectively), which binds to DSBs and recruits the Artemis/DNA-PK catalytic subunit (DNA-PKcs) complex and other NHEJ proteins [265,266]. The two DNA ends bound to DNA-PKcs are joined together and processed by Artemis/DNA-PKcs to make them compatible for joining [267,268]. POLμ and POLλ bind to the XRCC5/6/DNA complex and fill in nucleotides at the strand breaks [269]. Finally, the XRCC4-like-factor (XLF)/XRCC4/LIGIV complex ligates across the

DNA ends [270–272]. A-EJ works as a backup DSB repair mechanism in the absence of NHEJ proteins and it has been shown that NHEJ-deficient cells retain significant CSR activity [273–278]. Moreover, A-EJ probably does not compete with NHEJ and the choice between the two is likely determined by the density of DNA lesions [279]. A-EJ involves larger microhomologies between the DSB junctions than NHEJ, but the pathway is not well understood and may even represent several pathways [280,281]. The consensus initiation step for A-EJ is 5′ to 3′ resection at the DSB to expose ssDNA for annealing to a homologous sequence [282]. DNA endonuclease RBBP8 (RBBP8, also called C-terminal-interacting protein, CtIP) is used for the DNA end-resection step [276,283]. Base-pairing then occurs at regions with complementary microhomology and the ends are joined by an undetermined ligase [281].

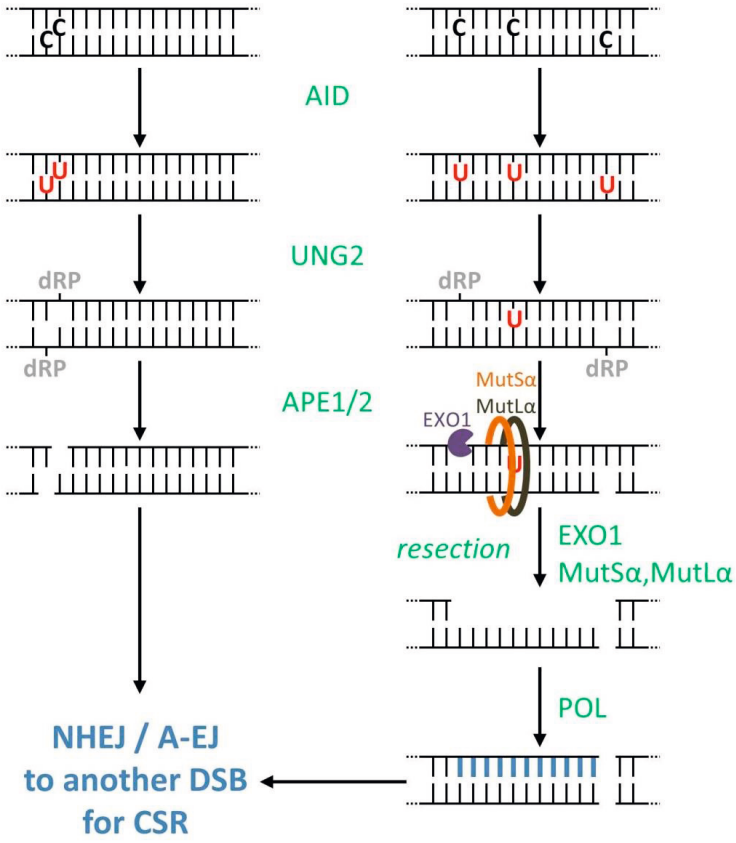


Figure 4: Class switch recombination. AID deaminates C to U in C loci of *Ig* S regions. The Us are excised by UNG2 and the AP-site nicked by APE1 or APE2, which may result in a DSB if two adjacent nicks occur in opposite strands. The MMR pathway can also process an AP-site and U in the strand opposite the APE1/2 nick, creating a blunt DSB. The DSBs are joined to other C loci by NHEJ or A-EJ. Modified from [223,238].

6. Host Defense by Uracilation of Exogenous Genomes: Genomic Uracil in Innate Immunity

APOBEC3A-G (A3A-A3G) constitute an innate barrier to retroviruses, endogenous retro-elements, and DNA viruses (reviewed in [284,285]). The biological function of APOBECs in this sense is the uracilation of exogenous DNA instead of cellular genomic DNA, so this section will only give a brief overview of A3 proteins in antiviral defense.

The restriction of human immunodeficiency virus (HIV) by A3G is well-characterized. A3G is incorporated into incoming virus particles in the cell through interaction with the nucleocapsid domain of the group specific antigen (Gag) protein in an

RNA-independent manner [286–292]. Then, A3G deaminates cytosines in the cDNA created by reverse transcription of viral RNA, the frequency of which can exceed 10 % of all cytosines, resulting in a staggering loss of genetic information and subsequent formation of defective virions during the subsequent replication cycle [48,293]. Most lentiviruses (including HIV) counteract A3G by expressing viral infectivity factor (Vif), which hijacks the E3 ubiquitin ligase Cullin-5 (CUL5) to target A3G for ubiquitinylation and proteosomal degradation [294–301]. The actual restriction of viral cDNA could be performed by UNG2 and APE1 processing of the uracils, but APOBEC-mediated viral restriction has also been shown to occur in the absence of UNG2 and SMUG1, so whether this actually occurs or to what extent remains in question [27,302–305].

Although not as well characterized as with HIV, A3 enzymes have been shown to restrict several other viruses. Human T-cell lymphotropic virus (HTLV) does not encode any Vif-like protein and cannot degrade A3 *in vitro* [306,307]. Despite this, there was no hypermutation in viral cDNA from HTLV-infected patients [306]. Two possibilities have been suggested to explain this discrepancy. First, the infrequent replication by reverse transcription in HTLV reduces the opportunities for A3 deamination to occur [308–311]. Second, elements in the C-terminus of the HTLV-1 nucleocapsid inhibit A3G packaging to the virus particle [312]. Hepatitis B virus (HBV) may also be susceptible to APOBEC deamination by A1, AID, and all A3 enzymes except A3D and A3E [313–319]. A3G restriction of HBV has been reported with a reduction of ~30-fold viral DNA in the presence of A3G [316,318,320]. Hepatitis C virus (HCV) replication can be inhibited by A3G *in vitro*, but no hypermutated sequences have been found, possibly because HCV is RNA-based during all phases of replication (*i.e.* no cDNA is produced) [321]. Nevertheless exogenous Vif-1 decreased A3G levels and increased the HCV replication rate *in vitro* [322]. Human papillomavirus has been shown to be susceptible to A3A, A3C, and A3H editing [323]. Finally, human herpesviruses (HHV) can be restricted by APOBECs with the identification of hypermutated viral DNA, and A3C overexpression has been shown to reduce viral load [314,324]. Although much work remains to be done in the characterization of A3 enzymes' (as well as other APOBECs') roles in viral restriction, it is clear that they are potent DNA damagers and therefore also potentially dangerous for the host genome.

7. Pathology of an Abundance or Shortage of Genomic Uracil

Given that work during the decade has clearly illustrated the role of genomic uracil as a necessary intermediate in immunity and a mutagenic DNA lesion, it is unsurprising that perturbations in the "uracilome" have pathological consequences (figure 5). In this section, I

will describe how a shortage of genomic uracil by AID deficiency leads to immunodeficiency and even autoimmunity. Then, I will discuss how an abundance of genomic uracil by uracil-DNA glycosylase deficiency, perturbations in pyrimidine biosynthesis, or dysfunctional enzymatic cytosine deamination can lead to both immunodeficiency and cancer.

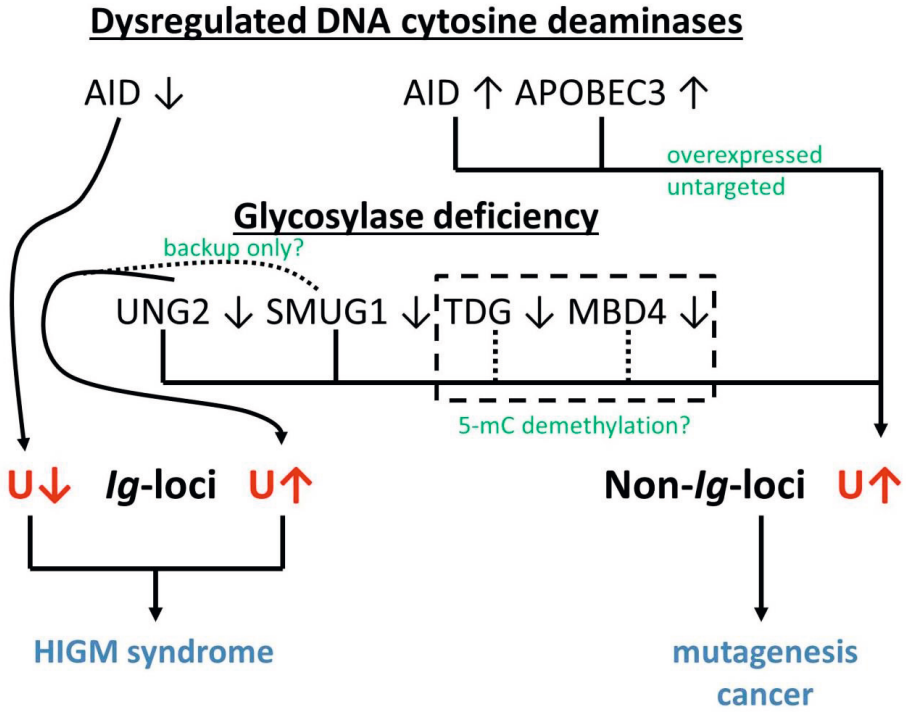


Figure 5: Overview of uracil-DNA glycosylase AID/APOBEC dysregulation, their impacts on genomic uracil levels, and their resulting pathogenicities.

7.1 Shortage of genomic uracil from AID deficiency

A shortage of genomic uracil by AID deficiency leads to immunodeficiency. Hyper-immunoglobulin M (HIGM) syndromes are a heterogeneous group of genetic disorders that result in defective CSR with or without defective SHM (reviewed in [325]). HIGM syndrome leads to bacterial infections of the respiratory and digestive tracts and lymphoid hyperplasia [325–328]. HIGM syndrome patients may also suffer from autoimmune disorders [328]. The most common forms of HIGM syndrome are a result of CD40 or CD40 ligand deficiencies [329], the lack of which precludes B-cell stimulation for CSR and SHM upon antigen recognition. Although very rare, AID deficiency results in HIGM with both defective CSR and SHM, usually in an autosomal recessive manner [52,53]. C-terminal mutations in *AID* have been reported to be autosomal dominant and impair CSR but not SHM [330,331]. AID contains a nuclear export signal (NES) in its C-terminus, so the autosomal dominance of

HIGM in patients with C-terminal mutated *AID* may suggest that either the inactivation of the NES leads to an accumulation of the mutant allele in the nucleus or AID acts in a multimerization or complex formation with not yet defined partners crucial for CSR [332–334]. Thus, in CD40(L)- or AID-deficient HIGH syndromes, a shortage of genomic uracil causes immune dysfunction resulting from a lack of high affinity antibodies.

7.2 Abundance of genomic uracil from deficient uracil repair or increased uracilation

An increased uracil burden is pathogenic. In this section, I will describe how UNG2 deficiency can lead to both HIGM and lymphomagenesis (7.2.1 and 7.2.2). Next, I will list the evidence linking the other uracil-DNA glycosylases to increased mutagenicity (7.2.3). Then, I will briefly explore how the alteration of the pyrimidine biosynthesis pathway by folate deficiency can lead to both an increased and reduced risk of cancer (7.2.4). Finally, I will outline recent advances showing how enzymatic deamination by AID/APOBECs has been clearly linked to a variety of cancers (7.2.5).

7.2.1 UNG2 deficiency: Hyper Immunoglobulin M Syndrome

A deficiency in DNA uracil repair by UNG can lead to both dysfunctional CSR and oncogenesis. Like those lacking AID, persons lacking functional UNG suffer from HIGM syndrome [327]. Imai *et al.* studied a small group of HIGM syndrome patients with mutations in the *UNG* gene and found them to have strongly impaired CSR and SHM skewed SHM with no quantitative deficiency [254]. The SHM-induced base insertion in these patients was skewed towards G-C instead of A-T, likely due to the absence of UNG, forcing replication over unrepaired U:G. Although this work only included three patients, the biological conclusions from the study fit data from studies performed on transgenic mice. $Ung^{-/-}$ mice have been shown to have greatly reduced CSR but competent SHM [215,335]. The little CSR activity remaining in these mice is attributed to uracil processing by SMUG1and MutSα; however, SMUG1 overexpression does not rescue CSR in $Ung^{-/-}$ mice, although it does increase CSR levels in $Ung^{-/-}Msh2^{-/-}$ mice to $Ung^{-/-}$ levels [210,336]. Furthermore, SMUG1 overexpression in UNG-competent mice leads to diminished CSR, suggesting that SMUG1 preferentially initiates the traditional error-free BER pathway, while UNG2 uracil excision leads to error-prone repair [255]. Thus, either an abundance of unrepaired uracil or its repair in an error-free manner are both deleterious to competent CSR and the immune response.

7.2.2 UNG2 deficiency: increased lymphomagenesis

UNG2 deficiency can also increase the global uracil load and lead to hyperplasia and lymphoma. $Ung^{-/-}$ mice show a 22-fold increase in B-cell lymphoma development over wild-type mice [337], which is linked to early-age lymphoid hyperplasia in splenic B-cells and

possibly an immunological imbalance characterized by differential interferon (IFN) γ, interleukin (IL) 6, and IL-2 levels [338]. Gene-specific mutation analysis in mice showed a 1.4-1.8 increase in the AID deamination targets and oncogenes B-cell lymphoma 6 protein (*Bcl-6*) and Myc proto-oncogene protein (*Myc*) [339]. Moreover, a high-AID-expressing lymph node tumor analyzed showed 3-fold increased mutation levels in both *Bcl-6* and *c-Myc* loci, but not the tumor suppressor gene *p53* that is not targeted by AID [339]. The C:G to T:A transition mutation frequency in $Ung^{-/-}$ mice was increased as well, which is consistent with the aberrant SHM caused by UNG-deficiency.

7.2.3 SMUG1, TDG, or MBD4 deficiencies: increased mutagenesis

Although considered a backup for UNG2 uracil excision in BER, SMUG1's role in uracil repair is not necessarily redundant. siRNA-mediated silencing of *Smug1* in mouse embryonic fibroblasts (MEFs) revealed a 2.4-fold increase in mutation frequency over wild-type MEFs, suggesting that UNG2 and SMUG1 are not fully complementary [340]. Furthermore, although $Ung^{-/-}Smug1^{-/-}$ mice will breed normally and remain healthy beyond 1 year of age, $Ung^{-/-}Smug1^{-/-}Msh2^{-/-}$ mice have greatly increased cancer predisposition and shortened lifespans, indicating when both base excision and mismatch repair pathways are defective, the mutagenic effects of spontaneous cytosine deamination are sufficient to increase cancer incidence without precluding mouse development [121].

MBD4 is important for mutation suppression. MBD4-deficient mice are fertile, develop normally, and have no increase in tumorigenesis, but show a two- to three-fold increase in C:G to T:A transition mutations at CpG sites in the small intestines [341,342]. Cancer-susceptible mice heterozygous for the Min allele in adenomatous polyposis coli (Apc^{min}) gene crossed with $Mbd4^{-/-}$ mice also showed an accelerated tumor formation with predominant CpG to TpG transition mutations as the *Apc* gene [341]. Furthermore, *MBD4* is mutated in 26-43 % of gastric, colorectal, endometrial, and pancreatic cancers that exhibit microsatellite instability, and the presence of an MBD4 mutant that binds 5-mCpG sites but lacks glycosylase activity has been shown to more than double the mutation frequency in colon cancer cells [343–347]. Thus, MBD4 repair of cytosine deamination at CpG sites is crucial to avoid mutagenesis and possibly cancer, although as with TDG it is unclear whether an abundance of uracil is the main cause.

Several studies have linked single nucleotide polymorphisms (SNPs) in uracil-DNA glycosylase genes with elevated uracil levels or increased cancer predisposition. Germline variations in *TDG* and *UNG* genes were linked to colon cancer predisposition, though their rarity is indicative of a limited role [348]. Furthermore, SNPs in *TDG* and *SMUG1* have been

associated with esophageal squamous cell carcinoma (ESCC), and several *SMUG1* SNPs modestly increase breast, bladder, and colon cancer risk [349–353]. Two SNPs in *SMUG1*and one in *UNG* were also shown to increase genomic blood uracil levels [354]. Finally, several SNPs in *MBD4* have been associated with increased risk of ESCC and lung, colon, and cervical cancers [355–360]. There are several conflicting reports regarding the risk of the uracil-DNA glycosylase SNPs and their associations, so I provided a more complete list thereof in Table 3. It is therefore likely that dysfunctional glycosylase activity plays at least a small role in oncogenesis, though the interplay between uracil repair-related genomic instability and other oncogenic factors is unclear.

Table 3: SNPs associated with uracil-DNA glycosylases [349–377].

Enzyme	SNP	Functional Consequence	Association	Reference
UNG	rs34259	intron variant	increase in uracil levels in DNA (whole blood)	Chanson (2009)
			no association with risk of breask cancer	Marian (2011)
			risk of rheumatoid arthritis	Lo (2012)
	rs246079	intron variant	risk of ESCC	Yin (2014)
			risk of lung cancer	Doherty (2013)
	rs3219218	intron variant	*no association* with risk of ESCC	Yin (2014)
	rs2337395	intron variant	risk of age-related machular degeneration	Blasiak (2012), Synowiec (2014)
	rs3890995	intron variant	disease progression in bladder cancer	Wei (2012)
	rs1018784	intron variant		
	rs3219266	intron variant	*no increase* in uracil levels in DNA (whole blood)	Chanson (2009)
	rs246085	intron variant		
SMUG1			increase in uracil levels in DNA (whole blood)	Chanson (2009)
	rs2029166	intron variant	risk of ESCC	Li (2013)
			risk of breast cancer	Marian (2011)
	rs7296239	intron variant	increase in uracil levels in DNA (whole blood)	Chanson (2009)
			risk of breast cancer	Marian (2011)
	rs2029167	intron variant	risk of bladder cancer	Xie (2015)
	rs2233921	intron variant	colorectal cancer survival	Pardini (2013)
	rs3087404	intron variant	risk of age-related machular degeneration	Blasiak (2012), Synowiec (2014)
	rs3136392	intron variant	*no increase* in uracil levels in DNA (whole blood)	Chanson (2009)
	rs1994356	intron variant		
TDG			genomic instability	Sjolund (2014)
			susceptibility to chromosomal damage after vinyl chloride exposure	Wen-Bin (2009)
	rs4135113	G199S	risk of ESCC	Yang (2014)
			no association with risk of ESCC	Li (2013)
			no association with risk of colon cancer	Curtin (2011), Lin (2014)
			no association with lung cancer	Krześniak (2004)
	rs4135054	intron variant	risk of ESCC	Li (2013)
			no increase in uracil levels in DNA (whole blood)	Chanson (2009)
	rs2700505	intron variant	risk of bladder cancer	Xie (2015)
	rs4135150	intron variant	risk of non-melanoma skin cancer	Ruczinski (2012)
			no increase in uracil levels in DNA (whole blood)	Chanson (2009)
	rs2888805	V367M	risk of non-melanoma skin cancer	Ruczinski (2012)
	rs167715	intron variant	risk of treatment-related mortality in posthematopoetic cell transplant	Thyagarajan (2010)
	rs2374327	intron variant		
	rs703657	intron variant		
	rs172814	intron variant		
	rs4135063	intron variant	*no increase* in uracil levels in DNA (whole blood)	Chanson (2009)
	rs2723877	intron variant		
	rs1086115	intron variant		
	rs1866074	intron variant		
MBD4			DNA repair capacity	Allione (2013)
	rs2005618	intron variant	risk of rheumatoid arthritis	Huang (2012)
			no association with risk of rheumatoid arthritis	Cheng (2012)
			no association with risk of ESCC	Yin (2014)
	rs10342	A273T	loss of protein-DNA interaction (*in silico*)	Allione (2013)
	rs3138373	intron variant	*no association* with risk of ESCC	Yin (2014)
			poor prognosis in non-small cell lung cancer	Dong (2012)
			risk of lung cancer	Shin (2006), Miao (2008)
	rs140693	E346K	risk of cervical cancer	Xiong (2012)
			risk of colon cancer	Song (2009)
			no association with risk of rheumatoid arthritis	Cheng (2012)
			no association with risk of immune thrombocytopenic purpura	Zhao (2010)
			risk of ESCC	Yin (2014)
	rs3138355	intron variant	risk of rheumatoid arthritis	Huang (2012)
			*no increase i*n uracil levels in DNA (whole blood)	Chanson (2009)
	rs2311394	intron variant	*no association* with risk of rheumatoid arthritis	Cheng (2012)
			*no increase i*n uracil levels in DNA (whole blood)	Chanson (2009)
	rs1047043	intron variant		
	rs4128190	intron variant	*no increase i*n uracil levels in DNA (whole blood)	Chanson (2009)
	rs140696	synonymous		

7.2.4 Folate deficiency: contradicting evidence for oncogenesis

Whether an increased uracil load by thymidylate synthase inhibition or dysfunction contributes to cancer risk is under contention. Folate is a necessary co-factor for TS activity and folic acid deficiency or folate antagonists such as methotrexate impair its activity [378,379]. TS inhibition is cytotoxic to some cancers, and one hypothesis for TS inhibitor-induced cytotoxicity is that futile cycles of glycosylase uracil excision, BER, and further dUTP misincorporation induce DNA fragmentation [13,380]. Were this the case, TS inhibitor sensitivity would increase with increased UNG expression, but this is not the case: cancer subtypes resistant to the TS inhibitor pemetrexed exhibit higher UNG expression than pemetrexed-sensitive subtypes, and UNG overexpression has been ineffective at sensitizing cells to other TS inhibitors [381,382]. Instead, UNG loss has been shown to enhance DSBs in colon cancer cell lines upon pemetrexed treatment, so it has been suggested that pemetrexed-induced uracil misincorporation is genotoxic by uracil accumulation near replication origins, replication fork stalling, fork collapse, DSB break formation, and cell death [383].

However, the clinical picture is less straightforward. Although folate depletion (leading to lowered TS activity) appears to increase carcinogenesis [384,385], supplemental folate intake and high folic acid levels have also been shown to increase cancer risk [386–388]. Another study showed that a low-folate diet did not increase tumor development in *Ung*$^{-/-}$ mice [389]. Finally, it was shown that increasing dietary folate both increased and decreased genomic uracil in mice colon and livers, respectively [390]. Thus, it remains unclear whether and to what extent genomic uracil is modulated by thymidylate synthase.

7.2.5 Enzymatic cytosine deamination: increased widespread carcinogenesis

DNA uracil abundance by enzymatic DNA cytosine deamination can also lead to oncogenesis. A general hallmark of many B-cell lymphomas is a translocation between the *Ig*-loci and a proto-oncogene, such as the *BCL1/Ig* translocation in Mantle zone lymphoma, *BCL2/Ig* translocation in follicular lymphoma, and *MYC/Ig* translocation in Burkitt's lymphoma [391,392]. The role of AID in these translocations has been well established: removal of AID has been shown to decrease the frequency of *MYC/IgH* translocations [393]. These translocations occur most frequently at transcription start sites, which is consistent with the accepted model of transcription-coupled deamination of single stranded DNA [394]. AID also shifts the incidence from pre-B-cell to more mature B-cell lymphomas in MYC-overexpressing mice and shifts the preference from Mantle cell lymphoma to diffuse large B-cell lymphoma in BCL6-overexpressing mice [395,396]. Furthermore, ablation of AID activity by knock-out in mice results in the accumulation of significantly fewer mutations

linked to B-cell lymophomagenesis [397]. These data suggest that AID deamination is lymphomagenic. In addition, AID is highly expressed in several lymphomas, so deamination may be important for therapy resistance and disease progression [398–400]. High AID levels have also been correlated with high genomic uracil levels in lymphoma cell lines and primary B-cell leukemia and lymphoma cells [259].

Enzymatic DNA cytosine deamination has also been linked to non-B-cell cancers. A 2012 study by Nik-Zainal *et al.* sequenced the genomes of 21 breast cancer tumor samples and discovered unique mutational signatures in clusters called "*kataegis*" (Greek for shower or thunderstorm) [401]. The *kataegis* clusters were attributed to APOBEC cytosine deamination and subsequent work showed that the mutational signatures observed were indeed consistent with the action of these enzymes [402]. Another study found that A3B was up-regulated in the majority of breast cancer tumor samples and cell lines analyzed and the tumors and cell lines with high A3B levels contained twice as many somatic hypermutations as those with low A3B levels [403]. The same study also showed that genomic uracil levels were decreased in a high-A3B-expressing cell line upon small hairpin RNA (shRNA) knock-down of A3B. Subsequent work has shown that: increased AID/APOBEC expression levels increase mutational frequency and *kataegis* in yeast in an UNG-dependent fashion [404]; APOBEC mutational signatures and *kataegis* are present in most cancers, including breast pancreas, lung, and liver cancer, medulloblastomas, chronic lymphocytic leukemia, B-cell lymphomas, and acute lymphocytic leukemia [405]; and *kataegis* clusters are present in multiple myeloma, ovarian cancer, osteosarcoma, and renal cell carcinoma [406–409]. Thus, an increase in enzymatic DNA uracilation has a clear link to cancer.

8. AIMS OF THE THESIS

The overarching goal of the articles presented in this thesis and my doctoral studies in general was to better understand the link between genomic uracil's role in cancer and adaptive immunity. Thus, we explored the following topics:

1. How does uracil repair differ between man and mouse? An overwhelming amount of mechanistic insights are gained from both *in vitro* and *in vivo* mouse studies, so it was crucial to identify what drawbacks a mouse model may have.
2. We aimed to establish a method to measure absolute global genomic uracil levels. A wide variety of methods have been employed to measure genomic uracil, but they have yielded an equally broad distribution of results, suggesting possible technical shortcomings in the methods. A method to accurately determine uracil levels was therefore crucial to make conclusions regarding its regulation.
3. What are the relative contributions of UNG and SMUG1 to general uracil repair in various tissues? Together, the two enzymes supposedly maintain a low genomic uracil burden, so understanding their relative contributions to uracil excision would help clarify their respective roles.
4. Do AID/APOBEC expression levels correlate with genomic uracil levels? Given the recent discovery of APOBEC mutational signatures in cancers, it was important to determine whether the global uracil burden was substantially increased with high AID/APOBEC expression or whether the uracils were transient, *i.e.* they were repaired too soon after introduction to measure.

9. SUMMARY OF RESULTS

9.1. Article I - Uracil-DNA glycosylase in base excision repair and adaptive immunity: species differences between man and mouse.

May 2011, *Journal of Biological Chemistry*

Doseth B, Visnes T, Wallenius A, Ericsson I, Sarno A, Pettersen HS, Flatberg A, Catterall T, Slupphaug G, Krokan HE, Kavli B

Mouse models are very commonly used in molecular biology, including for the elucidation of BER and SHM and CSR. In this article, we examined the species differences in the initiation of uracil repair between mouse and man to better understand the applicability of results derived from mouse models.

We measured the uracil excision activity using a standard oligonucleotide-based nicking assay in which a 19-mer oligonucleotide containing one uracil is removed by the uracil-DNA glycosylases in sample cell extracts and the strand cleaved. We used a variety of normal, embryonic, and cancer human and mouse cell lines and found that the mean human cell line uracil excision activity was nine to ten-fold higher than mouse using substrates with both U:A and U:G. Then, we measured complete BER in the cell lines using covalently closed circular DNA (cccDNA) substrates with a single U:A or U:G incubated with cell extracts and radiolabelled nucleotides that can be used to estimate repair. Human extracts showed higher U:A repair efficiency than mouse extracts, but there was no species difference in U:G repair. This suggests distinct U:A versus U:G BER mechanisms in man and mouse.

Next, we examined the relative contributions of UNG and SMUG1 in man and mouse by selectively inhibiting each enzyme. UNG was inhibited using the irreversible inhibitor uracil-DNA glycosylase inhibitor (Ugi) from a *B. subtilis* bacteriophage, and SMUG1 was inhibited with a neutralizing SMUG1 IgG [410]. U:G lesions had previously been reported to be mostly repaired by UNG2 in humans and SMUG1 in mice [100,411]. Inhibiting both UNG and SMUG1 completely abolished measurable uracil excision activity in all samples, suggesting a minor role for TDG and MBD4 in uracil repair; however, the enzymes should not be ruled out because the reaction conditions may have disfavored them and did not include APE1, which increases the turnover of the enzymes by releasing them from the AP site. UNG showed a six-fold higher uracil excision activity in human extracts than in mouse. Interestingly, SMUG1 neutralization increased U:G activity in the samples, which could be explained by SMUG1 having a high affinity for U:G but a relatively lower turnover than UNG. Therefore, SMUG1 can compete with UNG for the limited U:G substrate when uninhibited. Contrastingly, SMUG1 exhibited an 8-fold higher uracil excision activity in

mouse than in man. Furthermore, about 50 % of the total excision activity in mouse samples could be attributed to SMUG1, compared to only 1 % in human samples. Thus, UNG and SMUG1 have similar relative contributions in mouse and human cells, but UNG is clearly the predominant glycosylase in human cells.

We then quantified the relative abundances of UNG, SMUG1, and TDG in the cell lines to determine whether protein levels could explain the relative contributions of the enzymes to uracil excision activity. Due to their low abundance, all proteins except for human UNG required immunoprecipitation prior to western blot analysis. UNG was ~20-fold higher in human than in mouse cell lines, indicating that the higher relative UNG activity was a result of a higher abundance of the enzyme and not necessarily a more active enzyme. The large difference may be explained by alternative transcriptional regulation of UNG in mouse and man. Indeed, there is limited homology between human and mouse UNG promoter sequences, although many key transcription factor binding sites are conserved [412]. Also, human UNG2 is regulated by stepwise phosphorylations, but one phosphorylation site is not conserved in mouse UNG2, suggesting a less stringent regulation of the protein in mouse cells [107]. There was no significant difference in SMUG1 levels between human and mouse cell lines, and TDG was 3.4-fold higher in mouse cells. Thus, the reliance on SMUG1 in mouse cell lines was most likely a result of a lower UNG abundance.

Finally, we measured the contributions of UNG and SMUG1 to uracil excision in CSR. To this end, we stimulated *ex vivo* wild type and $Ung^{-/-}$ splenic B-cells from mice with LPS and IL-4 to induce *in vitro* CSR. As expected, there was no UNG activity in the $Ung^{-/-}$ B-cells, as well as no change in SMUG1 activity with UNG-deficiency. All excision activity was again abolished when inhibiting both UNG and SMUG1, suggesting that TDG and MBD4 are unlikely to play a role in CSR uracil excision. In contrast to a previous report, SMUG1 activity was not down-regulated upon stimulation, but rather it was increased by 40 % [255]; however, SMUG1 activity does decrease with stimulation if normalized to total protein in the extract instead of to the number of cells used. Normalization to total protein can be disadvantageous because cell morphology changes after stimulation, a difference in cytoplasmic proteins that can skew normalization. Thus, SMUG1 may also play a role in backup uracil excision during CSR, although it may recruit error-free BER proteins that inhibit CSR *in vivo*.

CSR is completely abolished in humans lacking functional UNG, but only 95 % reduced in mice. MMR plays a role in CSR, but it also depends on uracil processing by a glycosylase and is therefore unlikely the reason for why UNG deficiency has a more

pronounced effect in mice. SMUG1 overexpression partially rescues CSR in *Ung*$^{-/-}$ mice and *Ung*$^{-/-}$*Smug1*$^{-/-}$ mice are completely devoid of CSR [253,255]. Additionally, uracil in ssDNA at AID hotspots is 200-fold more favored by human UNG2 than human SMUG1 and SMUG1 prefers U:G in a dsDNA context. Finally, SMUG1 activity represented a larger portion of the uracil excision capacity in mouse cell lines and human cells expressed much more UNG than mouse cells, so mouse cells may rely more heavily on SMUG1 than on UNG2. Thus, SMUG1 functions as a backup for UNG2 in mouse CSR, but may not do so to the same extent in human cells.

In conclusion, we identified key differences between human and mouse cells suggesting that though UNG2 is the major uracil-DNA glycosylase in both organisms, SMUG1 likely plays a larger role in mouse cells. Care should therefore be taken before using conclusions derived using mouse models to elucidate human biology.

9.2. Article II - A robust, sensitive assay for genomic uracil determination by LC/MS/MS reveals lower levels than previously reported.

September 2013, *DNA Repair*

Galashevskaya A*, Sarno A*, Vågbø CB, Aas PA, Hagen L, Slupphaug G, Krokan HE

*(shared first authorship)

Uracil's roles in adaptive immunity and possible carcinogenicity underscore the need for a reliable assay to quantify it. Several groups have employed a variety of methods to assay whole-genome uracil, but the resulting broad variation in their results suggest that the methods contained technical shortcomings despite the heterogeneity of the samples measured. In this article, we set out to both identify and ameliorate any potential shortcomings in LC/MS-based uracil quantification.

We decided to measure dUrd by enzymatic hydrolysis of DNA to dNs instead of Ura by UNG excision for several reasons. First, dUrd contains a labile *N*-glycosylic bond between Ura and the deoxyribose moiety that requires substantially less energy to fragment than the heterocyclic Ura bonds, and less collision energy often tends to give lower background signals. Thus, measurement of dUrd by multiple reaction monitoring (MRM) on a triple-quadrupole mass spectrometer yielded a lower background than Ura. Second, the extent of Ura excision by UNG cannot easily be confirmed. One group found that restriction digestion of genomic DNA was essential to ensure complete Ura excision, presumably because the enzyme could more easily scan the shorter DNA fragments [413]. Conversely, the extent to which DNA is hydrolyzed can be easily measured by measuring dAdo/Cyd/Guo/Thd.

Naturally occurring ^{13}C-dCyd is isobaric with and yields an identical mass transition to dUrd, so they are indistinguishable by MRM. Given that dCyd is five to six orders of magnitude more abundant than dUrd, the small percentage of ^{13}C-dCyd presented completely obfuscated the dUrd peak using a normal reverse phase chromatography column, despite good chromatographic separation between the two nucleosides. We avoided this problem by using a reverse phase column with embedded weak ion-pairing groups (PrimeSep200) with which dUrd elutes before dCyd, avoiding obfuscation. We could not directly measure dUrd from hydrolyzed DNA using HPLC with a PrimeSep200 column directly coupled a MS/MS because dUrd elutes too near the void volume of the column, where ion suppression interferes with the signal. So, we set up a precursory HPLC step to separate dUrd from both dCyd and ion-suppressing contaminants. This had the added advantage of allowing direct quantification of total dNs by HPLC-UV. We found that measuring dUrd in this way gave a mean accuracy of 94.3%, intra- and inter-day CV values of 9.7 % and 10%, respectively, and a lower limit of quantification of 5 fmol dUrd.

We found that dUrd can be largely overestimated by co-purification of dUMP with DNA. Co-purified dUMP would then be dephosphorylated to dUrd during DNA hydrolysis. Alternatively, dCMP could be co-purified and deaminated. Note that this drawback is not present when measuring UNG-excised Ura. To eliminate co-purified dUMP, we pre-treated DNA samples with phosphatase to dephosphorylate dUMP and then precipitated the DNA. We hypothesized that dUrd would not carry over after the precipitation. Indeed, we found that up to 98% of the dUrd measured was removed after phosphatase treatment, confirming that dUMP (or dCMP) had been co-purified with DNA.

The rate of dCyd deamination is determined by temperature, pH, and DNA strandedness [17,19]. Thus, we wanted to avoid *in vitro* dCyd deamination by avoiding high heat and basic or acidic pH. Our final hydrolysis reaction was performed at 37 °C for 50 min and lay within pH 6 to 7.6. Other methods denature DNA by heating to 95 °C before a 3 h to 6 h DNA hydrolysis [24,25,414]. We found that heating DNA to 95 °C for 5 min resulted in a 1.7-fold increase in genomic dUrd. Furthermore, the hydrolysis step introduced 4.805×10^{-3} dUrd per 10^6 bp per min. We therefore included a control sample in our assay that has been deuracilated with UNG to measure the amount of *in vitro* generated dUrd. Subtracting the deuracilated control from the actual samples yields an accurate estimation of the actual genomic dUrd content in a sample.

Next, we aimed to validate the assay. We first compared dUrd content from DNA isolated using a spin column kit and phenol:chloroform:isoamyl alcohol, and found no

difference between the isolation methods. Next, we compared our assay to LC/MS/MS analysis of UNG-excised Ura (similar to [415]) and found a significant lack of reproducibility with the Ura excision method. Finally, we compared DNA from UNG2 proficient and deficient lymphoblastoid cell lines as well as *Ung*$^{+/+}$ and *Ung*$^{-/-}$ MEFs and found that UNG deficiency leads to a clear increase in genomic uracil levels.

From these results, we proposed that our method addresses the major sources of error in genomic uracil measurement.

9.3. Article III - UNG and SMUG1 efficiently complement each other in removing genomic uracil from mouse organs.

Manuscript

Sarno A*, Alsøe L*, Galashevskaya A, Tekin NB, Jobert L, SenGupta T, Carracedo S, Krokan HE, Nilsen H *(shared first authorship)

It is widely accepted that UNG2 is the primary uracil-DNA glycosylase with SMUG1 acting as its immediate backup. In paper I, we confirmed that this is indeed the case, although SMUG1 likely plays a larger role in mouse cells. In this article, we aimed to further clarify the respective roles of the two glycosylases, focusing on their effect on the genomic uracil load. Thus, we used *Ung*$^{-/-}$*Smug1*$^{-/-}$ mice to elucidate the relative roles of UNG and SMUG1.

First, we generated *Smug1*$^{-/-}$ mice and characterized them. The homozygous knockout mice showed no residual SMUG1 expression in the spleen by mRNA or western blot and the heterozygous mice had 50 % SMUG1 compared to wild type mice. Moreover, the mice were fertile and the *Smug1* knock-out gene was inherited at Mendelian ratios. SMUG1-deficiency led to no obvious pathological abnormalities in appearance or organ lesions. Blood cell count was also normal in these mice, but they exhibited a 20 % reduction in lymphocytes. We then crossed the *Smug*$^{-/-}$ mice with a previously-established *Ung*$^{-/-}$ strain, which also had no gross abnormalities except for CSR deficiency and a high incidence of lymphoma late in life [215,337]. The *Ung*$^{-/-}$*Smug1*$^{-/-}$ mice also showed no gross morphological abnormalities. Thus, we concluded that combined UNG/SMUG1 deficiency is well-tolerated in mice, at least in the time period examined here.

To confirm that SMUG1 is the major enzyme responsible for 5-hmU repair, we measured 5-hmU levels by LC/MS/MS and 5-hmU excision activity using the oligo nicking assay described in article I. Expectedly, there was no 5-hmU excision activity in *Smug*$^{-/-}$ MEFs. The wild type MEF and organ extracts showed a preference towards 5-hmU:G over 5-hmU:A and no activity on single-stranded 5-hmU substrate. The 5-hmU excision activities were almost identical to the uracil excision activities, indicating that SMUG1 excises the two

nucleosides at the same rate. 5-hmU levels were lowest in wild type organs and MEFs and increased in *Smug1*$^{+/-}$ and further in *Smug1*$^{-/-}$ organs, suggesting that a single copy of the *Smug1* gene (and subsequent 50 % expression relative to wild type) is not sufficient to maintain basal levels of 5-hmU. We measured 1.4- to 3.8-fold increases in 5-hmU over wild type in SMUG1-deficient organs. Furthermore, there was a slight increase in 5-hmU in MEFs after 22 passages, but it was independent of SMUG1 status. Lastly, there was no significant correlation between 5-hmU excision activity and 5-hmU levels.

We then explored the effect of SMUG1-deficiency on genomic uracil levels and uracil excision activity. There was a higher variation in uracil excision activity between tissue types than there was with 5hmU excision, possibly because of the contribution from UNG. The spleen and heart displayed the highest activities with U:G and ssU substrates, which also had similar activity profiles between organs. The only organ extract that had a different relative activity between U:G and ssU was the brain, the uracil excision activity of which was reduced using both U:G and U:A substrates. The reduction of uracil excision activity with SMUG1 deficiency in the brain combined with the low excision activity observed with ssU substrate regardless of SMUG1 status strongly suggests that SMUG1 is the major uracil-DNA glycosylase in the brain.

Finally, we measured the effect on uracil excision activity and genomic uracil accumulation on *Ung*$^{-/-}$ and *Ung*$^{-/-}$*Smug1*$^{-/-}$ mice. For the activity assays, we used wild type organ extracts pre-incubated with Ugi to simulate *Ung*$^{-/-}$ mice and found that uracil excision was completely abolished in all organ extracts with the ssU substrate and substantially reduced in all extracts save the brain with U:A and U:G substrates. The lack of any effect on uracil excision activity by UNG deficiency further indicates that SMUG1 is the major uracil-DNA glycosylase in the brain. Expectedly, there was no appreciable activity in any extract from *Ung*$^{-/-}$*Smug1*$^{-/-}$ mice, regardless of substrate. Unlike SMUG1, UNG-deficiency resulted in a two- to three-fold increase in genomic uracil levels over wild type in mouse organs, although they did not correlate with uracil excision activity. Strikingly, the *Ung*$^{-/-}$*Smug1*$^{-/-}$ exhibited a three- to 20-fold increase in genomic uracil levels over wild type. The heart, skeletal muscle, kidney, and lung samples showed an increase of 20- to 30-fold in genomic uracil over wild type and the liver a greater than 80-fold increase. The spleen and brain only exhibited three- and four-fold increases in genomic uracil, respectively. The synergistic increase in genomic uracil levels in *Ung*$^{-/-}$*Smug1*$^{-/-}$ mice demonstrates that SMUG1 can maintain genomic uracil at near wild type levels in the absence of UNG.

The spleen's relatively modest increase in genomic uracil was especially interesting given its high proliferative capacity. dUTP misincorporation has the potential to be a much more substantial source of genomic uracil compared to cytosine deamination, so more highly proliferative cells should accumulate more uracil than organs with lower proliferation. We did not observe such a clear correlation, although we only tested DNA isolated from organs, which have a heterogeneous cell composition. Although tempting to ascribe the large increase in genomic uracil to either dUTP misincorporation or cytosine deamination, more work needs to be performed to elucidate the relative contributions of the two mechanisms. In addition to proliferation rates, dUTP/dTTP ratios may vary between tissue and cell types, which would impact the dUTP misincorporation rate. Furthermore, ion levels and pH that can alter the spontaneous deamination rate are also likely to be tissue-dependent. TDG and MBD4 may also have contributed to uracil repair, despite the fact that no uracil excision activity was observed in *Ung*$^{-/-}$*Smug1*$^{-/-}$ organ extracts. The activity assay itself may not have been optimized for TDG and MBD4, as discussed in article I.

Thus, the main conclusion from this article is that while UNG is indeed the major glycosylase responsible for uracil repair in mice, SMUG1 alone can maintain genomic uracil at physiologically normal levels and is responsible for the majority of uracil excision in the brain. The relative contributions of dUTP misincorporation versus cytosine deamination remains to be explored, as well as to what extent SMUG1 can compensate for UNG in humans.

9.4. Article IV - AID expression in B-cell lymphomas causes accumulation of genomic uracil and a distinct AID mutational signature.

November 2014, *DNA Repair*

Pettersen HS, Galashevskaya A, Doseth B, Sousa MML, Sarno A, Visnes T, Aas PA, Liabakk NB, Slupphaug G, Sætrom P, Kavli B, Krokan HE

AID and APOBECs are associated with several mutational signatures in *kataegis* hotspots found in a wide variety of cancers (section 7.2.5). Although the presence of these mutational signatures implies at least the transient presence of genomic uracil, only one study had measured genomic uracil levels in APOBEC-expressing cells and they only did so to demonstrate the effect of shRNA knock-down of A3B [416]. In this article, we aimed to test whether there was a correlation between AID, APOBEC, and uracil-DNA glycosylase levels and genomic uracil. Note that an article containing very similar experiments and results was published the week we submitted our manuscript [259].

To find a correlation between AIC/APOBEC levels and genomic uracil, we measured genomic uracil by LC/MS/MS and AID/APOBEC expression by qPCR in 17 cancer cell lines for. There was a staggering 72-fold variation in genomic uracil levels between the cell lines, and we observed that uracil levels were higher in lymphoma cell lines (4.4- and 18-fold higher in lymphoma cell lines than in non-lymphoma cell lines and primary B-cells from healthy donors, respectively). Of all the DNA cytosine deaminases, AID showed the best correlation with genomic uracil ($R^2 = 0.70$). We confirmed the correlation with AID levels measured by western blot ($R^2 = 0.95$) and mass spectrometry ($R^2 = 0.65$).

Next, we tested whether AID attenuation by CSR stimulation, exogenous AID expression, or shRNA knockdown could alter the genomic uracil load in B-cell lines. AID overexpression in the mouse B-cell lymphoma cell line CH12F3 increased genomic uracil six-fold. Stimulation of the CH12F3 line to undergo CSR induced AID expression and increased genomic uracil four-fold 48 h after stimulation. The increase in genomic uracil post-stimulation was probably not due to an increase in dUTP misincorporation because CH12F3 cells proliferate more slowly upon stimulation. Thus, unless there was a large attenuation in the dUTP/dTTP ratio, the slower proliferation of the unstimulated cells would decrease their dUTP misincorporation relative to the stimulated cells. We also knocked down AID by shRNA in a cell line with high constitutive AID expression, reducing AID expression by 60 % and correspondingly reducing genomic uracil levels by 38 %. Thus, we showed that endogenous and exogenous AID expression increase genomic uracil levels *in cellulo*.

We explored uracil repair to determine to what extent it is responsible for the high genomic uracil levels in AID-expressing cell lines. Uracil excision activity assays (mean of U:G and U:A) had a statistically significant but weak inverse correlation ($R^2 = 0.52$) with genomic uracil. We further measured the four uracil-DNA glycosylases by mass spectrometry and found weak inverse correlations between UNG and SMUG1 levels and genomic uracil ($R^2 = 0.42$ and 0.28, respectively). These inverse correlations suggest that UNG and SMUG1 play a minor but significant role in the accumulation of AID-induced genomic uracil. The correlations were likely not higher because genomic uracil represents a balance of uracil introduction and repair. Uracilation in these cells seems to mostly be a product of AID deamination, whereas uracil repair regulated by a variety of factors including but not limited to uracil-DNA glycosylases, some of which are likely dysfunctional in cancer cell lines.

A method to distinguish between genomic uracil in U:A versus U:G contexts would be the best way to exclude dUTP misincorporation as a source of uracil in these cell lines, but no such method yet exists. As an alternative, we correlated cell doubling time with genomic

uracil and found that the lymphoma cell lines showed no correlation and the non-lymphoma cell lines had an inverse correlation between genomic uracil and cell doubling time. This suggests that dUTP minsincorporation may affect the genomic uracil load in cells with low AID expression, but has little to no impact in cells with high AID expression. In addition, we measured *in vitro* U:G BER activity on synchronized HeLa cells. We found that UNG was the major contributor to BER G1/S and G2/G1, though to a lesser extent and only 1.5-fold more than TDG. SMUG1 exhibited a small contribution to BER during all cell cycle phases. Thus, TDG and SMUG1 may have roles in BER of U:G mismatches in G1- and to a lesser extent in S. The relative roles of the glycosylases is likely similar in other cell lines, although this has not been demonstrated.

Alexandrov *et al.* analyzed sequencing data in from 7,042 cancers and found A3 *kataegis*-localized mutational signatures in acute lymphoblastic leukemia (ALL), lung adenocarcinomas, and breast, pancreas, and liver cancers [405]. We hypothesized that there may be different *kataegis* mutational signatures in B-cell lymphomas and chronic lymphocytic leukemia (CLL), so we re-analyzed the data accordingly. We found that B-cell lymphomas and CLL have an AID-specific mutational signature in *kataegis* regions (AG**C**T instead of T**C**A/T for other cancer types with *kataegis* clusters). This finding strongly links AID with genomic instabilities in B-cell malignancies.

In this article, we demonstrated a clear link between AID expression and genomic uracil as well as identified an AID mutational signature in CLL and B-cell lymphomas. The work we presented provides a strong piece of evidence to the theory that enzymatic DNA cytosine deamination is a major contributor to genomic instability in cancer.

10. DISCUSSION

10.1 SMUG1 plays a larger role in uracil repair in mice than in humans

There are significant differences in uracil processing between mouse and man that should be recognized before any conclusions are made using mouse models. Nevertheless, the overall uracil excision mechanisms between the two species seem to be similar enough for significant insight to be gained by using mouse models. The major difference between man and mouse seems to be the extent to which SMUG1 serves as a backup uracil-DNA glycosylase in mouse. UNG2 levels are lower in mice in general, so SMUG1 may act more as a partner with a large degree of overlapping function than a backup in mice. This is in part evidenced by the residual CSR activity in *Ung*$^{-/-}$ mice (5 % of wild type) that is much less pronounced in UNG-deficient humans [215]. Furthermore, we showed in article III that SMUG1 alone is sufficient to maintain physiologically low genomic uracil levels in mice. Thus, UNG deficiency, dysfunction, or deregulation may play a larger role in humans by inducing a more pronounced increase in the genomic uracil burden and therefore increasing overall genomic instability.

We showed in article III that either UNG or SMUG1 are sufficient to maintain genomic uracil at near wild type levels, but this has not been shown in humans. Given that UNG2 seems to play a larger role in human cells, its inactivation may predispose cells to uracilation more than in UNG-deficient mouse cells. Unlike *Ung*$^{-/-}$ mice, UNG-deficient humans likely still express some form of the enzyme, albeit mutated. Thus, it may be incorrect to assume that UNG2 dysfunction in humans could be mitigated by SMUG1 in the same way that it is in *Ung*$^{-/-}$ mice because mutated UNG2 may still outcompete SMUG1 in uracil binding, thereby negating its effects as a backup glycosylase. Such a situation may not be well reflected *in vitro* because of the relatively high abundance of uracil in most assays.

10.2 AID deamination causes uracil accumulation in B-cell malignancies

AID can be a major contributor to the genomic uracil load in B-cells. We (in article IV) and Shalhout *et al.* showed that AID and genomic uracil levels correlated well in lymphoma cell lines [259]. Moreover, genomic uracil levels could be attenuated by attenuating AID levels, *i.e.* exogenous AID overexpression or endogenous AID induction by cytokine stimulation increased genomic uracil levels, while AID knockdown decreased uracil levels. Uracil-DNA glycosylase expression and uracil excision activity did not correlate as well with genomic uracil as AID expression, so it is tempting to conclude that AID deamination overwhelms BER in the cell lines we tested. On the contrary, the lymphoma cell lines exhibited lower overall uracil excision activities, indicating that uracil excision was at

least partly compromised. Furthermore, the article by Shalhout *et al.* showed that genomic uracil levels are not increased after cytokine stimulation in primary B-cells. Thus, genomic uracil levels are a function of both uracil introduction (in this case by enzymatic deamination) and repair.

Enzymatic cytosine deamination also increases the genomic uracil burden in patients. The torrent of publications since 2012 showing mutational signatures attributed to AID and APOBECs in cancers provide convincing evidence that genomic uracil is involved in cancer development and progression [401–409,416,417]. Shalhout *et al.* also demonstrated that AID and genomic uracil levels correlate well in B-cells from CLL patients [259]. Although we did not directly test patient samples, we found a mutational signature specific to AID instead of A3 in published B-cell lymphoma and CLL sequence data, which substantiates the role AID in increasing genomic uracil levels in B-cell malignancies. Interestingly, Shalhout *et al.* also found that UNG2 and SMUG1 expression levels varied little between tumor samples, so AID deamination seems to overwhelm uracil repair in these cells. The combined evidence points to a role of deaminase-induced uracil in B-cell malignancies, though whether and the extent to which enzymatic deamination leads to cancer or is a byproduct of dysregulated DNA damage repair remain unknown.

It remains unclear whether uracil repair is a targeted event. In normal CSR, AID and UNG2 are recruited to S regions, so uracilation is likely a transient event in which uracils are quickly added by AID and partially removed by UNG2 [418,419]. Thus, both an attenuation of UNG2 or AID levels and a dysregulation of scaffolding or recruitment proteins can alter the genomic uracil burden. There is no evidence that APOBECs other than AID are recruited to specific DNA regions or that they co-localize with and uracil-DNA glycosylase, so their activity likely leaves uracils scattered throughout the genome in both ssDNA and dsDNA contexts. UNG2 or SMUG1 must then scan the entire genome for uracils before the next replication cycle, which may be too high a burden for the glycosylases to remove all the uracil. Thus, unless uracil-DNA glycosylases are targeted to uracilation sites like UNG2 during antibody maturation, even proficient BER may not be sufficient to remove uracils resulting from enzymatic deamination. The result of APOBEC or untargeted AID deamination could therefore be only trickle of uracils resulting from only partial repair of enzymatically deaminated cytosines. Furthermore, a cell may experience different levels of susceptibility to uracil accumulation, *e.g.* if glycosylase levels are low. The consequent genomic instability could help drive cancer development or progression. Future research

should therefore focus on elucidating the relative contributions of uracil introduction versus uracil repair in increasing the overall genomic uracil load.

10.3 The importance of accurate absolute genomic uracil quantification

In light of the apparent involvement of genomic uracil in cancer, it is crucial that groups employ standardized methods for genomic uracil measurement. Of the recent publications dealing with deaminase-induced *kataegis*, only Burns *et al.* measured genomic uracil [403]. They used UNG to excise uracil and measure it by LC/MS/MS on two breast cancer cell lines high in A3B expression. They found that very high A3B-expressing cell lines contained 15 to 20 Ura per 10^6 dN, whereas our results in article IV from high AID-expressing cell lines contained up to 4 dUrd per 10^6 dN, which was 72-fold higher than cells with low AID expression. The large difference between our results could be explained biologically: A3B may introduce more uracils than AID and/or they may not be excised as efficiently in the cells we tested. On the other hand, technical variations in our two methods may have influenced the results. For example, it has been reported by the Ames group that free uracil may contaminate samples during assay preparations (*e.g.* during vacuum centrifugation) [413,420]. Such a contamination may be systematic and therefore not affect the conclusions derived from the assay; however, a general overestimation of genomic uracil by one group precludes direct comparison of their results with another group's results. This is further exemplified by the article by Shalhout *et al.* that measured genomic uracil levels and AID expression in very similar samples (including some of the same cell lines) to our article IV [259]. They did not provide absolute uracil quantification, so although their results look similar to our own in that AID and genomic uracil levels correlate, albeit with a more pronounced difference between high and low uracil samples, we cannot directly compare our data.

Perhaps more strikingly, our own uracil measurements exhibited some variation. We reported that wild type MEFs contained ~0.1 and ~2 dUrd per 10^6 dN in articles II and III, respectively. The main difference in the two measurements was that UNG2 deuracilated DNA was not used as a control in article III. Thus, some the dUrd measured was likely generated *in vitro* during sample preparation. The high uracil values in *e.g.* $Ung^{-/-}Smug1^{-/-}$ organs were less affected by the overestimation because in assay uracilation is subtracted from (and not divided by) dUrd measurements to give the final value, but lower wild type, $Smug1^{+/-}$, and $Smug1^{-/-}$ values may have been overestimated and are thus not comparable to the uracil values in articles II and IV.

It is therefore important to strive towards a universally accepted "gold standard" uracil assay that accurately quantifies the absolute genomic uracil levels so that values may be compared between different groups. Currently, no direct comparisons can be made using data from different groups unless within a relative context (*i.e.* perturbed versus unperturbed samples). TDG is thought to only excise U:G, so it may be employed to distinguish between U:A and U:G given a high enough U:G specificity. TDG's turnover rate is much slower than that of UNG2, but it has been shown to increase *in vitro* when SUMOylated or with the addition of APE1, making it more suitable for uracil quantification [133–136,421]. Several groups have also made efforts to quantify genomic uracil within specific genes, although it is unclear whether their methods have the resolution to measure basal genomic uracil levels [422–424]. Regardless of this, it is clear that genomic uracil measurements will likely yield substantial insights into cancer biology, so standardizing the methods employed is crucial to this end. We have made substantial progress in this field, but there remains room for improvement.

10.4 Final thoughts: translating genomic uracil research to the clinic

The role of genomic uracil in cancer has potential applications in the clinic, but first several key questions need to be answered: at what stages of cancer is uracilation occurring, on what factors does it depend, and when is it pathogenic? In the short term, APOBEC, repair protein, and uracil levels could be compared to mutation data over time the better elucidate when uracil leads to mutation, as well as which factors affect its introduction and mutagenicity. Such a study would require large, well-stratified cohorts with multiple tumor samples gathered over time *e.g.* before and after treatment, disease progression, or even onset of disease. From this, it may be possible to partly elucidate what role uracilation is responsible for in cancer development, treatment response, or disease progression. Genomic uracil or factors that lead to uracilation like high APOBEC or low uracil-DNA glycosylase levels could then be employed as biomarkers that give prognostic insight into *e.g.* cancer risk, therapy response, or risk of disease progression. More optimistically, novel therapeutic strategies could target uracilation pathways to both increase or decrease genomic uracil. Decreasing genomic uracil in cancer or pre-cancer cells by *e.g.* inhibiting enzymatic deamination could decrease risk of disease progression by eliminating a source of genomic instability. Alternatively, increasing genomic uracil by *e.g.* TS-inhibition could target cancer cells already rich in uracil to overwhelm DNA damage repair and induce apoptosis.

The exciting developments in the field of genomic uracil in the last several years have propelled the field into the limelight. Here I have presented data indicating that the generation

genomic uracil is a complex phenomenon and its consequences affect both immunity and cancer. Moreover, the enzymes involved in genomic uracil regulation also have important roles in epigenetic regulation. Future research will likely take advantage of genomic uracil in disease prevention, diagnosis, and treatment.

11. REFERENCES

[1] E.C. Friedberg, G. Walker, W. Siede, DNA repair and mutagenesis, ASM Press, 1995. http://books.google.com/books?id=xVmZKUie2lkC.

[2] D.W. Mosbaugh, S.E. Bennett, Uracil-excision DNA repair, Prog Nucleic Acid Res Mol Biol. 48 (1994) 315–70.

[3] C.K. Mathews, DNA precursor metabolism and genomic stability, FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 20 (2006) 1300–1314. doi:10.1096/fj.06-5730rev.

[4] T.W. Traut, Physiological concentrations of purines and pyrimidines, Mol Cell Biochem. 140 (1994) 1–22.

[5] E. Wist, O. Unhjem, H. Krokan, Accumulation of small fragments of DNA in isolated HeLa cell nuclei due to transient incorporation of dUMP, Biochim. Biophys. Acta. 520 (1978) 253–270.

[6] S. Kaneda, J. Nalbantoglu, K. Takeishi, K. Shimizu, O. Gotoh, T. Seno, et al., Structural and functional analysis of the human thymidylate synthase gene, J. Biol. Chem. 265 (1990) 20277–20284.

[7] A.M. Gilles, E. Presecan, A. Vonica, I. Lascu, Nucleoside diphosphate kinase from human erythrocytes. Structural characterization of the two polypeptide chains responsible for heterogeneity of the hexameric enzyme, J. Biol. Chem. 266 (1991) 8784–8789.

[8] S.H. Huang, A. Tang, B. Drisco, S.Q. Zhang, R. Seeger, C. Li, et al., Human dTMP kinase: gene expression and enzymatic activity coinciding with cell cycle progression and cell growth, DNA Cell Biol. 13 (1994) 461–471.

[9] R. Olinski, M. Jurgowiak, T. Zaremba, Uracil in DNA-Its biological significance, Mutat Res. 705 (2010) 239–45. doi:10.1016/j.mrrev.2010.08.001.

[10] E. Chargaff, Chemical specificity of nucleic acids and mechanism of their enzymatic degradation, Experientia. 6 (1950) 201–9.

[11] J. Romiguier, V. Ranwez, E.J.P. Douzery, N. Galtier, Contrasting GC-content dynamics across 33 mammalian genomes: Relationship with life-history traits and chromosome sizes, Genome Res. 20 (2010) 1001–1009. doi:10.1101/gr.104372.109.

[12] S.I. Ahmad, S.H. Kirk, A. Eisenstark, Thymine metabolism and thymineless death in prokaryotes and eukaryotes, Annu. Rev. Microbiol. 52 (1998) 591–625. doi:10.1146/annurev.micro.52.1.591.

[13] R.D. Ladner, The role of dUTPase and uracil-DNA repair in cancer chemotherapy, Curr. Protein Pept. Sci. 2 (2001) 361–370.

[14] K.J. Kuong, A. Kuzminov, Disintegration of nascent replication bubbles during thymine starvation triggers RecA- and RecBCD-dependent replication origin destruction, J. Biol. Chem. 287 (2012) 23958–23970. doi:10.1074/jbc.M112.359687.

[15] W.E. Howe, D.W. Mount, Production of cells without deoxyribonucleic acid during thymidine starvation of lexA- cultures of Escherichia coli K-12., J. Bacteriol. 124 (1975) 1113–1121.

[16] B.A. Kunz, B.W. Glickman, Mechanism of mutation by thymine starvation in Escherichia coli: clues from mutagenic specificity., J. Bacteriol. 162 (1985) 859–864.

[17] T. Lindahl, B. Nyberg, Heat-induced deamination of cytosine residues in deoxyribonucleic acid, Biochemistry (Mosc.). 13 (1974) 3405–10.

[18] L.A. Frederico, T.A. Kunkel, B.R. Shaw, A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy, Biochemistry (Mosc.). 29 (1990) 2532–2537.

[19] T. Lindahl, Instability and decay of the primary structure of DNA, Nature. 362 (1993) 709–15. doi:10.1038/362709a0.

[20] S. Tamir, S. Burney, S.R. Tannenbaum, DNA damage by nitric oxide, Chem. Res. Toxicol. 9 (1996) 821–827. doi:10.1021/tx9600311.
[21] P.C. Dedon, S.R. Tannenbaum, Reactive nitrogen species in the chemical biology of inflammation, Arch. Biochem. Biophys. 423 (2004) 12–22.
[22] P. Lonkar, P.C. Dedon, Reactive species and DNA damage in chronic inflammation: reconciling chemical mechanisms and biological fates, Int. J. Cancer J. Int. Cancer. 128 (2011) 1999–2009. doi:10.1002/ijc.25815.
[23] V. Bouvard, R. Baan, K. Straif, Y. Grosse, B. Secretan, F. El Ghissassi, et al., A review of human carcinogens--Part B: biological agents, Lancet Oncol. 10 (2009) 321–322.
[24] M. Dong, C. Wang, W.M. Deen, P.C. Dedon, Absence of 2'-deoxyoxanosine and presence of abasic sites in DNA exposed to nitric oxide at controlled physiological concentrations, Chem Res Toxicol. 16 (2003) 1044–55. doi:10.1021/tx034046s.
[25] M. Dong, P.C. Dedon, Relatively small increases in the steady-state levels of nucleobase deamination products in DNA from human TK6 cells exposed to toxic levels of nitric oxide, Chem Res Toxicol. 19 (2006) 50–7. doi:10.1021/tx050252j.
[26] S.G. Conticello, The AID/APOBEC family of nucleic acid mutators, Genome Biol. 9 (2008) 229. doi:10.1186/gb-2008-9-6-229.
[27] V.C. Vieira, M.A. Soares, The Role of Cytidine Deaminases on Innate Immune Responses against Human Viral Infections, BioMed Res. Int. 2013 (2013) e683095. doi:10.1155/2013/683095.
[28] A. Moris, S. Murray, S. Cardinaud, AID and APOBECs span the gap between innate and adaptive immunity, Front. Microbiol. 5 (2014) 534. doi:10.3389/fmicb.2014.00534.
[29] J. Greeve, I. Altkemper, J.H. Dieterich, H. Greten, E. Windler, Apolipoprotein B mRNA editing in 12 different mammalian species: hepatic expression is reflected in low concentrations of apoB-containing plasma lipoproteins, J. Lipid Res. 34 (1993) 1367–1383.
[30] N. Navaratnam, J.R. Morrison, S. Bhattacharya, D. Patel, T. Funahashi, F. Giannoni, et al., The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase, J. Biol. Chem. 268 (1993) 20709–20712.
[31] B. Teng, C.F. Burant, N.O. Davidson, Molecular cloning of an apolipoprotein B messenger RNA editing protein, Science. 260 (1993) 1816–1819.
[32] W. Liao, S.H. Hong, B.H. Chan, F.B. Rudolph, S.C. Clark, L. Chan, APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family, Biochem. Biophys. Res. Commun. 260 (1999) 398–404. doi:10.1006/bbrc.1999.0925.
[33] C. Etard, U. Roostalu, U. Strähle, Lack of Apobec2-related proteins causes a dystrophic muscle phenotype in zebrafish embryos, J. Cell Biol. 189 (2010) 527–539. doi:10.1083/jcb.200912125.
[34] A. Vonica, A. Rosa, B.L. Arduini, A.H. Brivanlou, APOBEC2, a selective inhibitor of TGFβ signaling, regulates left-right axis specification during early embryogenesis, Dev. Biol. 350 (2011) 13–23. doi:10.1016/j.ydbio.2010.09.016.
[35] K. Rai, I.J. Huggins, S.R. James, A.R. Karpf, D.A. Jones, B.R. Cairns, DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45, Cell. 135 (2008) 1201–1212. doi:10.1016/j.cell.2008.11.042.
[36] I.B. Rogozin, M.K. Basu, I.K. Jordan, Y.I. Pavlov, E.V. Koonin, APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis, Cell Cycle Georget. Tex. 4 (2005) 1281–1285.
[37] A.G. Lada, C.F. Krick, S.G. Kozmin, V.I. Mayorov, T.S. Karpova, I.B. Rogozin, et al., Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast, Biochem. Biokhimiia. 76 (2011) 131–146.

[38] H.P. Bogerd, H.L. Wiegand, B.P. Doehle, K.K. Lueders, B.R. Cullen, APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells, Nucleic Acids Res. 34 (2006) 89–95. doi:10.1093/nar/gkj416.
[39] M. Kinomoto, T. Kanno, M. Shimura, Y. Ishizaka, A. Kojima, T. Kurata, et al., All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition, Nucleic Acids Res. 35 (2007) 2955–2964. doi:10.1093/nar/gkm181.
[40] F. Delebecque, R. Suspène, S. Calattini, N. Casartelli, A. Saïb, A. Froment, et al., Restriction of foamy viruses by APOBEC cytidine deaminases, J. Virol. 80 (2006) 605–614. doi:10.1128/JVI.80.2.605-614.2006.
[41] K. Yu, F.-T. Huang, M.R. Lieber, DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine, J. Biol. Chem. 279 (2004) 6496–6500. doi:10.1074/jbc.M311616200.
[42] R.P. Bennett, V. Presnyak, J.E. Wedekind, H.C. Smith, Nuclear Exclusion of the HIV-1 host defense factor APOBEC3G requires a novel cytoplasmic retention signal and is not dependent on RNA binding, J. Biol. Chem. 283 (2008) 7320–7327. doi:10.1074/jbc.M708567200.
[43] L. Lackey, Z.L. Demorest, A.M. Land, J.F. Hultquist, W.L. Brown, R.S. Harris, APOBEC3B and AID have similar nuclear import mechanisms, J. Mol. Biol. 419 (2012) 301–314. doi:10.1016/j.jmb.2012.03.011.
[44] M.M.H. Li, M. Emerman, Polymorphism in Human APOBEC3H Affects a Phenotype Dominant for Subcellular Localization and Antiviral Activity ▽, J. Virol. 85 (2011) 8197–8207. doi:10.1128/JVI.00624-11.
[45] R.C.L. Beale, S.K. Petersen-Mahrt, I.N. Watt, R.S. Harris, C. Rada, M.S. Neuberger, Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo, J. Mol. Biol. 337 (2004) 585–596. doi:10.1016/j.jmb.2004.01.046.
[46] A.E. Armitage, A. Katzourakis, T. de Oliveira, J.J. Welch, R. Belshaw, K.N. Bishop, et al., Conserved footprints of APOBEC3G on Hypermutated human immunodeficiency virus type 1 and human endogenous retrovirus HERV-K(HML2) sequences, J. Virol. 82 (2008) 8743–8761. doi:10.1128/JVI.00584-08.
[47] K.N. Bishop, R.K. Holmes, A.M. Sheehy, N.O. Davidson, S.-J. Cho, M.H. Malim, Cytidine deamination of retroviral DNA by diverse APOBEC proteins, Curr. Biol. CB. 14 (2004) 1392–1396. doi:10.1016/j.cub.2004.06.057.
[48] R.S. Harris, K.N. Bishop, A.M. Sheehy, H.M. Craig, S.K. Petersen-Mahrt, I.N. Watt, et al., DNA deamination mediates innate immunity to retroviral infection, Cell. 113 (2003) 803–809.
[49] E.W. Refsland, M.D. Stenglein, K. Shindo, J.S. Albin, W.L. Brown, R.S. Harris, Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction, Nucleic Acids Res. 38 (2010) 4274–4284. doi:10.1093/nar/gkq174.
[50] S. Wissing, M. Montano, J.L. Garcia-Perez, J.V. Moran, W.C. Greene, Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells, J. Biol. Chem. 286 (2011) 36427–36437. doi:10.1074/jbc.M111.251058.
[51] M. Muramatsu, V.S. Sankaranand, S. Anant, M. Sugai, K. Kinoshita, N.O. Davidson, et al., Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells, J Biol Chem. 274 (1999) 18470–6.

[52] M. Muramatsu, K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, T. Honjo, Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme, Cell. 102 (2000) 553–563.
[53] P. Revy, T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, et al., Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2), Cell. 102 (2000) 565–575.
[54] A. Gazumyan, A. Bothmer, I.A. Klein, M.C. Nussenzweig, K.M. McBride, Activation-induced cytidine deaminase in antibody diversification and chromosome translocation, Adv. Cancer Res. 113 (2012) 167–190. doi:10.1016/B978-0-12-394280-7.00005-1.
[55] Y. Matsumoto, H. Marusawa, K. Kinoshita, Y. Endo, T. Kou, T. Morisawa, et al., Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium, Nat. Med. 13 (2007) 470–476. doi:10.1038/nm1566.
[56] T. Chiba, H. Marusawa, A novel mechanism for inflammation-associated carcinogenesis; an important role of activation-induced cytidine deaminase (AID) in mutation induction, J. Mol. Med. Berl. Ger. 87 (2009) 1023–1027. doi:10.1007/s00109-009-0527-3.
[57] T. Muto, M. Muramatsu, M. Taniwaki, K. Kinoshita, T. Honjo, Isolation, Tissue Distribution, and Chromosomal Localization of the Human Activation-Induced Cytidine Deaminase (AID) Gene, Genomics. 68 (2000) 85–88. doi:10.1006/geno.2000.6268.
[58] R.W. Yen, P.M. Vertino, B.D. Nelkin, J.J. Yu, W. el-Deiry, A. Cumaraswamy, et al., Isolation and characterization of the cDNA encoding human DNA methyltransferase, Nucleic Acids Res. 20 (1992) 2287–2291.
[59] J.C. Shen, W.M. Rideout, P.A. Jones, High frequency mutagenesis by a DNA methyltransferase, Cell. 71 (1992) 1073–80.
[60] J.C. Shen, J.M. Zingg, A.S. Yang, C. Schmutte, P.A. Jones, A mutant HpaII methyltransferase functions as a mutator enzyme, Nucleic Acids Res. 23 (1995) 4275–82.
[61] J.M. Zingg, J.C. Shen, A.S. Yang, H. Rapoport, P.A. Jones, Methylation inhibitors can increase the rate of cytosine deamination by (cytosine-5)-DNA methyltransferase., Nucleic Acids Res. 24 (1996) 3267–3275.
[62] J.M. Zingg, J.C. Shen, P.A. Jones, Enzyme-mediated cytosine deamination by the bacterial methyltransferase M.MspI, Biochem. J. 332 (Pt 1) (1998) 223–230.
[63] R. Métivier, R. Gallais, C. Tiffoche, C. Le Péron, R.Z. Jurkowska, R.P. Carmouche, et al., Cyclical DNA methylation of a transcriptionally active promoter, Nature. 452 (2008) 45–50. doi:10.1038/nature06544.
[64] D.O. Zharkov, Base excision DNA repair, Cell. Mol. Life Sci. CMLS. 65 (2008) 1544–1565. doi:10.1007/s00018-008-7543-2.
[65] H.E. Krokan, M. Bjørås, Base excision repair, Cold Spring Harb. Perspect. Biol. 5 (2013) a012583. doi:10.1101/cshperspect.a012583.
[66] K.H. Almeida, R.W. Sobol, A unified view of base excision repair: lesion-dependent protein complexes regulated by post-translational modification, DNA Repair. 6 (2007) 695–711. doi:10.1016/j.dnarep.2007.01.009.
[67] A.B. Robertson, A. Klungland, T. Rognes, I. Leiros, DNA repair in mammalian cells: Base excision repair: the long and short of it, Cell. Mol. Life Sci. CMLS. 66 (2009) 981–993. doi:10.1007/s00018-009-8736-z.
[68] M. Akbari, J. Pena-Diaz, S. Andersen, N.B. Liabakk, M. Otterlei, H.E. Krokan, Extracts of proliferating and non-proliferating human cells display different base excision pathways and repair fidelity, DNA Repair Amst. 8 (2009) 834–43. doi:10.1016/j.dnarep.2009.04.002.

[69] S.L. Cravens, M. Hobson, J.T. Stivers, Electrostatic properties of complexes along a DNA glycosylase damage search pathway, Biochemistry (Mosc.). (2014). doi:10.1021/bi501011m.
[70] M. Slutsky, L.A. Mirny, Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential, Biophys. J. 87 (2004) 4021–4035. doi:10.1529/biophysj.104.050765.
[71] L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, A. Kosmrlj, How a protein searches for its site on DNA: the mechanism of facilitated diffusion, J. Phys. Math. Theor. 42 (2009) 434013. doi:10.1088/1751-8113/42/43/434013.
[72] G.V. Mechetin, D.O. Zharkov, Mechanisms of diffusional search for specific targets by DNA-dependent proteins, Biochem. Biokhimiia. 79 (2014) 496–505. doi:10.1134/S0006297914060029.
[73] J.D. Schonhoft, J.T. Stivers, Timing facilitated site transfer of an enzyme on DNA, Nat. Chem. Biol. 8 (2012) 205–210. doi:10.1038/nchembio.764.
[74] D.M. Gowers, G.G. Wilson, S.E. Halford, Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15883–15888. doi:10.1073/pnas.0505378102.
[75] N.P. Stanford, M.D. Szczelkun, J.F. Marko, S.E. Halford, One- and three-dimensional pathways for proteins to reach specific DNA sites, EMBO J. 19 (2000) 6546–6557. doi:10.1093/emboj/19.23.6546.
[76] B. Dalhus, J.K. Laerdahl, P.H. Backe, M. Bjørås, DNA base repair--recognition and initiation of catalysis, FEMS Microbiol. Rev. 33 (2009) 1044–1078. doi:10.1111/j.1574-6976.2009.00188.x.
[77] H.E. Krokan, R. Standal, G. Slupphaug, DNA glycosylases in the base excision repair of DNA, Biochem J. 325 (Pt 1) (1997) 1–16.
[78] D. Svilar, E.M. Goellner, K.H. Almeida, R.W. Sobol, Base excision repair and lesion-dependent subpathways for repair of oxidative DNA damage, Antioxid. Redox Signal. 14 (2011) 2491–2507. doi:10.1089/ars.2010.3466.
[79] P. Fortini, E. Dogliotti, Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways, DNA Repair. 6 (2007) 398–409. doi:10.1016/j.dnarep.2006.10.008.
[80] E. Dogliotti, P. Fortini, B. Pascucci, E. Parlanti, The mechanism of switching among multiple BER pathways, Prog. Nucleic Acid Res. Mol. Biol. 68 (2001) 3–27.
[81] M. Sukhanova, S. Khodyreva, O. Lavrik, Poly(ADP-ribose) polymerase 1 regulates activity of DNA polymerase beta in long patch base excision repair, Mutat. Res. 685 (2010) 80–89. doi:10.1016/j.mrfmmm.2009.08.009.
[82] L. Narciso, P. Fortini, D. Pajalunga, A. Franchitto, P. Liu, P. Degan, et al., Terminally differentiated muscle cells are defective in base excision DNA repair and hypersensitive to oxygen injury, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 17010–17015. doi:10.1073/pnas.0701743104.
[83] M. Bauer, M. Goldstein, M. Christmann, H. Becker, D. Heylmann, B. Kaina, Human monocytes are severely impaired in base and DNA double-strand break repair that renders them vulnerable to oxidative stress, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 21105–21110. doi:10.1073/pnas.1111919109.
[84] E.D. Tichy, L. Liang, L. Deng, J. Tischfield, S. Schwemberger, G. Babcock, et al., Mismatch and base excision repair proficiency in murine embryonic stem cells, DNA Repair. 10 (2011) 445–451. doi:10.1016/j.dnarep.2011.01.008.
[85] P. Fortini, E. Parlanti, O.M. Sidorkina, J. Laval, E. Dogliotti, The type of DNA glycosylase determines the base excision repair pathway in mammalian cells, J. Biol. Chem. 274 (1999) 15230–15236.

[86] R.W. Sobol, R. Prasad, A. Evenski, A. Baker, X.P. Yang, J.K. Horton, et al., The lyase activity of the DNA repair protein beta-polymerase protects from DNA-damage-induced cytotoxicity, Nature. 405 (2000) 807–810. doi:10.1038/35015598.
[87] S.L. Allinson, I.I. Dianova, G.L. Dianov, DNA polymerase beta is the major dRP lyase involved in repair of oxidative base lesions in DNA by mammalian cell extracts, EMBO J. 20 (2001) 6919–6926. doi:10.1093/emboj/20.23.6919.
[88] A.J. Podlutsky, I.I. Dianova, S.H. Wilson, V.A. Bohr, G.L. Dianov, DNA synthesis and dRPase activities of polymerase beta are both essential for single-nucleotide patch base excision repair in mammalian cell extracts, Biochemistry (Mosc.). 40 (2001) 809–813.
[89] M.J. Longley, R. Prasad, D.K. Srivastava, S.H. Wilson, W.C. Copeland, Identification of 5'-deoxyribose phosphate lyase activity in human DNA polymerase gamma and its role in mitochondrial base excision repair in vitro, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 12244–12248.
[90] K. Bebenek, A. Tissier, E.G. Frank, J.P. McDonald, R. Prasad, S.H. Wilson, et al., 5'-Deoxyribose phosphate lyase activity of human DNA polymerase iota in vitro, Science. 291 (2001) 2156–2159. doi:10.1126/science.1058386.
[91] E.K. Braithwaite, R. Prasad, D.D. Shock, E.W. Hou, W.A. Beard, S.H. Wilson, DNA polymerase lambda mediates a back-up base excision repair activity in extracts of mouse embryonic fibroblasts, J. Biol. Chem. 280 (2005) 18469–18475. doi:10.1074/jbc.M411864200.
[92] R. Prasad, M.J. Longley, F.S. Sharief, E.W. Hou, W.C. Copeland, S.H. Wilson, Human DNA polymerase theta possesses 5'-dRP lyase activity and functions in single-nucleotide base excision repair in vitro, Nucleic Acids Res. 37 (2009) 1868–1877. doi:10.1093/nar/gkp035.
[93] Y. Kubota, R.A. Nash, A. Klungland, P. Schar, D.E. Barnes, T. Lindahl, Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein., EMBO J. 15 (1996) 6662–6670.
[94] Y. Gao, S. Katyal, Y. Lee, J. Zhao, J.E. Rehg, H.R. Russell, et al., DNA ligase III is critical for mtDNA integrity but not Xrcc1-mediated nuclear DNA repair, Nature. 471 (2011) 240–244. doi:10.1038/nature09773.
[95] D. Simsek, A. Furda, Y. Gao, J. Artus, E. Brunet, A.-K. Hadjantonakis, et al., Crucial role for DNA ligase III in mitochondria but not in Xrcc1-dependent repair, Nature. 471 (2011) 245–248. doi:10.1038/nature09794.
[96] B. Pascucci, M. Stucki, Z.O. Jónsson, E. Dogliotti, U. Hübscher, Long patch base excision repair with purified human proteins. DNA ligase I as patch size mediator for DNA polymerases delta and epsilon, J. Biol. Chem. 274 (1999) 33696–33702.
[97] H.E. Krokan, P. Sætrom, P.A. Aas, H.S. Pettersen, B. Kavli, G. Slupphaug, Error-free versus mutagenic processing of genomic uracil—Relevance to cancer, DNA Repair. 19 (2014) 38–47. doi:10.1016/j.dnarep.2014.03.028.
[98] B. Doseth, C. Ekre, G. Slupphaug, H.E. Krokan, B. Kavli, Strikingly different properties of uracil-DNA glycosylases UNG2 and SMUG1 may explain divergent roles in processing of genomic uracil, DNA Repair. 11 (2012) 587–593. doi:10.1016/j.dnarep.2012.03.003.
[99] H. Nilsen, I. Rosewell, P. Robins, C.F. Skjelbred, S. Andersen, G. Slupphaug, et al., Uracil-DNA Glycosylase (UNG)-Deficient Mice Reveal a Primary Role of the Enzyme during DNA Replication, Mol. Cell. 5 (2000) 1059–1065. doi:10.1016/S1097-2765(00)80271-3.
[100] B. Kavli, O. Sundheim, M. Akbari, M. Otterlei, H. Nilsen, F. Skorpen, 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 (2002) 39926–36. doi:10.1074/jbc.M207107200.
[101] H. Nilsen, M. Otterlei, T. Haug, K. Solum, T.A. Nagelhus, F. Skorpen, et al., Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene., Nucleic Acids Res. 25 (1997) 750–755.
[102] T. Haug, F. Skorpen, P.A. Aas, V. Malm, C. Skjelbred, H.E. Krokan, Regulation of expression of nuclear and mitochondrial forms of human uracil-DNA glycosylase., Nucleic Acids Res. 26 (1998) 1449–1457.
[103] G. Slupphaug, I. Eftedal, B. Kavli, S. Bharati, N.M. Helle, T. Haug, 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 (Mosc.). 34 (1995) 128–38.
[104] I. Eftedal, P.H. Guddal, G. Slupphaug, G. Volden, H.E. Krokan, Consensus sequences for good and poor removal of uracil from double stranded DNA by uracil-DNA glycosylase, Nucleic Acids Res. 21 (1993) 2095–2101.
[105] H. Nilsen, S.P. Yazdankhah, I. Eftedal, H.E. Krokan, 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 (1995) 205–209.
[106] M. Dizdaroglu, A. Karakaya, P. Jaruga, G. Slupphaug, H.E. Krokan, Novel activities of human uracil DNA N-glycosylase for cytosine-derived products of oxidative DNA damage, Nucleic Acids Res. 24 (1996) 418–422.
[107] L. Hagen, B. Kavli, M.M. Sousa, K. Torseth, N.B. Liabakk, O. Sundheim, et al., Cell cycle-specific UNG2 phosphorylations regulate protein turnover, activity and association with RPA, EMBO J. 27 (2008) 51–61. doi:10.1038/sj.emboj.7601958.
[108] M. Akbari, M. Otterlei, J. Peña-Diaz, P.A. Aas, B. Kavli, N.B. Liabakk, et al., Repair of U/G and U/A in DNA by UNG2-associated repair complexes takes place predominantly by short-patch repair both in proliferating and growth-arrested cells, Nucleic Acids Res. 32 (2004) 5486–5498. doi:10.1093/nar/gkh872.
[109] E. Parlanti, G. Locatelli, G. Maga, E. Dogliotti, Human base excision repair complex is physically associated to DNA replication and cell cycle regulatory proteins, Nucleic Acids Res. 35 (2007) 1569–1577. doi:10.1093/nar/gkl1159.
[110] M. Otterlei, E. Warbrick, T.A. Nagelhus, T. Haug, G. Slupphaug, M. Akbari, et al., Post-replicative base excision repair in replication foci, EMBO J. 18 (1999) 3834–3844. doi:10.1093/emboj/18.13.3834.
[111] G. Mer, A. Bochkarev, R. Gupta, E. Bochkareva, L. Frappier, C.J. Ingles, et al., Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA, Cell. 103 (2000) 449–456.
[112] R.J. Boorstein, A. Cummings, D.R. Marenstein, M.K. Chan, Y. Ma, T.A. Neubert, et al., Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1, J. Biol. Chem. 276 (2001) 41991–41997. doi:10.1074/jbc.M106953200.
[113] A. Masaoka, M. Matsubara, R. Hasegawa, T. Tanaka, S. Kurisu, H. Terato, et al., Mammalian 5-formyluracil-DNA glycosylase. 2. Role of SMUG1 uracil-DNA glycosylase in repair of 5-formyluracil and other oxidized and deaminated base lesions, Biochemistry (Mosc.). 42 (2003) 5003–5012. doi:10.1021/bi0273213.
[114] A. Darwanto, J.A. Theruvathu, J.L. Sowers, D.K. Rogstad, T. Pascal, W. Goddard, et al., Mechanisms of base selection by human single-stranded selective monofunctional

uracil-DNA glycosylase, J. Biol. Chem. 284 (2009) 15835–15846. doi:10.1074/jbc.M807846200.
[115] B. Kavli, S. Andersen, M. Otterlei, N.B. Liabakk, K. Imai, A. Fischer, 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 (2005) 2011–2021. doi:10.1084/jem.20050042.
[116] H.S. Pettersen, O. Sundheim, K.M. Gilljam, G. Slupphaug, H.E. Krokan, B. Kavli, Uracil–DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms, Nucleic Acids Res. 35 (2007) 3879–3892. doi:10.1093/nar/gkm372.
[117] E. Zarakowska, D. Gackowski, M. Foksinski, R. Olinski, Are 8-oxoguanine (8-oxoGua) and 5-hydroxymethyluracil (5-hmUra) oxidatively damaged DNA bases or transcription (epigenetic) marks?, Mutat. Res. Genet. Toxicol. Environ. Mutagen. 764-765 (2014) 58–63. doi:10.1016/j.mrgentox.2013.09.002.
[118] J.U. Guo, Y. Su, C. Zhong, G. Ming, H. Song, Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain, Cell. 145 (2011) 423–434. doi:10.1016/j.cell.2011.03.022.
[119] M.R. Branco, G. Ficz, W. Reik, Uncovering the role of 5-hydroxymethylcytosine in the epigenome, Nat. Rev. Genet. 13 (2012) 7–13. doi:10.1038/nrg3080.
[120] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S.K. Laube, D. Eisen, et al., Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA, Nat. Chem. Biol. 10 (2014) 574–581. doi:10.1038/nchembio.1532.
[121] K. Kemmerich, F.A. Dingler, C. Rada, M.S. Neuberger, Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice, Nucleic Acids Res. 40 (2012) 6016–6025. doi:10.1093/nar/gks259.
[122] R.J. Boorstein, L.N. Chiu, G.W. Teebor, Phylogenetic evidence of a role for 5-hydroxymethyluracil-DNA glycosylase in the maintenance of 5-methylcytosine in DNA, Nucleic Acids Res. 17 (1989) 7653–7661.
[123] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, et al., Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates, PLoS ONE. 5 (2010) e15367. doi:10.1371/journal.pone.0015367.
[124] A. Szwagierczak, S. Bultmann, C.S. Schmidt, F. Spada, H. Leonhardt, Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA, Nucleic Acids Res. 38 (2010) e181. doi:10.1093/nar/gkq684.
[125] S. Kriaucionis, N. Heintz, The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain, Science. 324 (2009) 929–930. doi:10.1126/science.1169786.
[126] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, et al., Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair, Cell. 146 (2011) 67–79. doi:10.1016/j.cell.2011.06.020.
[127] C.S. Nabel, H. Jia, Y. Ye, L. Shen, H.L. Goldschmidt, J.T. Stivers, et al., AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation, Nat. Chem. Biol. 8 (2012) 751–758. doi:10.1038/nchembio.1042.
[128] G. Rangam, K.-M. Schmitz, A.J.A. Cobb, S.K. Petersen-Mahrt, AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud, PloS One. 7 (2012) e43279. doi:10.1371/journal.pone.0043279.
[129] F.-M. Boisvert, S. van Koningsbruggen, J. Navascués, A.I. Lamond, The multifunctional nucleolus, Nat. Rev. Mol. Cell Biol. 8 (2007) 574–585. doi:10.1038/nrm2184.

[130] L. Jobert, H.K. Skjeldam, B. Dalhus, A. Galashevskaya, C.B. Vågbø, M. Bjørås, et al., The human base excision repair enzyme SMUG1 directly interacts with DKC1 and contributes to RNA quality control, Mol. Cell. 49 (2013) 339–345. doi:10.1016/j.molcel.2012.11.010.
[131] P. Neddermann, J. Jiricny, Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells, Proc. Natl. Acad. Sci. U. S. A. 91 (1994) 1642–1646.
[132] P. Neddermann, P. Gallinari, T. Lettieri, D. Schmid, O. Truong, J.J. Hsuan, et al., Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase, J. Biol. Chem. 271 (1996) 12767–12774.
[133] U. Hardeland, M. Bentele, J. Jiricny, P. Schär, Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis, J. Biol. Chem. 275 (2000) 33449–33456. doi:10.1074/jbc.M005095200.
[134] U. Hardeland, R. Steinacher, J. Jiricny, P. Schär, Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover, EMBO J. 21 (2002) 1456–1464. doi:10.1093/emboj/21.6.1456.
[135] R. Steinacher, P. Schär, Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation, Curr. Biol. CB. 15 (2005) 616–623. doi:10.1016/j.cub.2005.02.054.
[136] U. Hardeland, M. Bentele, T. Lettieri, R. Steinacher, J. Jiricny, P. Schär, Thymine DNA glycosylase, Prog. Nucleic Acid Res. Mol. Biol. 68 (2001) 235–253.
[137] A. Maiti, A.C. Drohat, Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites, J. Biol. Chem. 286 (2011) 35334–35338. doi:10.1074/jbc.C111.284620.
[138] Y.-F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, et al., Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA, Science. 333 (2011) 1303–1307. doi:10.1126/science.1210944.
[139] M.T. Bennett, M.T. Rodgers, A.S. Hebert, L.E. Ruslander, L. Eisele, A.C. Drohat, Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability, J. Am. Chem. Soc. 128 (2006) 12510–12519. doi:10.1021/ja0634829.
[140] R.M. Kohli, Y. Zhang, TET enzymes, TDG and the dynamics of DNA demethylation, Nature. 502 (2013) 472–479. doi:10.1038/nature12750.
[141] B. Hendrich, U. Hardeland, H.H. Ng, J. Jiricny, A. Bird, The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites, Nature. 401 (1999) 301–304. doi:10.1038/45843.
[142] B.A. Manvilla, A. Maiti, M.C. Begley, E.A. Toth, A.C. Drohat, Crystal structure of human methyl-binding domain IV glycosylase bound to abasic DNA, J. Mol. Biol. 420 (2012) 164–175. doi:10.1016/j.jmb.2012.04.028.
[143] A.B. Sjolund, A.G. Senejani, J.B. Sweasy, MBD4 and TDG: multifaceted DNA glycosylases with ever expanding biological roles, Mutat. Res. 743-744 (2013) 12–25. doi:10.1016/j.mrfmmm.2012.11.001.
[144] V. Valinluck, P. Liu, J.I. Kang, A. Burdzy, L.C. Sowers, 5-Halogenated pyrimidine lesions within a CpG sequence context mimic 5-methylcytosine by enhancing the binding of the methyl-CpG-binding domain of methyl-CpG-binding protein 2 (MeCP2), Nucleic Acids Res. 33 (2005) 3057–3064. doi:10.1093/nar/gki612.
[145] J.P. Henderson, J. Byun, J. Takeshita, J.W. Heinecke, Phagocytes produce 5-chlorouracil and 5-bromouracil, two mutagenic products of myeloperoxidase, in human inflammatory tissue, J. Biol. Chem. 278 (2003) 23522–23528. doi:10.1074/jbc.M303928200.

[146] D.P. Turner, S. Cortellino, J.E. Schupp, E. Caretti, T. Loh, T.J. Kinsella, et al., The DNA N-glycosylase MED1 exhibits preference for halogenated pyrimidines and is involved in the cytotoxicity of 5-iododeoxyuridine, Cancer Res. 66 (2006) 7686–7693. doi:10.1158/0008-5472.CAN-05-4488.
[147] M.A. Aziz, J.E. Schupp, T.J. Kinsella, Modulation of the activity of methyl binding domain protein 4 (MBD4/MED1) while processing iododeoxyuridine generated DNA mispairs, Cancer Biol. Ther. 8 (2009) 1156–1163.
[148] F. Petronzelli, A. Riccio, G.D. Markham, S.H. Seeholzer, M. Genuardi, M. Karbowski, et al., Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): fundamental role of the catalytic domain, J. Cell. Physiol. 185 (2000) 473–480. doi:10.1002/1097-4652(200012)185:3<473::AID-JCP19>3.0.CO;2-#.
[149] B. Zhu, Y. Zheng, H. Angliker, S. Schwarz, S. Thiry, M. Siegmann, et al., 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence, Nucleic Acids Res. 28 (2000) 4157–4165.
[150] M.-S. Kim, T. Kondo, I. Takada, M.-Y. Youn, Y. Yamamoto, S. Takahashi, et al., DNA demethylation in hormone-induced transcriptional derepression, Nature. 461 (2009) 1007–1012. doi:10.1038/nature08456.
[151] G. Bridge, S. Rashid, S.A. Martin, DNA mismatch repair and oxidative DNA damage: implications for cancer biology and treatment, Cancers. 6 (2014) 1597–1614. doi:10.3390/cancers6031597.
[152] G.-M. Li, Mechanisms and functions of DNA mismatch repair, Cell Res. 18 (2008) 85–98. doi:10.1038/cr.2007.115.
[153] J. Genschel, S.J. Littman, J.T. Drummond, P. Modrich, Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha, J. Biol. Chem. 273 (1998) 19895–19901.
[154] R.D. Kolodner, G.T. Marsischky, Eukaryotic DNA mismatch repair, Curr. Opin. Genet. Dev. 9 (1999) 89–96.
[155] G.M. Li, P. Modrich, Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 1950–1954.
[156] G. Plotz, J. Raedle, A. Brieger, J. Trojan, S. Zeuzem, hMutS? forms an ATP-dependent complex with hMutL? and hMutL? on DNA, Nucleic Acids Res. 30 (2002) 711–718.
[157] E. Cannavo, G. Marra, J. Sabates-Bellver, M. Menigatti, S.M. Lipkin, F. Fischer, et al., Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair, Cancer Res. 65 (2005) 10759–10766. doi:10.1158/0008-5472.CAN-05-2528.
[158] Y. Zhang, F. Yuan, S.R. Presnell, K. Tian, Y. Gao, A.E. Tomkinson, et al., Reconstitution of 5'-directed human mismatch repair in a purified system, Cell. 122 (2005) 693–705. doi:10.1016/j.cell.2005.06.027.
[159] J. Genschel, P. Modrich, Mechanism of 5'-directed excision in human mismatch repair, Mol. Cell. 12 (2003) 1077–1086.
[160] F.A. Kadyrov, L. Dzantiev, N. Constantin, P. Modrich, Endonucleolytic function of MutLalpha in human mismatch repair, Cell. 126 (2006) 297–308. doi:10.1016/j.cell.2006.05.039.
[161] L. Dzantiev, N. Constantin, J. Genschel, R.R. Iyer, P.M. Burgers, P. Modrich, A defined human system that supports bidirectional mismatch-provoked excision, Mol. Cell. 15 (2004) 31–41. doi:10.1016/j.molcel.2004.06.016.

[162] K. Wei, A.B. Clark, E. Wong, M.F. Kane, D.J. Mazur, T. Parris, et al., Inactivation of Exonuclease 1 in mice results in DNA mismatch repair defects, increased cancer susceptibility, and male and female sterility, Genes Dev. 17 (2003) 603–614. doi:10.1101/gad.1060603.
[163] J. Jiricny, The multifaceted mismatch-repair system, Nat. Rev. Mol. Cell Biol. 7 (2006) 335–346. doi:10.1038/nrm1907.
[164] H. Wang, J.B. Hays, Signaling from DNA mispairs to mismatch-repair excision sites despite intervening blockades, EMBO J. 23 (2004) 2126–2133. doi:10.1038/sj.emboj.7600153.
[165] M.S. Junop, G. Obmolova, K. Rausch, P. Hsieh, W. Yang, Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair, Mol. Cell. 7 (2001) 1–12.
[166] D.J. Allen, A. Makhov, M. Grilley, J. Taylor, R. Thresher, P. Modrich, et al., MutS mediates heteroduplex loop formation by a translocation mechanism, EMBO J. 16 (1997) 4467–4476. doi:10.1093/emboj/16.14.4467.
[167] R. Fishel, Mismatch repair, molecular switches, and signal transduction, Genes Dev. 12 (1998) 2096–2101.
[168] S. Gradia, S. Acharya, R. Fishel, The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch, Cell. 91 (1997) 995–1005.
[169] J. Jiang, L. Bai, J.A. Surtees, Z. Gemici, M.D. Wang, E. Alani, Detection of high-affinity and sliding clamp modes for MSH2-MSH6 by single-molecule unzipping force analysis, Mol. Cell. 20 (2005) 771–781. doi:10.1016/j.molcel.2005.10.014.
[170] M.L. Mendillo, D.J. Mazur, R.D. Kolodner, Analysis of the interaction between the Saccharomyces cerevisiae MSH2-MSH6 and MLH1-PMS1 complexes with DNA using a reversible DNA end-blocking system, J. Biol. Chem. 280 (2005) 22245–22257. doi:10.1074/jbc.M407545200.
[171] T. Iyama, D.M. Wilson, DNA repair mechanisms in dividing and non-dividing cells, DNA Repair. 12 (2013) 620–636. doi:10.1016/j.dnarep.2013.04.015.
[172] A. Pluciennik, L. Dzantiev, R.R. Iyer, N. Constantin, F.A. Kadyrov, P. Modrich, PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 16066–16071. doi:10.1073/pnas.1010662107.
[173] S.A. Nick McElhinny, G.E. Kissling, T.A. Kunkel, Differential correction of lagging-strand replication errors made by DNA polymerases {alpha} and {delta}, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 21070–21075. doi:10.1073/pnas.1013048107.
[174] E.D. Larson, D.W. Bednarski, N. Maizels, High-fidelity correction of genomic uracil by human mismatch repair activities, BMC Mol. Biol. 9 (2008) 94. doi:10.1186/1471-2199-9-94.
[175] E.D. Larson, M.L. Duquette, W.J. Cummings, R.J. Streiff, N. Maizels, MutSalpha binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions, Curr. Biol. CB. 15 (2005) 470–474. doi:10.1016/j.cub.2004.12.077.
[176] T.M. Wilson, A. Vaisman, S.A. Martomo, P. Sullivan, L. Lan, F. Hanaoka, et al., MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A:T mutations in antibody genes, J. Exp. Med. 201 (2005) 637–645. doi:10.1084/jem.20042066.
[177] H. Wang, C.W. Lawrence, G.M. Li, J.B. Hays, Specific binding of human MSH2.MSH6 mismatch-repair protein heterodimers to DNA incorporating thymine- or uracil-containing UV light photoproducts opposite mismatched bases, J. Biol. Chem. 274 (1999) 16894–16900.

[178] M. Meyers, M.W. Wagner, A. Mazurek, C. Schmutte, R. Fishel, D.A. Boothman, DNA mismatch repair-dependent response to fluoropyrimidine-generated damage, J. Biol. Chem. 280 (2005) 5516–5526. doi:10.1074/jbc.M412105200.
[179] K.P. Murphy, Janeway's Immunobiology, n.d.
[180] D.G. Schatz, P.C. Swanson, V(D)J recombination: mechanisms of initiation, Annu. Rev. Genet. 45 (2011) 167–202. doi:10.1146/annurev-genet-110410-132552.
[181] C.D. Allen, T. Okada, J.G. Cyster, Germinal Center Organization and Cellular Dynamics, Immunity. 27 (2007) 190–202. doi:10.1016/j.immuni.2007.07.009.
[182] K.P. Lam, R. Kühn, K. Rajewsky, In vivo ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death, Cell. 90 (1997) 1073–1083.
[183] B. Schiemann, J.L. Gommerman, K. Vora, T.G. Cachero, S. Shulga-Morskaya, M. Dobles, et al., An essential role for BAFF in the normal development of B cells through a BCMA-independent pathway, Science. 293 (2001) 2111–2114. doi:10.1126/science.1061964.
[184] S. Han, K. Hathcock, B. Zheng, T.B. Kepler, R. Hodes, G. Kelsoe, Cellular interaction in germinal centers. Roles of CD40 ligand and B7-2 in established germinal centers, J. Immunol. Baltim. Md 1950. 155 (1995) 556–567.
[185] T. Kawabe, T. Naka, K. Yoshida, T. Tanaka, H. Fujiwara, S. Suematsu, et al., The immune responses in CD40-deficient mice: impaired immunoglobulin class switching and germinal center formation, Immunity. 1 (1994) 167–178.
[186] J.M. Di Noia, M.S. Neuberger, Molecular mechanisms of antibody somatic hypermutation, Annu. Rev. Biochem. 76 (2007) 1–22. doi:10.1146/annurev.biochem.76.061705.090740.
[187] J.U. Peled, F.L. Kuang, M.D. Iglesias-Ussel, S. Roa, S.L. Kalis, M.F. Goodman, et al., The biochemistry of somatic hypermutation, Annu. Rev. Immunol. 26 (2008) 481–511. doi:10.1146/annurev.immunol.26.021607.090236.
[188] Y.J. Liu, D.Y. Mason, G.D. Johnson, S. Abbot, C.D. Gregory, D.L. Hardie, et al., Germinal center cells express bcl-2 protein after activation by signals which prevent their entry into apoptosis, Eur. J. Immunol. 21 (1991) 1905–1910. doi:10.1002/eji.1830210819.
[189] U. Klein, Y. Tu, G.A. Stolovitzky, J.L. Keller, J. Haddad, V. Miljkovic, et al., Transcriptional analysis of the B cell germinal center reaction, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 2639–2644. doi:10.1073/pnas.0437996100.
[190] I.C. MacLennan, Germinal centers, Annu. Rev. Immunol. 12 (1994) 117–139. doi:10.1146/annurev.iy.12.040194.001001.
[191] G. Billian, C. Bella, P. Mondière, T. Defrance, Identification of a tonsil IgD+ B cell subset with phenotypical and functional characteristics of germinal center B cells, Eur. J. Immunol. 26 (1996) 1712–1719. doi:10.1002/eji.1830260808.
[192] J. Feuillard, D. Taylor, M. Casamayor-Palleja, G.D. Johnson, I.C. MacLennan, Isolation and characteristics of tonsil centroblasts with reference to Ig class switching, Int. Immunol. 7 (1995) 121–130.
[193] Y.-J. Liu, D.E. Joshua, G.T. Williams, C.A. Smith, J. Gordon, I.C.M. MacLennan, Mechanism of antigen-driven selection in germinal centres, Nature. 342 (1989) 929–931. doi:10.1038/342929a0.
[194] K. Rajewsky, Clonal selection and learning in the antibody system, Nature. 381 (1996) 751–758. doi:10.1038/381751a0.
[195] D. Paus, T.G. Phan, T.D. Chan, S. Gardam, A. Basten, R. Brink, Antigen recognition strength regulates the choice between extrafollicular plasma cell and germinal center B cell differentiation, J. Exp. Med. 203 (2006) 1081–1091. doi:10.1084/jem.20060087.

[196] T.G. Phan, D. Paus, T.D. Chan, M.L. Turner, S.L. Nutt, A. Basten, et al., High affinity germinal center B cells are actively selected into the plasma cell compartment, J. Exp. Med. 203 (2006) 2419–2424. doi:10.1084/jem.20061254.
[197] M.S. Neuberger, R.S. Harris, J. Di Noia, S.K. Petersen-Mahrt, Immunity through DNA deamination, Trends Biochem Sci. 28 (2003) 305–12.
[198] H. Arakawa, J. Hauschild, J.-M. Buerstedde, Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion, Science. 295 (2002) 1301–1306. doi:10.1126/science.1067308.
[199] E.S. Tang, A. Martin, Immunoglobulin gene conversion: synthesizing antibody diversification and DNA repair, DNA Repair. 6 (2007) 1557–1571. doi:10.1016/j.dnarep.2007.05.002.
[200] G.M. Edelman, Antibody structure and molecular immunology, Science. 180 (1973) 830–840.
[201] G.M. Edelman, Antibody structure and molecular immunology, Scand. J. Immunol. 34 (1991) 1–22.
[202] S. Kracker, A. Durandy, Insights into the B cell specific process of immunoglobulin class switch recombination, Immunol. Lett. 138 (2011) 97–103. doi:10.1016/j.imlet.2011.02.004.
[203] J. Stavnezer, C.E. Schrader, IgH Chain Class Switch Recombination: Mechanism and Regulation, J. Immunol. Baltim. Md 1950. 193 (2014) 5370–5378. doi:10.4049/jimmunol.1401849.
[204] T. Kataoka, T. Kawakami, N. Takahashi, T. Honjo, Rearrangement of immunoglobulin gamma 1-chain gene and mechanism for heavy-chain class switch., Proc. Natl. Acad. Sci. U. S. A. 77 (1980) 919–923.
[205] N. Maizels, Immunoglobulin gene diversification, Annu. Rev. Genet. 39 (2005) 23–46. doi:10.1146/annurev.genet.39.073003.110544.
[206] M. Shapiro-Shelef, K. Calame, Regulation of plasma-cell development, Nat. Rev. Immunol. 5 (2005) 230–242. doi:10.1038/nri1572.
[207] J. Alinikula, O. Lassila, Gene interaction network regulates plasma cell differentiation, Scand. J. Immunol. 73 (2011) 512–519. doi:10.1111/j.1365-3083.2011.02556.x.
[208] K.-P. Nera, O. Lassila, Pax5--a critical inhibitor of plasma cell fate, Scand. J. Immunol. 64 (2006) 190–199. doi:10.1111/j.1365-3083.2006.01809.x.
[209] S.A. Omori, R.C. Rickert, Phosphatidylinositol 3-kinase (PI3K) signaling and regulation of the antibody response, Cell Cycle Georget. Tex. 6 (2007) 397–402.
[210] J.M. Di Noia, G.T. Williams, D.T.Y. Chan, J.-M. Buerstedde, G.S. Baldwin, M.S. Neuberger, Dependence of antibody gene diversification on uracil excision, J. Exp. Med. 204 (2007) 3209–3219. doi:10.1084/jem.20071768.
[211] R.W. Maul, P.J. Gearhart, AID and somatic hypermutation, Adv Immunol. 105 (2010) 159–91. doi:10.1016/S0065-2776(10)05006-6.
[212] S.K. Petersen-Mahrt, R.S. Harris, M.S. Neuberger, AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification, Nature. 418 (2002) 99–103. doi:10.1038/nature00862.
[213] A. Longacre, T. Sun, R.E. Goldsby, B.D. Preston, U. Storb, Ig gene somatic hypermutation in mice defective for DNA polymerase delta proofreading, Int. Immunol. 15 (2003) 477–481.
[214] P. Casali, Z. Pal, Z. Xu, H. Zan, DNA repair in antibody somatic hypermutation, Trends Immunol. 27 (2006) 313–321. doi:10.1016/j.it.2006.05.001.
[215] C. Rada, G.T. Williams, H. Nilsen, D.E. Barnes, T. Lindahl, M.S. Neuberger, Immunoglobulin Isotype Switching Is Inhibited and Somatic Hypermutation Perturbed

in UNG-Deficient Mice, Curr. Biol. 12 (2002) 1748–1755. doi:10.1016/S0960-9822(02)01215-0.
[216] H. Arakawa, G.-L. Moldovan, H. Saribasak, N.N. Saribasak, S. Jentsch, J.-M. Buerstedde, A Role for PCNA Ubiquitination in Immunoglobulin Hypermutation, PLoS Biol. 4 (2006) e366. doi:10.1371/journal.pbio.0040366.
[217] J.G. Jansen, P. Langerak, A. Tsaalbi-Shtylik, P. van den Berk, H. Jacobs, N. de Wind, Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice, J. Exp. Med. 203 (2006) 319–323. doi:10.1084/jem.20052227.
[218] K. Masuda, R. Ouchida, Y. Li, X. Gao, H. Mori, J.-Y. Wang, A critical role for REV1 in regulating the induction of C:G transitions and A:T mutations during Ig gene hypermutation, J. Immunol. Baltim. Md 1950. 183 (2009) 1846–1850. doi:10.4049/jimmunol.0901240.
[219] A.-L. Ross, J.E. Sale, The catalytic activity of REV1 is employed during immunoglobulin gene diversification in DT40, Mol. Immunol. 43 (2006) 1587–1594. doi:10.1016/j.molimm.2005.09.017.
[220] P. Auerbach, R.A.O. Bennett, E.A. Bailey, H.E. Krokan, B. Demple, Mutagenic specificity of endogenously generated abasic sites in Saccharomyces cerevisiae chromosomal DNA, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 17711–17716. doi:10.1073/pnas.0504643102.
[221] J.R. Nelson, C.W. Lawrence, D.C. Hinkle, Deoxycytidyl transferase activity of yeast REV1 protein, Nature. 382 (1996) 729–731. doi:10.1038/382729a0.
[222] P.H.L. Krijger, P.C.M. van den Berk, N. Wit, P. Langerak, J.G. Jansen, C.-A. Reynaud, et al., PCNA ubiquitination-independent activation of polymerase η during somatic hypermutation and DNA damage tolerance, DNA Repair. 10 (2011) 1051–1059. doi:10.1016/j.dnarep.2011.08.005.
[223] P.H.L. Krijger, A. Tsaalbi-Shtylik, N. Wit, P.C.M. van den Berk, N. de Wind, H. Jacobs, Rev1 is essential in generating G to C transversions downstream of the Ung2 pathway but not the Msh2+Ung2 hybrid pathway, Eur. J. Immunol. 43 (2013) 2765–2770. doi:10.1002/eji.201243191.
[224] C. Kano, F. Hanaoka, J.-Y. Wang, Analysis of mice deficient in both REV1 catalytic activity and POLH reveals an unexpected role for POLH in the generation of C to G and G to C transversions during Ig gene hypermutation, Int. Immunol. 24 (2012) 169–174. doi:10.1093/intimm/dxr109.
[225] P.H.L. Krijger, P. Langerak, P.C.M. van den Berk, H. Jacobs, Dependence of nucleotide substitutions on Ung2, Msh2, and PCNA-Ub during somatic hypermutation, J. Exp. Med. 206 (2009) 2603–2611. doi:10.1084/jem.20091707.
[226] J. Stavnezer, E.K. Linehan, M.R. Thompson, G. Habboub, A.J. Ucher, T. Kadungure, et al., Differential expression of APE1 and APE2 in germinal centers promotes error-prone repair and A:T mutations during somatic hypermutation, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 9217–9222. doi:10.1073/pnas.1405590111.
[227] P. Garg, P.M. Burgers, Ubiquitinated proliferating cell nuclear antigen activates translesion DNA polymerases eta and REV1, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 18361–18366. doi:10.1073/pnas.0505949102.
[228] Y. Ide, D. Tsuchimoto, Y. Tominaga, Y. Iwamoto, Y. Nakabeppu, Characterization of the genomic structure and expression of the mouse Apex2 gene, Genomics. 81 (2003) 47–57.
[229] F. Delbos, S. Aoufouchi, A. Faili, J.-C. Weill, C.-A. Reynaud, DNA polymerase eta is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse, J. Exp. Med. 204 (2007) 17–23. doi:10.1084/jem.20062131.

[230] F. Delbos, A. De Smet, A. Faili, S. Aoufouchi, J.-C. Weill, C.-A. Reynaud, Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse, J. Exp. Med. 201 (2005) 1191–1196. doi:10.1084/jem.20050292.
[231] A. Faili, A. Stary, F. Delbos, S. Weller, S. Aoufouchi, A. Sarasin, et al., A backup role of DNA polymerase kappa in Ig gene hypermutation only takes place in the complete absence of DNA polymerase eta, J. Immunol. Baltim. Md 1950. 182 (2009) 6353–6359. doi:10.4049/jimmunol.0900177.
[232] S.A. Martomo, W.W. Yang, R.P. Wersto, T. Ohkumo, Y. Kondo, M. Yokoi, et al., Different mutation signatures in DNA polymerase eta- and MSH6-deficient mice suggest separate roles in antibody diversification, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 8656–8661. doi:10.1073/pnas.0501852102.
[233] X. Zeng, D.B. Winter, C. Kasmer, K.H. Kraemer, A.R. Lehmann, P.J. Gearhart, DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes, Nat. Immunol. 2 (2001) 537–541. doi:10.1038/88740.
[234] C. Rada, M.R. Ehrenstein, M.S. Neuberger, C. Milstein, Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting, Immunity. 9 (1998) 135–141.
[235] H. Saribasak, R.W. Maul, Z. Cao, W.W. Yang, D. Schenten, S. Kracker, et al., DNA polymerase ζ generates tandem mutations in immunoglobulin variable regions, J. Exp. Med. 209 (2012) 1075–1081. doi:10.1084/jem.20112234.
[236] M. Wiesendanger, B. Kneitz, W. Edelmann, M.D. Scharff, Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern, J. Exp. Med. 191 (2000) 579–584.
[237] P.D. Bardwell, C.J. Woo, K. Wei, Z. Li, A. Martin, S.Z. Sack, et al., Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice, Nat. Immunol. 5 (2004) 224–229. doi:10.1038/ni1031.
[238] R.W. Maul, P.J. Gearhart, Refining the Neuberger Model: uracil processing by activated B cells, Eur. J. Immunol. (2014) n/a–n/a. doi:10.1002/eji.201444813.
[239] J. Stavnezer, J.E.J. Guikema, C.E. Schrader, Mechanism and regulation of class switch recombination, Annu. Rev. Immunol. 26 (2008) 261–292. doi:10.1146/annurev.immunol.26.021607.090248.
[240] W. Dunnick, G.Z. Hertz, L. Scappino, C. Gritzmacher, DNA sequences at immunoglobulin switch region recombination sites., Nucleic Acids Res. 21 (1993) 365–372.
[241] I.M. Min, L.R. Rothlein, C.E. Schrader, J. Stavnezer, E. Selsing, Shifts in targeting of class switch recombination sites in mice that lack μ switch region tandem repeats or Msh2, J. Exp. Med. 201 (2005) 1885–1890. doi:10.1084/jem.20042491.
[242] A.G. Betz, C. Milstein, A. González-Fernández, R. Pannell, T. Larson, M.S. Neuberger, Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region, Cell. 77 (1994) 239–248.
[243] I.B. Rogozin, N.A. Kolchanov, Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis, Biochim. Biophys. Acta. 1171 (1992) 11–18.
[244] G.S. Shapiro, K. Aviszus, D. Ikle, L.J. Wysocki, Predicting regional mutability in antibody V genes based solely on di- and trinucleotide sequence composition, J. Immunol. Baltim. Md 1950. 163 (1999) 259–268.

[245] S.K. Dickerson, E. Market, E. Besmer, F.N. Papavasiliou, AID Mediates Hypermutation by Deaminating Single Stranded DNA, J. Exp. Med. 197 (2003) 1291–1296. doi:10.1084/jem.20030481.
[246] J. Chaudhuri, M. Tian, C. Khuong, K. Chua, E. Pinaud, F.W. Alt, Transcription-targeted DNA deamination by the AID antibody diversification enzyme, Nature. 422 (2003) 726–730. doi:10.1038/nature01574.
[247] P. Pham, R. Bransteitter, J. Petruska, M.F. Goodman, Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation, Nature. 424 (2003) 103–107. doi:10.1038/nature01760.
[248] J. Hasbold, A.B. Lyons, M.R. Kehry, P.D. Hodgkin, Cell division number regulates IgG1 and IgE switching of B cells following stimulation by CD40 ligand and IL-4, Eur. J. Immunol. 28 (1998) 1040–1051. doi:10.1002/(SICI)1521-4141(199803)28:03<1040::AID-IMMU1040>3.0.CO;2-9.
[249] P.D. Hodgkin, J.H. Lee, A.B. Lyons, B cell differentiation and isotype switching is related to division cycle number, J. Exp. Med. 184 (1996) 277–281.
[250] U. Basu, F.-L. Meng, C. Keim, V. Grinstein, E. Pefanis, J. Eccleston, et al., The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates, Cell. 144 (2011) 353–363. doi:10.1016/j.cell.2011.01.001.
[251] J.S. Rush, M. Liu, V.H. Odegard, S. Unniraman, D.G. Schatz, Expression of activation-induced cytidine deaminase is regulated by cell division, providing a mechanistic basis for division-linked class switch recombination, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 13242–13247. doi:10.1073/pnas.0502779102.
[252] C.E. Schrader, E.K. Linehan, S.N. Mochegova, R.T. Woodland, J. Stavnezer, Inducible DNA breaks in Ig S regions are dependent on AID and UNG, J. Exp. Med. 202 (2005) 561–568. doi:10.1084/jem.20050872.
[253] F.A. Dingler, K. Kemmerich, M.S. Neuberger, C. Rada, Uracil excision by endogenous SMUG1 glycosylase promotes efficient Ig class switching and impacts on A:T substitutions during somatic mutation, Eur. J. Immunol. 44 (2014) 1925–1935. doi:10.1002/eji.201444482.
[254] K. Imai, G. Slupphaug, W.I. Lee, P. Revy, S. Nonoyama, N. Catalan, et al., Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination, Nat Immunol. 4 (2003) 1023–8. doi:10.1038/ni974.
[255] J.M. Di Noia, C. Rada, M.S. Neuberger, SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair, EMBO J. 25 (2006) 585–595. doi:10.1038/sj.emboj.7600939.
[256] J.E.J. Guikema, E.K. Linehan, D. Tsuchimoto, Y. Nakabeppu, P.R. Strauss, J. Stavnezer, et al., APE1- and APE2-dependent DNA breaks in immunoglobulin class switch recombination, J. Exp. Med. 204 (2007) 3017–3026. doi:10.1084/jem.20071289.
[257] J. Xu, A. Husain, W. Hu, T. Honjo, M. Kobayashi, APE1 is dispensable for S-region cleavage but required for its repair in class switch recombination, Proc. Natl. Acad. Sci. U. S. A. (2014). doi:10.1073/pnas.1420221111.
[258] S. Masani, L. Han, K. Yu, Apurinic/apyrimidinic endonuclease 1 is the essential nuclease during immunoglobulin class switch recombination, Mol. Cell. Biol. 33 (2013) 1468–1473. doi:10.1128/MCB.00026-13.
[259] S. Shalhout, D. Haddad, A. Sosin, T.C. Holland, A. Al-Katib, A. Martin, et al., Genomic Uracil Homeostasis during Normal B Cell Maturation and Loss of this Balance During B Cell Cancer Development, Mol. Cell. Biol. (2014) MCB.00589–14. doi:10.1128/MCB.00589-14.

[260] D.A. Kaminski, J. Stavnezer, Stimuli that enhance IgA class switching increase histone 3 acetylation at Sα, but poorly stimulate sequential switching from IgG2b, Eur. J. Immunol. 37 (2007) 240–251. doi:10.1002/eji.200636645.
[261] X. Wu, J. Stavnezer, DNA polymerase β is able to repair breaks in switch regions and plays an inhibitory role during immunoglobulin class switch recombination, J. Exp. Med. 204 (2007) 1677–1689. doi:10.1084/jem.20070756.
[262] J. Jiricny, Postreplicative Mismatch Repair, Cold Spring Harb. Perspect. Biol. 5 (2013) a012633. doi:10.1101/cshperspect.a012633.
[263] S. Schaetzlein, R. Chahwan, E. Avdievich, S. Roa, K. Wei, R.L. Eoff, et al., Mammalian Exo1 encodes both structural and catalytic functions that play distinct roles in essential biological processes, Proc. Natl. Acad. Sci. 110 (2013) E2470–E2479. doi:10.1073/pnas.1308512110.
[264] J.M.M. van Oers, S. Roa, U. Werling, Y. Liu, J. Genschel, H. Hou, et al., PMS2 endonuclease activity has distinct biological functions and is essential for genome maintenance, Proc. Natl. Acad. Sci. 107 (2010) 13384–13389. doi:10.1073/pnas.1008589107.
[265] T.M. Gottlieb, S.P. Jackson, The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen, Cell. 72 (1993) 131–142.
[266] S.A. Nick McElhinny, C.M. Snowden, J. McCarville, D.A. Ramsden, Ku recruits the XRCC4-ligase IV complex to DNA ends, Mol. Cell. Biol. 20 (2000) 2996–3003.
[267] Y. Ma, U. Pannicke, K. Schwarz, M.R. Lieber, Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination, Cell. 108 (2002) 781–794.
[268] Y. Ma, K. Schwarz, M.R. Lieber, The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps, DNA Repair. 4 (2005) 845–851. doi:10.1016/j.dnarep.2005.04.013.
[269] Y. Ma, H. Lu, B. Tippin, M.F. Goodman, N. Shimazaki, O. Koiwai, et al., A biochemically defined system for mammalian nonhomologous DNA end joining, Mol. Cell. 16 (2004) 701–713. doi:10.1016/j.molcel.2004.11.017.
[270] J. Gu, H. Lu, B. Tippin, N. Shimazaki, M.F. Goodman, M.R. Lieber, XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps, EMBO J. 26 (2007) 1010–1023. doi:10.1038/sj.emboj.7601559.
[271] J. Gu, H. Lu, A.G. Tsai, K. Schwarz, M.R. Lieber, Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence, Nucleic Acids Res. 35 (2007) 5755–5762. doi:10.1093/nar/gkm579.
[272] U. Grawunder, M. Wilm, X. Wu, P. Kulesza, T.E. Wilson, M. Mann, et al., Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells, Nature. 388 (1997) 492–495. doi:10.1038/41358.
[273] C. Boboila, M. Jankovic, C.T. Yan, J.H. Wang, D.R. Wesemann, T. Zhang, et al., Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absence of ligase 4 and Ku70, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 3034–3039. doi:10.1073/pnas.0915067107.
[274] L. Han, K. Yu, Altered kinetics of nonhomologous end joining and class switch recombination in ligase IV–deficient B cells, J. Exp. Med. 205 (2008) 2745–2753. doi:10.1084/jem.20081623.
[275] J. Eccleston, C. Yan, K. Yuan, F.W. Alt, E. Selsing, Mismatch repair proteins MSH2, MLH1, and EXO1 are important for class-switch recombination events occurring in B cells that lack nonhomologous end joining, J. Immunol. Baltim. Md 1950. 186 (2011) 2336–2343. doi:10.4049/jimmunol.1003104.

[276] C.T. Yan, C. Boboila, E.K. Souza, S. Franco, T.R. Hickernell, M. Murphy, et al., IgH class switching and translocations use a robust non-classical end-joining pathway, Nature. 449 (2007) 478–482. doi:10.1038/nature06020.
[277] P. Soulas-Sprauel, G.L. Guyader, P. Rivera-Munoz, V. Abramowski, C. Olivier-Martin, C. Goujet-Zalc, et al., Role for DNA repair factor XRCC4 in immunoglobulin class switch recombination, J. Exp. Med. 204 (2007) 1717–1727. doi:10.1084/jem.20070255.
[278] Q. Pan-Hammarström, A.-M. Jones, A. Lähdesmäki, W. Zhou, R.A. Gatti, L. Hammarström, et al., Impact of DNA ligase IV on nonhomologous end joining pathways during class switch recombination in human cells, J. Exp. Med. 201 (2005) 189–194. doi:10.1084/jem.20040772.
[279] E.M. Cortizas, A. Zahn, M.E. Hajjar, A.-M. Patenaude, J.M. Di Noia, R.E. Verdun, Alternative end-joining and classical nonhomologous end-joining pathways repair different types of double-strand breaks during class-switch recombination, J. Immunol. Baltim. Md 1950. 191 (2013) 5751–5763. doi:10.4049/jimmunol.1301300.
[280] F.W. Alt, Y. Zhang, F.-L. Meng, C. Guo, B. Schwer, Mechanisms of Programmed DNA Lesions and Genomic Instability in the Immune System, Cell. 152 (2013) 417–429. doi:10.1016/j.cell.2013.01.007.
[281] S.F. Bunting, A. Nussenzweig, End-joining, translocations and cancer, Nat. Rev. Cancer. 13 (2013) 443–454. doi:10.1038/nrc3537.
[282] M. McVey, S.E. Lee, MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings, Trends Genet. TIG. 24 (2008) 529–538. doi:10.1016/j.tig.2008.08.007.
[283] N. Bennardo, A. Cheng, N. Huang, J.M. Stark, Alternative-NHEJ Is a Mechanistically Distinct Pathway of Mammalian Chromosome Break Repair, PLoS Genet. 4 (2008) e1000110. doi:10.1371/journal.pgen.1000110.
[284] I. Narvaiza, S. Landry, M.D. Weitzman, APOBEC3 proteins and genomic stability: the high cost of a good defense, Cell Cycle Georget. Tex. 11 (2012) 33–38. doi:10.4161/cc.11.1.18706.
[285] Y.-L. Chiu, W.C. Greene, The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements, Annu. Rev. Immunol. 26 (2008) 317–353. doi:10.1146/annurev.immunol.26.021607.090350.
[286] S. Cen, F. Guo, M. Niu, J. Saadatmand, J. Deflassieux, L. Kleiman, The Interaction between HIV-1 Gag and APOBEC3G, J. Biol. Chem. 279 (2004) 33177–33184. doi:10.1074/jbc.M402062200.
[287] T.M. Alce, W. Popik, APOBEC3G Is Incorporated into Virus-like Particles by a Direct Interaction with HIV-1 Gag Nucleocapsid Protein, J. Biol. Chem. 279 (2004) 34083–34086. doi:10.1074/jbc.C400235200.
[288] K. Luo, B. Liu, Z. Xiao, Y. Yu, X. Yu, R. Gorelick, et al., Amino-Terminal Region of the Human Immunodeficiency Virus Type 1 Nucleocapsid Is Required for Human APOBEC3G Packaging, J. Virol. 78 (2004) 11841–11852. doi:10.1128/JVI.78.21.11841-11852.2004.
[289] V. Zennou, D. Perez-Caballero, H. Göttlinger, P.D. Bieniasz, APOBEC3G Incorporation into Human Immunodeficiency Virus Type 1 Particles, J. Virol. 78 (2004) 12058–12061. doi:10.1128/JVI.78.21.12058-12061.2004.
[290] M. Douaisi, S. Dussart, M. Courcoul, G. Bessou, R. Vigne, E. Decroly, HIV-1 and MLV Gag proteins are sufficient to recruit APOBEC3G into virus-like particles, Biochem. Biophys. Res. Commun. 321 (2004) 566–573. doi:10.1016/j.bbrc.2004.07.005.

[291] A. Schäfer, H.P. Bogerd, B.R. Cullen, Specific packaging of APOBEC3G into HIV-1 virions is mediated by the nucleocapsid domain of the gag polyprotein precursor, Virology. 328 (2004) 163–168. doi:10.1016/j.virol.2004.08.006.
[292] T. Wang, W. Zhang, C. Tian, B. Liu, Y. Yu, L. Ding, et al., Distinct viral determinants for the packaging of human cytidine deaminases APOBEC3G and APOBEC3C, Virology. 377 (2008) 71–79. doi:10.1016/j.virol.2008.04.012.
[293] D. Lecossier, F. Bouchonnet, F. Clavel, A.J. Hance, Hypermutation of HIV-1 DNA in the Absence of the Vif Protein, Science. 300 (2003) 1112–1112. doi:10.1126/science.1083338.
[294] S.G. Conticello, R.S. Harris, M.S. Neuberger, The Vif Protein of HIV Triggers Degradation of the Human Antiretroviral DNA Deaminase APOBEC3G, Curr. Biol. 13 (2003) 2009–2013. doi:10.1016/j.cub.2003.10.034.
[295] S. Kao, M.A. Khan, E. Miyagi, R. Plishka, A. Buckler-White, K. Strebel, The Human Immunodeficiency Virus Type 1 Vif Protein Reduces Intracellular Expression and Inhibits Packaging of APOBEC3G (CEM15), a Cellular Inhibitor of Virus Infectivity, J. Virol. 77 (2003) 11398–11407. doi:10.1128/JVI.77.21.11398-11407.2003.
[296] B. Liu, X. Yu, K. Luo, Y. Yu, X.-F. Yu, Influence of Primate Lentiviral Vif and Proteasome Inhibitors on Human Immunodeficiency Virus Type 1 Virion Packaging of APOBEC3G, J. Virol. 78 (2004) 2072–2081. doi:10.1128/JVI.78.4.2072-2081.2004.
[297] M. Marin, K.M. Rose, S.L. Kozak, D. Kabat, HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation, Nat. Med. 9 (2003) 1398–1403. doi:10.1038/nm946.
[298] A. Mehle, B. Strack, P. Ancuta, C. Zhang, M. McPike, D. Gabuzda, Vif Overcomes the Innate Antiviral Activity of APOBEC3G by Promoting Its Degradation in the Ubiquitin-Proteasome Pathway, J. Biol. Chem. 279 (2004) 7792–7798. doi:10.1074/jbc.M313093200.
[299] A.M. Sheehy, N.C. Gaddis, J.D. Choi, M.H. Malim, Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein, Nature. 418 (2002) 646–650. doi:10.1038/nature00939.
[300] K. Stopak, C. de Noronha, W. Yonemoto, W.C. Greene, HIV-1 Vif Blocks the Antiviral Activity of APOBEC3G by Impairing Both Its Translation and Intracellular Stability, Mol. Cell. 12 (2003) 591–601. doi:10.1016/S1097-2765(03)00353-8.
[301] X. Yu, Y. Yu, B. Liu, K. Luo, W. Kong, P. Mao, et al., Induction of APOBEC3G Ubiquitination and Degradation by an HIV-1 Vif-Cul5-SCF Complex, Science. 302 (2003) 1056–1060. doi:10.1126/science.1089591.
[302] B. Schröfelbauer, Q. Yu, S.G. Zeitlin, N.R. Landau, Human Immunodeficiency Virus Type 1 Vpr Induces the Degradation of the UNG and SMUG Uracil-DNA Glycosylases, J. Virol. 79 (2005) 10978–10987. doi:10.1128/JVI.79.17.10978-10987.2005.
[303] B. Yang, K. Chen, C. Zhang, S. Huang, H. Zhang, Virion-associated Uracil DNA Glycosylase-2 and Apurinic/Apyrimidinic Endonuclease Are Involved in the Degradation of APOBEC3G-edited Nascent HIV-1 DNA, J. Biol. Chem. 282 (2007) 11667–11675. doi:10.1074/jbc.M606864200.
[304] S.M. Kaiser, M. Emerman, 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 (2006) 875–882. doi:10.1128/JVI.80.2.875-882.2006.
[305] M.-A. Langlois, M.S. Neuberger, Human APOBEC3G Can Restrict Retroviral Infection in Avian Cells and Acts Independently of both UNG and SMUG1, J. Virol. 82 (2008) 4660–4664. doi:10.1128/JVI.02469-07.

[306] F. Navarro, B. Bollman, H. Chen, R. König, Q. Yu, K. Chiles, et al., Complementary function of the two catalytic domains of APOBEC3G, Virology. 333 (2005) 374–386. doi:10.1016/j.virol.2005.01.011.
[307] M. Ooms, A. Krikoni, A.K. Kress, V. Simon, C. Münk, APOBEC3A, APOBEC3B, and APOBEC3H Haplotype 2 Restrict Human T-Lymphotropic Virus Type 1, J. Virol. 86 (2012) 6097–6108. doi:10.1128/JVI.06570-11.
[308] M. Cavrois, S. Wain-Hobson, A. Gessain, Y. Plumelle, E. Wattel, Adult T-cell leukemia/lymphoma on a background of clonally expanding human T-cell leukemia virus type-1-positive cells, Blood. 88 (1996) 4646–4650.
[309] M. Cavrois, I. Leclercq, O. Gout, A. Gessain, S. Wain-Hobson, E. Wattel, Persistent oligoclonal expansion of human T-cell leukemia virus type 1-infected circulating cells in patients with Tropical spastic paraparesis/HTLV-1 associated myelopathy, Oncogene. 17 (1998) 77–82.
[310] M. Matsuoka, K.-T. Jeang, Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation, Nat. Rev. Cancer. 7 (2007) 270–280. doi:10.1038/nrc2111.
[311] E. Wattel, J.-P. Vartanian, C. Pannetier, S. Wain-Hobson, Clonal expansion of human T-cell leukemia virus type I-infected cells in asymptomatic and symptomatic carriers without malignancy, J. Virol. 69 (1995) 2863–2868.
[312] D. Derse, S.A. Hill, G. Princler, P. Lloyd, G. Heidecker, Resistance of human T cell leukemia virus type 1 to APOBEC3G restriction is mediated by elements in nucleocapsid, Proc. Natl. Acad. Sci. 104 (2007) 2915–2920. doi:10.1073/pnas.0609444104.
[313] M.C. Gonzalez, R. Suspène, M. Henry, D. Guétard, S. Wain-Hobson, J.-P. Vartanian, Human APOBEC1 cytidine deaminase edits HBV DNA, Retrovirology. 6 (2009) 96. doi:10.1186/1742-4690-6-96.
[314] R. Suspène, M.-M. Aynaud, S. Koch, D. Pasdeloup, M. Labetoulle, B. Gaertner, et al., Genetic Editing of Herpes Simplex Virus 1 and Epstein-Barr Herpesvirus Genomes by Human APOBEC3 Cytidine Deaminases in Culture and In Vivo, J. Virol. 85 (2011) 7594–7602. doi:10.1128/JVI.00290-11.
[315] T.F. Baumert, C. Rösler, M.H. Malim, F. von Weizsäcker, Hepatitis B virus DNA is subject to extensive editing by the human deaminase APOBEC3C, Hepatol. Baltim. Md. 46 (2007) 682–689. doi:10.1002/hep.21733.
[316] J. Köck, H.E. Blum, Hypermutation of hepatitis B virus genomes by APOBEC3G, APOBEC3C and APOBEC3H, J. Gen. Virol. 89 (2008) 1184–1191. doi:10.1099/vir.0.83507-0.
[317] M. Henry, D. Guétard, R. Suspène, C. Rusniok, S. Wain-Hobson, J.-P. Vartanian, Genetic editing of HBV DNA by monodomain human APOBEC3 cytidine deaminases and the recombinant nature of APOBEC3G, PloS One. 4 (2009) e4277. doi:10.1371/journal.pone.0004277.
[318] C. Rösler, J. Köck, M. Kann, M.H. Malim, H.E. Blum, T.F. Baumert, et al., APOBEC-mediated interference with hepadnavirus production, Hepatol. Baltim. Md. 42 (2005) 301–309. doi:10.1002/hep.20801.
[319] J.-P. Vartanian, M. Henry, A. Marchio, R. Suspène, M.-M. Aynaud, D. Guétard, et al., Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis, PLoS Pathog. 6 (2010) e1000928. doi:10.1371/journal.ppat.1000928.
[320] P. Turelli, B. Mangeat, S. Jost, S. Vianin, D. Trono, Inhibition of hepatitis B virus replication by APOBEC3G, Science. 303 (2004) 1829. doi:10.1126/science.1092066.
[321] D. Moradpour, F. Penin, C.M. Rice, Replication of hepatitis C virus, Nat. Rev. Microbiol. 5 (2007) 453–463. doi:10.1038/nrmicro1645.

[322] Z.-G. Peng, Z.-Y. Zhao, Y.-P. Li, Y.-P. Wang, L.-H. Hao, B. Fan, et al., Host apolipoprotein b messenger RNA-editing enzyme catalytic polypeptide-like 3G is an innate defensive factor and drug target against hepatitis C virus, Hepatology. 53 (2011) 1080–1089. doi:10.1002/hep.24160.
[323] J.-P. Vartanian, D. Guétard, M. Henry, S. Wain-Hobson, Evidence for Editing of Human Papillomavirus DNA by APOBEC3 in Benign and Precancerous Lesions, Science. 320 (2008) 230–233. doi:10.1126/science.1153201.
[324] P. Gee, Y. Ando, H. Kitayama, S.P. Yamamoto, Y. Kanemura, H. Ebina, et al., APOBEC1-Mediated Editing and Attenuation of Herpes Simplex Virus 1 DNA Indicate That Neurons Have an Antiviral Role during Herpes Simplex Encephalitis, J. Virol. 85 (2011) 9726–9736. doi:10.1128/JVI.05288-11.
[325] E.G. Davies, A.J. Thrasher, Update on the hyper immunoglobulin M syndromes, Br. J. Haematol. 149 (2010) 167–180. doi:10.1111/j.1365-2141.2010.08077.x.
[326] S. Fagarasan, M. Muramatsu, K. Suzuki, H. Nagaoka, H. Hiai, T. Honjo, Critical roles of activation-induced cytidine deaminase in the homeostasis of gut flora, Science. 298 (2002) 1424–1427. doi:10.1126/science.1077336.
[327] S. Kracker, P. Gardes, F. Mazerolles, A. Durandy, Immunoglobulin class switch recombination deficiencies, Clin. Immunol. 135 (2010) 193–203. doi:10.1016/j.clim.2010.01.012.
[328] P. Quartier, J. Bustamante, O. Sanal, A. Plebani, M. Debré, A. Deville, et al., Clinical, immunologic and genetic analysis of 29 patients with autosomal recessive hyper-IgM syndrome due to Activation-Induced Cytidine Deaminase deficiency, Clin. Immunol. Orlando Fla. 110 (2004) 22–29. doi:10.1016/j.clim.2003.10.007.
[329] U. Korthäuer, D. Graf, H.W. Mages, F. Brière, M. Padayachee, S. Malcolm, et al., Defective expression of T-cell CD40 ligand causes X-linked immunodeficiency with hyper-IgM, Nature. 361 (1993) 539–541. doi:10.1038/361539a0.
[330] K. Imai, Y. Zhu, P. Revy, T. Morio, S. Mizutani, A. Fischer, et al., Analysis of class switch recombination and somatic hypermutation in patients affected with autosomal dominant hyper-IgM syndrome type 2, Clin. Immunol. Orlando Fla. 115 (2005) 277–285. doi:10.1016/j.clim.2005.02.003.
[331] V.-T. Ta, H. Nagaoka, N. Catalan, A. Durandy, A. Fischer, K. Imai, et al., AID mutant analyses indicate requirement for class-switch-specific cofactors, Nat. Immunol. 4 (2003) 843–848. doi:10.1038/ni964.
[332] S. Kracker, P. Gardës, A. Durandy, Inherited defects of immunoglobulin class switch recombination, Adv. Exp. Med. Biol. 685 (2010) 166–174.
[333] A.-M. Patenaude, A. Orthwein, Y. Hu, V.A. Campo, B. Kavli, A. Buschiazzo, et al., Active nuclear import and cytoplasmic retention of activation-induced deaminase, Nat. Struct. Mol. Biol. 16 (2009) 517–527. doi:10.1038/nsmb.1598.
[334] T. Doi, L. Kato, S. Ito, R. Shinkura, M. Wei, H. Nagaoka, et al., The C-terminal region of activation-induced cytidine deaminase is responsible for a recombination function other than DNA cleavage in class switch recombination, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 2758–2763. doi:10.1073/pnas.0813253106.
[335] A. Zahn, M. Daugan, S. Safavi, D. Godin, C. Cheong, A. Lamarre, et al., Separation of function between isotype switching and affinity maturation in vivo during acute immune responses and circulating autoantibodies in UNG-deficient mice, J. Immunol. Baltim. Md 1950. 190 (2013) 5949–5960. doi:10.4049/jimmunol.1202711.
[336] C. Rada, J.M. Di Noia, M.S. Neuberger, Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation, Mol. Cell. 16 (2004) 163–171. doi:10.1016/j.molcel.2004.10.011.

[337] H. Nilsen, G. Stamp, S. Andersen, G. Hrivnak, H.E. Krokan, T. Lindahl, et al., Gene-targeted mice lacking the Ung uracil-DNA glycosylase develop B-cell lymphomas, Oncogene. 22 (2003) 5381–5386. doi:10.1038/sj.onc.1206860.
[338] S. Andersen, M. Ericsson, H.Y. Dai, J. Pena-Diaz, G. Slupphaug, H. Nilsen, et al., Monoclonal B-cell hyperplasia and leukocyte imbalance precede development of B-cell malignancies in uracil-DNA glycosylase deficient mice, DNA Repair Amst. 4 (2005) 1432–41. doi:10.1016/j.dnarep.2005.08.004.
[339] H. Nilsen, Q. An, T. Lindahl, Mutation frequencies and AID activation state in B-cell lymphomas from Ung-deficient mice, Oncogene. 24 (2005) 3063–3066. doi:10.1038/sj.onc.1208480.
[340] Q. An, P. Robins, T. Lindahl, D.E. Barnes, C --> T mutagenesis and gamma-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases, EMBO J. 24 (2005) 2205–2213. doi:10.1038/sj.emboj.7600689.
[341] C.B. Millar, J. Guy, O.J. Sansom, J. Selfridge, E. MacDougall, B. Hendrich, et al., Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice, Science. 297 (2002) 403–405. doi:10.1126/science.1073354.
[342] E. Wong, K. Yang, M. Kuraguchi, U. Werling, E. Avdievich, K. Fan, et al., Mbd4 inactivation increases C->T transition mutations and promotes gastrointestinal tumor formation, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 14937–14942. doi:10.1073/pnas.232579299.
[343] Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency, Publ. Online 04 January 2000 Doi101038sjonc1203229. 18 (2000). doi:10.1038/sj.onc.1203229.
[344] S.A. Bader, M. Walker, D.J. Harrison, A human cancer-associated truncation of MBD4 causes dominant negative impairment of DNA repair in colon cancer cells, Br. J. Cancer. 96 (2007) 660–666. doi:10.1038/sj.bjc.6603592.
[345] C. Miquel, S. Jacob, S. Grandjouan, A. Aimé, J. Viguier, J.-C. Sabourin, et al., Frequent alteration of DNA damage signalling and repair pathways in human colorectal cancers with microsatellite instability, Oncogene. 26 (2007) 5919–5926. doi:10.1038/sj.onc.1210419.
[346] A. Riccio, L.A. Aaltonen, A.K. Godwin, A. Loukola, A. Percesepe, R. Salovaara, et al., The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability, Nat. Genet. 23 (1999) 266–268. doi:10.1038/15443.
[347] T. Yamada, T. Koyama, S. Ohwada, K. Tago, I. Sakamoto, S. Yoshimura, et al., Frameshift mutations in the MBD4/MED1 gene in primary gastric cancer with high-frequency microsatellite instability, Cancer Lett. 181 (2002) 115–120. doi:10.1016/S0304-3835(02)00043-5.
[348] P. Broderick, T. Bagratuni, J. Vijayakrishnan, S. Lubbe, I. Chandler, R.S. Houlston, Evaluation of NTHL1, NEIL1, NEIL2, MPG, TDG, UNG and SMUG1 genes in familial colorectal cancer predisposition, BMC Cancer. 6 (2006) 243. doi:10.1186/1471-2407-6-243.
[349] W.-Q. Li, N. Hu, P.L. Hyland, Y. Gao, Z.-M. Wang, K. Yu, et al., Genetic variants in DNA repair pathway genes and risk of esophageal squamous cell carcinoma and gastric adenocarcinoma in a Chinese population, Carcinogenesis. 34 (2013) 1536–1542. doi:10.1093/carcin/bgt094.
[350] C. Marian, M. Tao, J.B. Mason, D.S. Goerlitz, J. Nie, A. Chanson, et al., Single nucleotide polymorphisms in uracil-processing genes, intake of one-carbon nutrients and breast cancer risk, Eur. J. Clin. Nutr. 65 (2011) 683–689. doi:10.1038/ejcn.2011.29.

[351] X. Yang, H. Zhu, Q. Qin, Y. Yang, Y. Yang, H. Cheng, et al., Genetic variants and risk of esophageal squamous cell carcinoma: A GWAS-based pathway analysis, Gene. (2014). doi:10.1016/j.gene.2014.11.049.
[352] H. Xie, Y. Gong, J. Dai, X. Wu, J. Gu, Genetic variations in base excision repair pathway and risk of bladder cancer: A case-control study in the United States, Mol. Carcinog. 54 (2015) 50–57. doi:10.1002/mc.22073.
[353] B. Pardini, F. Rosa, E. Barone, C. Di Gaetano, J. Slyskova, J. Novotny, et al., Variation within 3'-UTRs of base excision repair genes and response to therapy in colorectal cancer patients: A potential modulation of microRNAs binding, Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 19 (2013) 6044–6056. doi:10.1158/1078-0432.CCR-13-0314.
[354] A. Chanson, L.D. Parnell, E.D. Ciappio, Z. Liu, J.W. Crott, K.L. Tucker, et al., Polymorphisms in uracil-processing genes, but not one-carbon nutrients, are associated with altered DNA uracil concentrations in an urban Puerto Rican population, Am. J. Clin. Nutr. 89 (2009) 1927–1936. doi:10.3945/ajcn.2009.27429.
[355] M.C. Shin, S.J. Lee, J.E. Choi, S.I. Cha, C.H. Kim, W.K. Lee, et al., Glu346Lys polymorphism in the methyl-CpG binding domain 4 gene and the risk of primary lung cancer, Jpn. J. Clin. Oncol. 36 (2006) 483–488. doi:10.1093/jjco/hyl055.
[356] R. Miao, H. Gu, H. Liu, Z. Hu, G. Jin, H. Wang, et al., Tagging single nucleotide polymorphisms in MBD4 are associated with risk of lung cancer in a Chinese population, Lung Cancer Amst. Neth. 62 (2008) 281–286. doi:10.1016/j.lungcan.2008.03.027.
[357] B. Hao, H. Wang, K. Zhou, Y. Li, X. Chen, G. Zhou, et al., Identification of genetic variants in base excision repair pathway and their associations with risk of esophageal squamous cell carcinoma, Cancer Res. 64 (2004) 4378–4384. doi:10.1158/0008-5472.CAN-04-0372.
[358] J. Dong, Z. Hu, Y. Shu, S. Pan, W. Chen, Y. Wang, et al., Potentially functional polymorphisms in DNA repair genes and non-small-cell lung cancer survival: a pathway-based analysis, Mol. Carcinog. 51 (2012) 546–552. doi:10.1002/mc.20819.
[359] X.-D. Xiong, X.-P. Luo, X. Liu, X. Jing, L.-Q. Zeng, M. Lei, et al., The MBD4 Glu346Lys polymorphism is associated with the risk of cervical cancer in a Chinese population, Int. J. Gynecol. Cancer Off. J. Int. Gynecol. Cancer Soc. 22 (2012) 1552–1556. doi:10.1097/IGC.0b013e31826e22e4.
[360] J.H. Song, E.J. Maeng, Z. Cao, S.Y. Kim, S.W. Nam, J.Y. Lee, et al., The Glu346Lys polymorphism and frameshift mutations of the Methyl-CpG Binding Domain 4 gene in gastrointestinal cancer, Neoplasma. 56 (2009) 343–347.
[361] S.-F. Lo, L. Wan, C.-M. Huang, H.-C. Lin, S.-Y. Chen, S.-C. Liu, et al., Genetic polymorphisms of the DNA repair gene UNG are associated with the susceptibility of rheumatoid arthritis, Rheumatol. Int. 32 (2012) 3723–3727. doi:10.1007/s00296-011-2185-3.
[362] J. Yin, Y. Sang, L. Zheng, L. Wang, L. Yuan, C. Liu, et al., Uracil-DNA glycosylase (UNG) rs246079 G/A polymorphism is associated with decreased risk of esophageal cancer in a Chinese population, Med. Oncol. Northwood Lond. Engl. 31 (2014) 272. doi:10.1007/s12032-014-0272-5.
[363] J.A. Doherty, L.C. Sakoda, M.M. Loomis, M.J. Barnett, L. Julianto, M.D. Thornquist, et al., DNA repair genotype and lung cancer risk in the beta-carotene and retinol efficacy trial, Int. J. Mol. Epidemiol. Genet. 4 (2013) 11–34.
[364] J. Blasiak, E. Synowiec, A. Salminen, K. Kaarniranta, Genetic variability in DNA repair proteins in age-related macular degeneration, Int. J. Mol. Sci. 13 (2012) 13378–13397. doi:10.3390/ijms131013378.

[365] E. Synowiec, D. Wysokinski, M. Zaras, U. Kolodziejska, E. Stoczynska-Fidelus, K. Janik, et al., Association between polymorphism of the DNA repair SMUG1 and UNG genes and age-related macular degeneration, Retina Phila. Pa. 34 (2014) 38–47. doi:10.1097/IAE.0b013e31829477d8.
[366] H. Wei, A. Kamat, M. Chen, H.-L. Ke, D.W. Chang, J. Yin, et al., Association of polymorphisms in oxidative stress genes with clinical outcomes for bladder cancer treated with Bacillus Calmette-Guérin, PloS One. 7 (2012) e38533. doi:10.1371/journal.pone.0038533.
[367] A. Sjolund, A.A. Nemec, N. Paquet, A. Prakash, P. Sung, S. Doublié, et al., A germline polymorphism of thymine DNA glycosylase induces genomic instability and cellular transformation, PLoS Genet. 10 (2014) e1004753. doi:10.1371/journal.pgen.1004753.
[368] M. Wen-Bin, W. Wei, Q. Yu-Lan, J. Fang, X. Zhao-Lin, Micronucleus occurrence related to base excision repair gene polymorphisms in Chinese workers occupationally exposed to vinyl chloride monomer, J. Occup. Environ. Med. Am. Coll. Occup. Environ. Med. 51 (2009) 578–585. doi:10.1097/JOM.0b013e3181990d19.
[369] K. Curtin, C.M. Ulrich, W.S. Samowitz, R.K. Wolff, D.J. Duggan, K.W. Makar, et al., Candidate pathway polymorphisms in one-carbon metabolism and risk of rectal tumor mutations, Int. J. Mol. Epidemiol. Genet. 2 (2011) 1–8.
[370] A.Y. Liu, D. Scherer, E. Poole, J.D. Potter, K. Curtin, K. Makar, et al., Gene-diet-interactions in folate-mediated one-carbon metabolism modify colon cancer risk, Mol. Nutr. Food Res. 57 (2013) 721–734. doi:10.1002/mnfr.201200180.
[371] M. Krześniak, D. Butkiewicz, A. Samojedny, M. Chorazy, M. Rusin, Polymorphisms in TDG and MGMT genes - epidemiological and functional study in lung cancer patients from Poland, Ann. Hum. Genet. 68 (2004) 300–312. doi:10.1046/j.1529-8817.2004.00079.x.
[372] I. Ruczinski, T.J. Jorgensen, Y.Y. Shugart, Y.B. Schaad, B. Kessing, J. Hoffman-Bolton, et al., A population-based study of DNA repair gene variants in relation to non-melanoma skin cancer as a marker of a cancer-prone phenotype, Carcinogenesis. 33 (2012) 1692–1698. doi:10.1093/carcin/bgs170.
[373] B. Thyagarajan, B. Lindgren, S. Basu, S. Nagaraj, M.D. Gross, D.J. Weisdorf, et al., Association between genetic variants in the base excision repair pathway and outcomes after hematopoietic cell transplantations, Biol. Blood Marrow Transplant. J. Am. Soc. Blood Marrow Transplant. 16 (2010) 1084–1089. doi:10.1016/j.bbmt.2010.03.001.
[374] A. Allione, S. Guarrera, A. Russo, F. Ricceri, R. Purohit, A. Pagnani, et al., Inter-individual variation in nucleotide excision repair pathway is modulated by non-synonymous polymorphisms in ERCC4 and MBD4 genes, Mutat. Res. 751-752 (2013) 49–54. doi:10.1016/j.mrfmmm.2013.08.005.
[375] C.-M. Huang, P.-H. Huang, C.-L. Chen, L. Wan, C.-H. Tsai, S.-C. Liu, et al., MBD4 gene is associated with rheumatoid arthritis in Chinese patients in Taiwan, Rheumatol. Int. 32 (2012) 117–122. doi:10.1007/s00296-010-1545-8.
[376] J. Cheng, H. Zhang, C. Zhuang, R. Liu, Peptidylarginine deiminase type 4 and methyl-CpG binding domain 4 polymorphisms in Chinese patients with rheumatoid arthritis, J. Rheumatol. 39 (2012) 1159–1165. doi:10.3899/jrheum.120007.
[377] H. Zhao, W. Du, D. Wang, D. Gu, F. Xue, J. Ge, et al., Single nucleotide polymorphism in the methyl-CpG binding domain 4 gene and the risk for immune thrombocytopenic purpura in Chinese population, Platelets. 21 (2010) 132–136. doi:10.3109/09537100903474365.
[378] M. Goulian, B. Bleile, B.Y. Tseng, Methotrexate-induced misincorporation of uracil into DNA, Proc. Natl. Acad. Sci. U. S. A. 77 (1980) 1956–1960.

[379] M. Goulian, B. Bleile, B.Y. Tseng, The effect of methotrexate on levels of dUTP in animal cells, J. Biol. Chem. 255 (1980) 10630–10637.
[380] H.A. Ingraham, L. Dickey, M. Goulian, DNA fragmentation and cytotoxicity from increased cellular deoxyuridylate, Biochemistry (Mosc.). 25 (1986) 3225–3230.
[381] L.D. Weeks, P. Fu, S.L. Gerson, Uracil-DNA glycosylase expression determines human lung cancer cell sensitivity to pemetrexed, Mol. Cancer Ther. 12 (2013) 2248–2260. doi:10.1158/1535-7163.MCT-13-0172.
[382] S.J. Welsh, S. Hobbs, G.W. Aherne, Expression of uracil DNA glycosylase (UDG) does not affect cellular sensitivity to thymidylate synthase (TS) inhibition, Eur. J. Cancer Oxf. Engl. 1990. 39 (2003) 378–387.
[383] L.D. Weeks, G.E. Zentner, P.C. Scacheri, S.L. Gerson, Uracil DNA glycosylase (UNG) loss enhances DNA double strand break formation in human cancer cells exposed to pemetrexed, Cell Death Dis. 5 (2014) e1045. doi:10.1038/cddis.2013.477.
[384] S.W. Choi, J.B. Mason, Folate and carcinogenesis: an integrated scheme, J. Nutr. 130 (2000) 129–132.
[385] E.S. Schernhammer, S. Ogino, C.S. Fuchs, Folate and vitamin B6 intake and risk of colon cancer in relation to p53 expression, Gastroenterology. 135 (2008) 770–780. doi:10.1053/j.gastro.2008.06.033.
[386] B.F. Cole, J.A. Baron, R.S. Sandler, R.W. Haile, D.J. Ahnen, R.S. Bresalier, et al., Folic acid for the prevention of colorectal adenomas: a randomized clinical trial, JAMA. 297 (2007) 2351–2359. doi:10.1001/jama.297.21.2351.
[387] K.J. Rycyna, D.J. Bacich, D.S. O'Keefe, Opposing roles of folate in prostate cancer, Urology. 82 (2013) 1197–1203. doi:10.1016/j.urology.2013.07.012.
[388] L.F. Petersen, N.T. Brockton, A. Bakkar, S. Liu, J. Wen, A.M. Weljie, et al., Elevated physiological levels of folic acid can increase in vitro growth and invasiveness of prostate cancer cells, BJU Int. 109 (2012) 788–795. doi:10.1111/j.1464-410X.2011.10437.x.
[389] H.G. Linhart, A. Troen, G.W. Bell, E. Cantu, W.-H. Chao, E. Moran, et al., Folate deficiency induces genomic uracil misincorporation and hypomethylation but does not increase DNA point mutations, Gastroenterology. 136 (2009) 227–235.e3. doi:10.1053/j.gastro.2008.10.016.
[390] K.-C. Kim, H. Jang, J. Sauer, E.M. Zimmerly, Z. Liu, A. Chanson, et al., Folate supplementation differently affects uracil content in DNA in the mouse colon and liver, Br. J. Nutr. 105 (2011) 688–693. doi:10.1017/S0007114510004332.
[391] R. Küppers, Mechanisms of B-cell lymphoma pathogenesis, Nat. Rev. Cancer. 5 (2005) 251–262. doi:10.1038/nrc1589.
[392] R. Küppers, R. Dalla-Favera, Mechanisms of chromosomal translocations in B cell lymphomas, Oncogene. 20 (2001) 5580–5594. doi:10.1038/sj.onc.1204640.
[393] Y. Dorsett, D.F. Robbiani, M. Jankovic, B. Reina-San-Martin, T.R. Eisenreich, M.C. Nussenzweig, A role for AID in chromosome translocations between c-myc and the IgH variable region, J. Exp. Med. 204 (2007) 2225–2232. doi:10.1084/jem.20070884.
[394] R. Chiarle, Y. Zhang, R.L. Frock, S.M. Lewis, B. Molinie, Y.-J. Ho, et al., Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells, Cell. 147 (2011) 107–119. doi:10.1016/j.cell.2011.07.049.
[395] A. Kotani, N. Kakazu, T. Tsuruyama, I. Okazaki, M. Muramatsu, K. Kinoshita, et al., Activation-induced cytidine deaminase (AID) promotes B cell lymphomagenesis in Emu-cmyc transgenic mice, Proc. Natl. Acad. Sci. 104 (2007) 1616–1620. doi:10.1073/pnas.0610732104.

[396] L. Pasqualucci, G. Bhagat, M. Jankovic, M. Compagno, P. Smith, M. Muramatsu, et al., AID is required for germinal center-derived lymphomagenesis, Nat. Genet. 40 (2008) 108–112. doi:10.1038/ng.2007.35.
[397] M. Liu, J.L. Duke, D.J. Richter, C.G. Vinuesa, C.C. Goodnow, S.H. Kleinstein, et al., Two levels of protection for the B cell genome during somatic hypermutation, Nature. 451 (2008) 841–845. doi:10.1038/nature06547.
[398] J. Greeve, A. Philipsen, K. Krause, W. Klapper, K. Heidorn, B.E. Castle, et al., Expression of activation-induced cytidine deaminase in human B-cell non-Hodgkin lymphomas, Blood. 101 (2003) 3574–3580. doi:10.1182/blood-2002-08-2424.
[399] M.S. Hardianti, E. Tatsumi, M. Syampurnawati, K. Furuta, K. Saigo, S. Kawano, et al., Expression of activation-induced cytidine deaminase (AID) in Burkitt lymphoma cells: rare AID-negative cell lines with the unmutated rearranged VH gene, Leuk. Lymphoma. 45 (2004) 155–160.
[400] L.A. Smit, R.J. Bende, J. Aten, J.E.J. Guikema, W.M. Aarts, C.J.M. van Noesel, Expression of activation-induced cytidine deaminase is confined to B-cell non-Hodgkin's lymphomas of germinal-center phenotype, Cancer Res. 63 (2003) 3894–3898.
[401] S. Nik-Zainal, L.B. Alexandrov, D.C. Wedge, P. Van Loo, C.D. Greenman, K. Raine, et al., Mutational processes molding the genomes of 21 breast cancers, Cell. 149 (2012) 979–993. doi:10.1016/j.cell.2012.04.024.
[402] A.G. Lada, A. Dhar, R.J. Boissy, M. Hirano, A.A. Rubel, I.B. Rogozin, et al., AID/APOBEC cytosine deaminase induces genome-wide kataegis, Biol. Direct. 7 (2012) 47; discussion 47. doi:10.1186/1745-6150-7-47.
[403] M.B. Burns, L. Lackey, M.A. Carpenter, A. Rathore, A.M. Land, B. Leonard, et al., APOBEC3B is an enzymatic source of mutation in breast cancer, Nature. 494 (2013) 366–370. doi:10.1038/nature11881.
[404] B.J. Taylor, S. Nik-Zainal, Y.L. Wu, L.A. Stebbings, K. Raine, P.J. Campbell, et al., DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis, eLife. 2 (2013) e00534. doi:10.7554/eLife.00534.
[405] L.B. Alexandrov, S. Nik-Zainal, D.C. Wedge, S.A.J.R. Aparicio, S. Behjati, A.V. Biankin, et al., Signatures of mutational processes in human cancer, Nature. 500 (2013) 415–421. doi:10.1038/nature12477.
[406] N. Bolli, H. Avet-Loiseau, D.C. Wedge, P. Van Loo, L.B. Alexandrov, I. Martincorena, et al., Heterogeneity of genomic evolution and mutational profiles in multiple myeloma, Nat. Commun. 5 (2014) 2997. doi:10.1038/ncomms3997.
[407] Z. Chen, J.H. Wang, Generation and repair of AID-initiated DNA lesions in B lymphocytes, Front. Med. 8 (2014) 201–216. doi:10.1007/s11684-014-0324-4.
[408] C.F. Davis, C.J. Ricketts, M. Wang, L. Yang, A.D. Cherniack, H. Shen, et al., The somatic genomic landscape of chromophobe renal cell carcinoma, Cancer Cell. 26 (2014) 319–330. doi:10.1016/j.ccr.2014.07.014.
[409] M. Hoogstraat, M.S. de Pagter, G.A. Cirkel, M.J. van Roosmalen, T.T. Harkins, K. Duran, et al., Genomic and transcriptomic plasticity in treatment-na?ve ovarian cancer, Genome Res. 24 (2014) 200–211. doi:10.1101/gr.161026.113.
[410] C.D. Mol, A.S. Arvai, R.J. Sanderson, G. Slupphaug, B. Kavli, H.E. Krokan, et al., Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA, Cell. 82 (1995) 701–708.
[411] H. Nilsen, K.A. Haushalter, P. Robins, D.E. Barnes, G.L. Verdine, T. Lindahl, Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase, EMBO J. 20 (2001) 4278–4286. doi:10.1093/emboj/20.15.4278.

[412] H. Nilsen, K.S. Steinsbekk, M. Otterlei, G. Slupphaug, P.A. Aas, H.E. Krokan, Analysis of uracil-DNA glycosylases from the murine Ung gene reveals differential expression in tissues and in embryonic development and a subcellular sorting pattern that differs from the human homologues, Nucleic Acids Res. 28 (2000) 2277–2285.
[413] S.T. Mashiyama, C.M. Hansen, E. Roitman, S. Sarmiento, J.E. Leklem, T.D. Shultz, et al., An assay for uracil in human DNA at baseline: effect of marginal vitamin B6 deficiency, Anal Biochem. 372 (2008) 21–31. doi:10.1016/j.ab.2007.08.034.
[414] A. Chango, A.M. Abdel Nour, C. Niquet, F.J. Tessier, Simultaneous determination of genomic DNA methylation and uracil misincorporation, Med Princ Pr. 18 (2009) 81–4. doi:10.1159/000189803.
[415] A.D. Bulgar, L.D. Weeks, Y. Miao, S. Yang, Y. Xu, C. Guo, et al., Removal of uracil by uracil DNA glycosylase limits pemetrexed cytotoxicity: overriding the limit with methoxyamine to inhibit base excision repair, Cell Death Dis. 3 (2012) e252. doi:10.1038/cddis.2011.135.
[416] M.B. Burns, N.A. Temiz, R.S. Harris, Evidence for APOBEC3B mutagenesis in multiple human cancers, Nat. Genet. 45 (2013) 977–983. doi:10.1038/ng.2701.
[417] L. Klemm, C. Duy, I. Iacobucci, S. Kuchen, G. von Levetzow, N. Feldhahn, et al., The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia, Cancer Cell. 16 (2009) 232–245. doi:10.1016/j.ccr.2009.07.030.
[418] Z. Xu, Z. Fulop, G. Wu, E.J. Pone, J. Zhang, T. Mai, et al., 14-3-3 adaptor proteins recruit AID to 5'-AGCT-3'-rich switch regions for class switch recombination, Nat. Struct. Mol. Biol. 17 (2010) 1124–1135. doi:10.1038/nsmb.1884.
[419] H. Zan, C.A. White, L.M. Thomas, T. Mai, G. Li, Z. Xu, et al., Rev1 recruits ung to switch regions and enhances du glycosylation for immunoglobulin class switch DNA recombination, Cell Rep. 2 (2012) 1220–1232. doi:10.1016/j.celrep.2012.09.029.
[420] S.T. Mashiyama, C. Courtemanche, I. Elson-Schwab, J. Crott, B.L. Lee, C.N. Ong, et al., Uracil in DNA, determined by an improved assay, is increased when deoxynucleosides are added to folate-deficient cultured human lymphocytes, Anal Biochem. 330 (2004) 58–69. doi:10.1016/j.ab.2004.03.065.
[421] A.R. Weber, D. Schuermann, P. Schär, Versatile Recombinant SUMOylation System for the Production of SUMO-Modified Protein, PLoS ONE. 9 (2014) e102157. doi:10.1371/journal.pone.0102157.
[422] R.W. Maul, H. Saribasak, S.A. Martomo, R.L. McClure, W. Yang, A. Vaisman, et al., Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions, Nat. Immunol. 12 (2011) 70–76. doi:10.1038/ni.1970.
[423] D.S. Bryan, M. Ransom, B. Adane, K. York, J.R. Hesselberth, High resolution mapping of modified DNA nucleobases using excision repair enzymes, Genome Res. 24 (2014) 1534–1542. doi:10.1101/gr.174052.114.
[424] A. Horváth, B.G. Vértessy, A one-step method for quantitative determination of uracil in DNA by real-time PCR, Nucleic Acids Res. 38 (2010) e196. doi:10.1093/nar/gkq815.

Article I

Is not included due to copyright

Article II

DNA Repair 12 (2013) 699–706

Contents lists available at SciVerse ScienceDirect

DNA Repair

journal homepage: www.elsevier.com/locate/dnarepair

A robust, sensitive assay for genomic uracil determination by LC/MS/MS reveals lower levels than previously reported☆

Anastasia Galashevskaya[1], Antonio Sarno[1], Cathrine B. Vågbø, Per A. Aas, Lars Hagen, Geir Slupphaug, Hans E. Krokan*

Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, NO-7489 Trondheim, Norway

ARTICLE INFO

Article history:
Received 18 February 2013
Received in revised form 6 May 2013
Accepted 9 May 2013
Available online 3 June 2013

Keywords:
Uracil in DNA
Uracil DNA glycosylase
DNA damage
Base excision repair
Adaptive immunity
Activation-induced cytidine deaminase

ABSTRACT

Considerable progress has been made in understanding the origins of genomic uracil and its role in genome stability and host defense; however, the main question concerning the basal level of uracil in DNA remains disputed. Results from assays designed to quantify genomic uracil vary by almost three orders of magnitude. To address the issues leading to this inconsistency, we explored possible shortcomings with existing methods and developed a sensitive LC/MS/MS-based method for the absolute quantification of genomic 2′-deoxyuridine (dUrd). To this end, DNA was enzymatically hydrolyzed to 2′-deoxyribonucleosides and dUrd was purified in a preparative HPLC step and analyzed by LC/MS/MS. The standard curve was linear over four orders of magnitude with a quantification limit of 5 fmol dUrd. Control samples demonstrated high inter-experimental accuracy (94.3%) and precision (CV 9.7%). An alternative method that employed UNG2 to excise uracil from DNA for LC/MS/MS analysis gave similar results, but the intra-assay variability was significantly greater. We quantified genomic dUrd in $Ung^{+/+}$ and $Ung^{-/-}$ mouse embryonic fibroblasts and human lymphoblastoid cell lines carrying *UNG* mutations. DNA-dUrd is 5-fold higher in $Ung^{-/-}$ than in $Ung^{+/+}$ fibroblasts and 11-fold higher in UNG2 dysfunctional than in UNG2 functional lymphoblastoid cells. We report approximately 400–600 dUrd per human or murine genome in repair-proficient cells, which is lower than results using other methods and suggests that genomic uracil levels may have previously been overestimated.

1. Introduction

Deamination of 2′-deoxycytidine (dCyd) and misincorporation of 2′-deoxyuridine 5′-monophosphate (dUMP) are the major sources of 2′-deoxyuridine (dUrd)/uracil (U) in the mammalian genome [1]. The former creates U:G mismatches and occurs spontaneously, mainly *via* direct nucleophilic attack of the hydroxyl ion on the protonated base under physiological conditions, by exposure to various chemicals, or enzymatically by activation induced cytidine deaminase (AID), APOBEC1, and possibly other members in the APOBEC family [2,3]. Unrepaired U:G mismatches result in C to T transitions during replication, the most frequent type of mutation in human cancers [4]. Alternatively, dUMP misincorporation creates U:A pairs, depends on the dTTP/dUTP ratio at the time of DNA replication, and is governed by thymidylate synthase and dUTPase activities [5]. U:A pairs may be cytotoxic due to altered binding of transcription factors and indirectly mutagenic through generation of abasic sites [6–9].

Genomic dUrd is generally treated as a lesion that can be corrected by base excision repair with mismatch repair as a likely backup for U:G mismatches [1,10,11]. Nevertheless, dUrd is also a key intermediate in adaptive immunity. In this process, dUrd is generated by AID-mediated dCyd deamination, which targets variable and switch regions of immunoglobulin genes in B-cells during somatic hypermutation (SHM) and class switch recombination (CSR), respectively [12]. This is a risky process because off-target deamination may cause mutations and translocations, ultimately culminating in B-cell lymphomas [13–15]. Importantly, the translocations occur at the DNA damage sites [16]. Furthermore, infection- and/or inflammatory cytokine-driven AID expression may contribute to carcinogenesis in epithelial cells [17–19].

The emerging significance of genomic uracil thus calls for an accurate and reliable method for its quantification. Most established methods are relative, which precludes comparisons between experimental batches and different laboratories [6,12,20–25].

Abbreviations: LC/MS/MS, liquid chromatography coupled to tandem mass spectrometry; UNG, uracil-DNA glycosylase encoded by the *UNG*-gene; dCyd/dUrd/Dn, 2′-deoxycytidine/2′-deoxyuridine/2′-deoxyribonucleoside.

* Corresponding author. Tel.: +47 72573074; fax: +47 72576400.
E-mail address: hans.krokan@ntnu.no (H.E. Krokan).
[1] Joint first authors.

1568-7864/$ – see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.dnarep.2013.05.002

Direct quantification of absolute levels of genomic uracil can be achieved using mass spectrometry. There are two main approaches: detection of U excised from DNA by UNG and detection of dUrd after enzymatic hydrolysis of DNA to 2′-deoxyribonucleosides (dNs) [26–32]. Both strategies are seemingly straightforward, but a wide variation in estimates has been reported, ranging from 3×10^3 to 4×10^6 uracils per mammalian genome [31,33]. It has been suggested that the inconsistency in reported genomic uracil levels may be due to differences in sample type, but may also emanate from technical shortcomings of the employed methods [33].

Here we present an improved LC/MS/MS-based method for the absolute quantification of dUrd in DNA and discuss drawbacks of related methods. We explore the issues that may lead to inaccurate estimation of genomic U and ameliorate them by introducing steps for specimen clean-up and chromatographic modifications. Additionally, we compare dUrd quantification by DNA hydrolysis to U quantification by UNG excision. We lastly applied our method to quantify genomic dUrd in *Ung*$^{+/+}$ and *Ung*$^{-/-}$ mouse embryonic fibroblasts, as well as human lymphoblastoid cell lines derived from hyper-IgM patients carrying *UNG* mutations. We measured genomic uracil values lower than those previously reported, indicating that previous methods may have overestimated genomic uracil.

2. Material and methods

2.1. Reagents

2′-Deoxyuridine, 2′-deoxycytidine, 2′-deoxyadenosine, 2′-deoxyguanosine, thymidine, alkaline phosphatase, nuclease P1, and BSA were from Sigma–Aldrich (Steinheim, Germany); DNase I was from Roche Applied Science (Mannheim, Germany); UltraPure™ salmon sperm DNA was from Invitrogen Corporation (Carlsbad, CA, USA). Recombinant uracil-DNA glycosylase (UNGΔ84) was purified in-house as previously described [34]. [2-^{13}C;1,3-$^{15}N_2$]-2′-deoxyuridine was from C/D/N Isotopes (Pointe-Claire, Quebec, Canada).

2.2. Cell lines

Ung$^{+/+}$ and *Ung*$^{-/-}$ mouse embryonic fibroblast cell lines [35] were cultured in DMEM 4500 mg/l D-glucose, supplemented with 0.29 mg/ml L-glutamine, 10% fetal bovine serum, 100 U/ml penicillin, 0.1 μg/ml streptomycin, and 2.5 μ/ml amphotericin B in a humidified 5% CO_2 incubator at 37 °C. Epstein–Barr virus immortalized human lymphoblastoid cell lines [36], a gift from Dr. Anne Durandy (Institut National de la Santé et de la Recherche Médicale, Paris, France), were cultured in RPMI–1640 + GlutaMax™-I medium supplemented with 10% heat-inactivated bovine serum, 100 U/ml penicillin, 0.1 mg/ml streptomycin, and 2.5 μg/ml amphotericin B.

2.3. DNA isolation and removal of intracellular 2′-deoxyribonucleotides

Cells (10^6/80 μl) were lysed in 10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.5% SDS, 2.5 mM DTT, 0.25 μg/μl proteinase K, 0.1 μg/μl RNase A and incubated at 37 °C for 1 h with shaking at 250 rpm. Genomic DNA was extracted in phenol:chloroform:isoamyl alcohol (25:24:1) and chloroform:isoamyl alcohol (24:1), then precipitated by adding 0.3 volume equivalents of 10 M ammonium acetate (pH 7.9) and one volume equivalent of 100% isopropanol, washed once in 70% ethanol, and buffered with 100 mM ammonium bicarbonate (pH 7.6) and 1 mM $MgCl_2$. Where indicated, DNA was isolated from cell pellets using the DNeasy® Blood and Tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions except for increasing the RNase A concentration to 0.1 μg/μl and decreasing the temperature during incubation with AL buffer from 56 to 37 °C. Potentially co-isolated intracellular 2′-deoxyribonucleotides were dephosphorylated by incubation with alkaline phosphatase (pH 7.6) from *Escherichia coli* (0.2 U/μl) in 100 mM ammonium bicarbonate for 30 min and DNA precipitated with isopropanol as described above.

2.4. DNA hydrolysis to 2′-deoxyribonucleosides

DNA was enzymatically hydrolyzed to dNs. Prior to hydrolysis, a control DNA sample was deuracilated by treatment with UNG to control for uracil generated *in vitro* during the assay. To this end, up to 15 μg DNA were buffered with 20 mM Tris-HCl (pH 7.5), 60 mM NaCl, 1 mM DTT, 1 mg/ml BSA in a reaction volume of 30 μl and treated with 0.075 U UNGΔ84 at 37 °C for 1 h. The DNA was isopropanol precipitated as described in 2.3 and resuspended in 30 μl 100 mM ammonium acetate (pH 6.0), 10 mM $MgCl_2$, and 1 mM $CaCl_2$ containing 2 U DNase I and 0.2 U nuclease P1 and incubated for 30 min at 37 °C. As an internal standard [2-^{13}C;1,3-$^{15}N_2$]-2′-deoxyuridine was used. The samples were then buffered in ammonium bicarbonate (pH 7.6) to a final concentration of 100 mM, and incubated for 20 min at 37 °C with 0.1 U alkaline phosphatase from *E. coli*. To precipitate contaminants that could potentially clog the HPLC column, three volume equivalents of ice-cold acetonitrile were added to the samples, which were then centrifuged (16,100 × *g*, 20 min, 4 °C). The supernatants were transferred to new tubes and vacuum centrifuged until dry at room temperature. The samples were finally dissolved in 100 μl water containing 10% acetonitrile.

2.5. Preparative purification of 2′-deoxyuridine

dUrd was purified by HPLC prior to LC/MS/MS analysis. The purification was performed using a reverse-phase column with weak acidic ion-pairing groups (Primesep 200, 2.1 mm × 150 mm, 5 μm, SIELC Technologies, Prospect Heights, IL), kept at 35 °C, on an Agilent 1200 series HPLC system, equipped with a G1365D multiple wavelength detector (Agilent Technologies, Waldbronn, Germany). Samples were maintained at 4 °C prior to injection. Each sample was injected in triplicate with an injection volume of 30 μl. The gradient used consisted of solvent A (water, 0.1% formic acid) and B (methanol, 0.1% formic acid) starting at 10% B for 0.5 min, ramping to 60% B over 6 min, holding at 60% B for 4 min, and re-equilibrating with 10% B for 10 min at a flow rate of 0.200 ml/min. dNs were quantified by measuring absorption at 260 nm. The fractions containing dUrd and IS were collected ±1 min with a Foxy R2 fraction collection system (Teledyne ISCO, Lincoln, NE, USA) and vacuum centrifuged until dry at room temperature. The samples were dissolved in 25 μl water containing 5% methanol.

2.6. Uracil excision

Uracil was excised from DNA for direct analysis by LC/MS/MS to compare uracil excision with DNA hydrolysis as in an alternative strategy for DNA-uracil quantification. The uracil excision and quantification protocol was modified from Bulgar et al. [26]. Up to 15 μg DNA were buffered with 20 mM Tris-HCl (pH 7.5), 10 mM NaCl, 1 mM DTT, 1 mg/ml BSA in a reaction volume of 40 μl and treated with 0.075 U UNGΔ84 at 37 °C for 1 h. The NaCl concentration used was different from that used for DNA deuracilation described above to avoid signal loss by ion suppression during LC/MS/MS. [2-^{13}C,$^{15}N_2$]-Uracil was used as internal standard. After incubation with UNG, 500 μl ice-cold acetonitrile were added to the

samples and they were then centrifuged (16,100 × *g*, 20 min, 4 °C). The supernatants were transferred to new tubes and vacuum centrifuged until dry at room temperature. The samples were finally dissolved in 40 μl 10% 2 mM ammonium formate 90% acetonitrile.

2.7. LC/MS/MS instrumentation and conditions

Both dUrd and uracil were quantified using an LC-20AD HPLC system (Shimadzu Corporation, Kyoto, Japan) coupled to an API 5000 triple-quadrupole mass spectrometer (Applied Biosystems, Carlsbad, CA, USA) operated under the multiple reaction monitoring (MRM) mode.

dUrd was quantified using a Zorbax SB-C18 reverse phase column at room temperature (2.1 mm × 150 mm, 3.5 μm, Agilent Technologies, Santa Clara, CA, USA), protected with a Zorbax Reliance guard-column (4.6 mm × 12.5 mm, Agilent Technologies). The injection volume was 20 μl. The gradient used contained solvent A (water, 0.1% formic acid) and B (methanol, 0.1% formic acid) starting at 5% B for 0.5 min, ramping to 90% B over 6 min, holding at 90% B for 1.5 min and re-equilibrating with 5% B for 5 min at a flow rate of 0.300 ml/min. Mass spectrometry detection was performed using positive electrospray ionization, monitoring the mass transitions 229.2 → 113.0 and 232.2 → 116.0 for 2′-deoxyuridine and [2-^{13}C,1,3-^{15}N$_2$]-2′-deoxyuridine, respectively.

For the alternative uracil-release method, uracil was quantified using a hydrophobic interaction liquid chromatography column (2.1 mm × 100 mm, 3.5 μm, Atlantis HILIC Silica column, Waters Corporation, Milford, MA, USA). The injection volume was 10 μl and the HPLC was run at 0.200 ml/min isocratically with 95% acetonitrile and 5% 2 mM ammonium formate. Detection was performed using negative electrospray ionization, monitoring the mass transitions 110.9 → 41.9 and 114.1 → 43.9 for uracil and [2-^{13}C,1,3-^{15}N$_2$]-uracil, respectively.

3. Results

3.1. Method development

3.1.1. MS/MS analysis

We used tandem mass spectrometry to validate our method's specificity. MS/MS spectra revealed ions with m/z values of 113.0 and 116.0, which correspond to the uracil and isotopically labeled base in the internal standard (IS), respectively. The *m*/*z* values 117.0, 99.0, and 81.1 were found in both dUrd and IS and correspond to 2-deoxyribose and 2-deoxyribose without one or two water molecules, respectively (Fig. 1A).

3.1.2. A precursory HPLC step is essential for sample purity

The analysis of dUrd is complicated by naturally occurring [^{13}C]-2′-deoxycytidine ([^{13}C]-dCyd), which is isobaric with dUrd. Although dUrd and dCyd are apparently well separated by reverse-phase chromatography, the relative abundance of dCyd over dUrd in DNA is so high that the [^{13}C]-dCyd peak tail (~1.1% of all carbon) will obfuscate the dUrd peak, consequently interfering with the subsequent MS analysis. To circumvent this problem, we used a reverse-phase column with embedded weak acidic ion-pairing groups (hereafter referred to by its brand name, Primesep 200), with which dUrd elutes well before dCyd (Fig. 1B). However, dUrd is weakly retained in the column and elutes near or with the void volume, resulting in ion suppression from ions present in the reaction buffer, which compete for ionization with the analyte of interest (dUrd, data not shown). To avoid this, we employed a precursory HPLC step with a Primesep 200 column to rid the samples of dCyd and increase sensitivity and then analyzed the dUrd concentration with a reverse-phase C18 column coupled to a mass spectrometer. We also tested a standard C18 column for the precursory HPLC step, but found that enough dCyd co-eluted with the dUrd fraction that dCyd deamination occurred

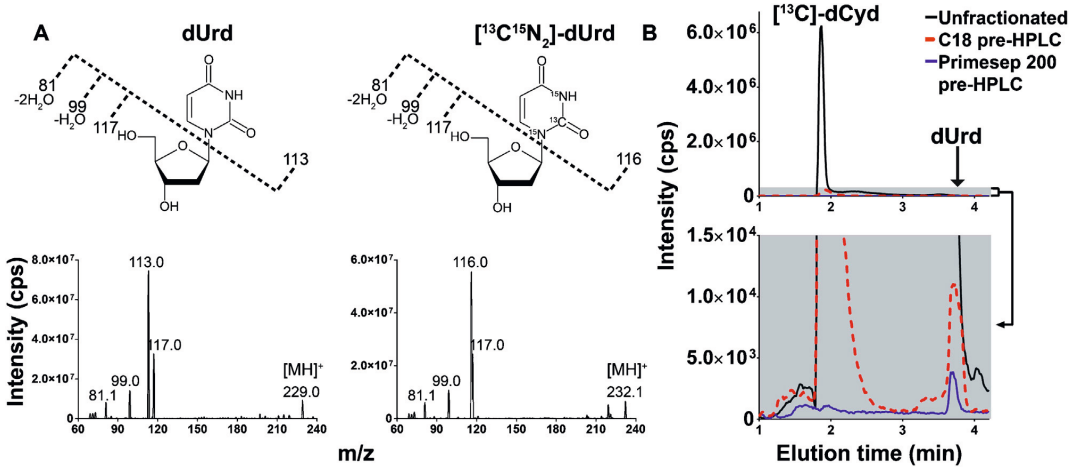


Fig. 1. Optimized LC/MS/MS conditions ensure method specificity. (A) MS/MS spectra of 2′-deoxyuridine (*m*/*z* 229) and [2-^{13}C,1,3-^{15}N]-2′-deoxyuridine (*m*/*z* 232.1) showing parent [MH]$^+$ to product ion transitions. The proposed origins of key fragments are indicated. Note that the collision energy was tuned to acquire spectra with more fragments to demonstrate the fragmentation pattern of dUrd. The final settings were optimized for the specific mass transitions analyzed. (B) Effect of precursory HPLC step with PrimeSep 200 and standard reverse phase C18 columns on LC/MS/MS chromatograms. Note that both chromatograms represent the same data displayed with a different *y*-axis scale. In the lower panel, the range to 1.5 × 10^4 has been expanded to visualize chromatographic tailing and the problems related to [^{13}C]-dCyd when using C18 column for pre-HPLC. The dUrd peak is obscured by the [^{13}C]-dCyd peak tail in the absence of fractionation (dashed red line). Using both C18 (solid black line) and Primesep 200 (solid blue line) columns overcome dUrd peak obfuscation by the [^{13}C]-dCyd peak, but the C18 column retains some dCyd, leading to a [^{13}C]-dCyd peak in the LC/MS/MS step as well as a higher dUrd peak, probably due to dCyd deamination.

Table 1
Summary of statistics for method validation. Deuracilated salmon sperm DNA samples were spiked with 5, 15, and 100 fmol dUrd to determine accuracy and intra/inter-day precision.

dUrd spike (fmol)	Accuracy (% theoretical value)	Intra-day precision (CV %)	Inter-day precision (CV %)
5	94.0	13.6	15.0
15	97.0	13.1	12.4
100	91.9	2.6	2.6
Mean	94.3	9.7	10.0
n	18	6	18

between the precursory HPLC step and the LC/MS/MS analysis (Fig. 1B).

An additional advantage of employing a precursory HPLC step is that it provides a convenient opportunity to quantify all dNs prior to LC/MS/MS analysis, thereby allowing accurate quantification of dUrd per dNs. We compared the DNA concentration measured by spectrophotometry of 5 μg salmon sperm DNA with the calculated concentration by HPLC-UV on three separate days and found 99.9% recovery after hydrolysis with a CV of 8.34%.

3.1.3. Determination of range, linearity, detection limit, precision, and accuracy

We determined the range, linearity, and detection limit for dUrd quantifications by making standard curves in both water and deuracilated DNA. Triplicate standard curves of dUrd in water containing 5–200 fmol dUrd and 40 fmol IS were analyzed on three different days ($r^2 = 1.0000$), demonstrating near perfect linearity (Supplementary Fig. 1). Deuracilated DNA prepared by UNG-treatment and isopropanol precipitation of 5 μg salmon sperm DNA was spiked with 5, 15, and 100 fmol dUrd and assayed in sets of six replicates. The mean observed accuracy for these samples was 94.3%, and the intra- and inter-day CV values were 9.7 and 10%, respectively. The lower limit of quantification was found to be 5 fmol dUrd (CV 15% $n = 18$). The data are summarized in Table 1.

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.dnarep.2013.05.002.

3.1.4. Sample contamination with intracellular 2′-deoxyribonucleotides causes overestimation of genomic dUrd

We tested whether cellular dUMP and dCMP could possibly interfere with dUrd analysis due to co-purification with DNA. Importantly, dCMP and dCyd (as well as dCyd in ssDNA) are deaminated more than two orders of magnitude faster than dCyd in dsDNA [3]. To this end, we pre-treated DNA samples with alkaline phosphatase and then precipitated the DNA. Our hypothesis was that dUMP and dCMP would co-purify with DNA to a larger extent than dUrd. Indeed, we found that up to 98% of measured dUrd in commercially prepared DNA was removed after phosphatase treatment and precipitation (Fig. 2A). DNA isolated in our laboratory showed similar results (data not shown).

3.1.5. Overcoming dCyd deamination during sample work-up

Three main factors have been demonstrated to affect cytosine deamination in purified DNA samples: temperature, pH, and the degree to which DNA is denatured [3,37,38]. Taking this into consideration, we made efforts to minimize dCyd deamination during sample work-up and analysis. Several methods used by other laboratories involve heat-denaturation of DNA prior to enzymatic hydrolysis [27]. We found that DNA denaturation by heating to 95 °C for 5–20 min increases the dUrd signal approximately 1.7- and 2.7-fold, respectively (Fig. 2B). To avoid deamination during work-up and analyses, we optimized reaction time and buffer conditions, concluding with enzymatic hydrolysis at pH 6–7.6 and 37 °C for 50 min using DNase I, nuclease P1, and alkaline phosphatase. To test the rate at which dUrd is introduced under these reaction conditions, we assayed the amount of dUrd generated during sample analysis over time. We found a constant deamination rate of $4.805 \times 10^{-3} \pm 5.9 \times 10^{-5}$ dUrd/10^6 bp/min ($R^2 = 0.9964$, $n = 12$, Fig. 2C). This corresponds to 1.059×10^{-2} dUrd/10^6 dCyd/min, which is in line with previously reported values of dCyd deamination rates of 2.6×10^{-2}, 1.2×10^{-3}, and 4.8×10^{-5} dUrd/10^6 dCyd/min for deoxyribonucleosides, single-stranded DNA, and double-stranded DNA, respectively [3]. Subtracting the deuracilated DNA control from the normal samples yielded a constant value regardless of the time point (0.66 dUrd/10^6 bp); however, the variation between replicate experiments increased with reaction time due to increasing background. Thus, we included a deuracilated DNA control in all sample batches to control for *in vitro*-generated dUrd.

It has been reported that alkaline phosphatase contained measurable dCyd deaminase activity [28,29]. We substituted dCyd for DNA to the equivalent of ~2 μg (10.5 nmol) and carried out mock hydrolysis with all enzymes, only alkaline phosphatase, and no enzymes. The amount of dUrd per dCyd in the untreated samples was statistically indistinguishable from that of the samples containing either all enzymes or only alkaline phosphatase (data not shown), which strongly suggests that none of the enzyme preparations employed contained dCyd deaminase activity under our reaction conditions. We therefore did not employ dCyd deaminase inhibitors.

3.1.6. dUrd quantification by DNA hydrolysis is more robust than U quantification by U excision

Several groups have employed UNG to excise uracil for GC or LC/MS analysis [26,32]. To compare this strategy to the dUrd method, we used UNG to excise U from DNA and measured U by a hydrophobic interaction chromatography column coupled to the same mass spectrometer used for dUrd quantification. First, we spiked U into deuracilated DNA and determined that the limit of quantification for this assay was 5 fmol. Then, we measured genomic U in DNA that had been heated to 95 °C. The results were similar to those obtained by DNA hydrolysis (Fig. 2B). We also assayed genomic uracil using both the DNA hydrolysis and U excision on DNA isolated using either phenol:chloroform:isoamyl alcohol isolation or a column-based kit (Supplementary Fig. 2). The level of genomic dUrd was similar regardless of the DNA isolation method when assayed using the DNA hydrolysis method, but significantly different in UNG2 deficient cells when using the U excision method ($P = 0.0275$, $n = 3$). This indicates that DNA hydrolysis is both more robust and reproducible than the U excision method.

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.dnarep.2013.05.002.

3.2. Genomic uracil in human and mouse cells proficient or deficient in UNG-activity

We tested the biological applicability of our method by comparing the levels of genomic dUrd in mammalian cell lines. First, we compared two lymphoblastoid cell lines: one with UNG-deficiency derived from a patient with a homozygous mutation substituting Ser with Phe (UNG2-F251S) and one with functional UNG derived from an individual with a heterozygous mutation substituting Arg with Cys (UNG2-R88C) [36]. The UNG2-R88C mutation has recently been reported in the NCBI SNP database (rs151095402) with a frequency of the C/T heterozygote of 0.003 in a cohort of >1500 individuals in the NHLBI Exome Sequencing Project. Furthermore, the UNG2-R88C cell line's overall uracil excision activity has been measured and is comparable to that in other UNG-WT human tissues

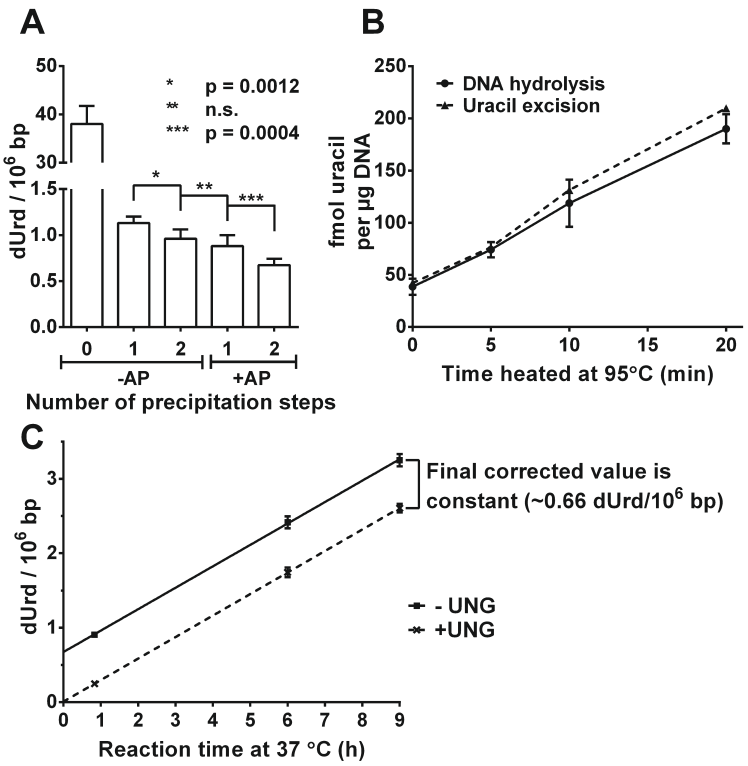


Fig. 2. Sample contamination with intracellular 2′-deoxyribonucleotides and *in vitro* dCyd deamination leads to overestimation of genomic dUrd. (A) Alkaline phosphatase (AP) pretreatment of commercially prepared DNA followed by repeated isopropanol precipitation steps prior to DNA hydrolysis decreases the final genomic dUrd value. (B) Denaturation of salmon sperm DNA by heating at 95 °C in water induces dCyd deamination and increases genomic dUrd or U content in time-dependent manner. (C) Prolongation of sample work-up procedure increases the amount of measured dUrd. Salmon sperm DNA samples as well as deuracilated controls were hydrolyzed at pH 6–7.6 and 37 °C for 50 min, 6 h, and 9 h. *In vitro* dCyd deamination occurs at the constant rate of 4.805×10^{-3} dUrd/10^6 bp/min. Results represent triplicate experiments ±SD.

and cell lines, whereas the UNG2-F251S is devoid of *in vitro* uracil excision activity [39,40]. We assayed genomic dUrd in these human cell lines in three separate experiments and found an 11-fold higher level of dUrd per base pair in the UNG2-F251S line (1.10 ± 0.13 dUrd/10^6 bp, CV 11.6%), as compared with the UNG2-R88C line (0.105 ± 0.014 dUrd/10^6 bp, CV 13%) (Fig. 3A).

We also quantified genomic dUrd in Ung-proficient and Ung-deficient mouse embryonic fibroblasts (MEFs) in triplicate experiments (Fig. 3B). We found a 5-fold higher genomic dUrd level in the $Ung^{-/-}$ line (0.344 ± 0.023 dUrd/10^6 bp, CV 6.76%) as compared with the $Ung^{+/+}$ line (0.072 ± 0.006 dUrd/10^6 bp, CV 8.59%). These experiments also suggest that other uracil-DNA glycosylases (*e.g.* SMUG1, TDG, and MBD4 [1]) cannot compensate for the lack of uracil-DNA glycosylase activity in the absence of UNG2.

4. Discussion

Although great progress has been made in understanding the mechanisms of base excision repair, quantitative information on the genomic content of the DNA base lesions and intermediates involved has yielded highly divergent results. As examples, measurements of genomic 8-oxo-7,8-dihydroguanine, uracil, and abasic sites have given results varying by orders of magnitude for each lesion [33,41–45].

Here, we have made efforts to improve quantification of genomic uracil by mass spectrometry and find that the content is lower than previously reported. Accurate quantification of genomic uracil is important to understand its processing, whether present as a lesion or as an essential intermediate in antibody affinity maturation. The interplay between these two fields forms the link between adaptive immunity and oncogenesis that has

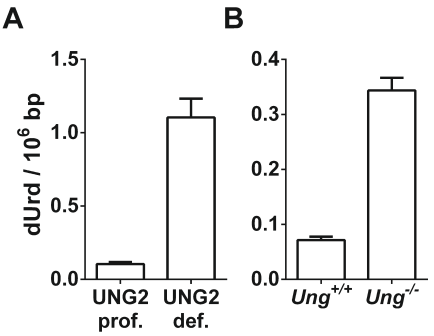


Fig. 3. Quantification of genomic dUrd in $Ung^{+/+}$ and $Ung^{-/-}$ mouse embryonic fibroblasts and human lymphoblastoid cell lines carrying *UNG* mutations. (A) UNG2 dysfunctional lymphoblastoid cells (UNG2-F251S) had 11-fold higher genomic dUrd level than lymphoblastoid cells with functional UNG2 (UNG2-R88C). (B) Genomic dUrd level was 5-fold higher in $Ung^{-/-}$ than in $Ung^{+/+}$ mouse embryonic fibroblasts. Results represent triplicate experiments ±SD.

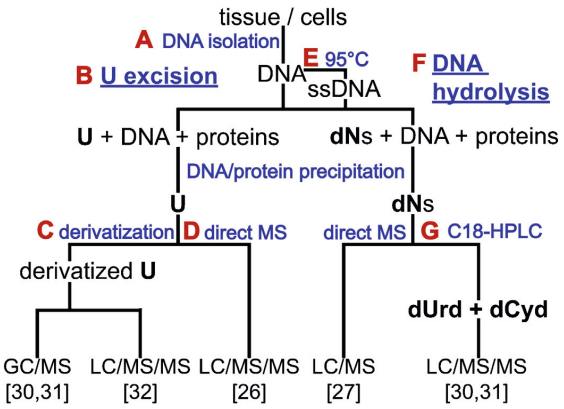


Fig. 4. Overview of possible errors in the methods for absolute quantification of genomic U/dUrd. (A) Intracellular 2′-deoxyribonucleotides can co-elute with DNA and be subsequently included in quantification. (B) Unspecific contaminants are usually more abundant with decreasing molecular weight of the precursor ions. (C) Differential derivatization of standards *versus* samples may lead to inaccuracies, and the efficiency of derivatization is not controlled. (D) Inaccurate determination of DNA concentration may compromise quantification. The extent of the uracil excision reaction is not monitored. (E) Denaturation of DNA by heating to 95 °C deaminates dCyd and overestimates the final genomic dUrd measurement. (F) Deamination of dCyd occurs at 37 °C and neutral pH. Extended incubation time during sample work-up may artificially increase the amount of dUrd. (G) dCyd elutes before dUrd with reverse-phase chromatography and may therefore contaminate the dUrd fraction due to peak tailing. dCyd may then be deaminated prior to MS/MS analysis.

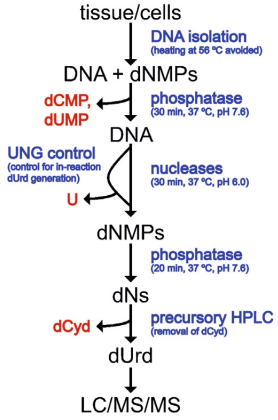


Fig. 5. Summary of improved genomic dUrd quantification. DNA isolation is improved by avoiding sample heating at 56 °C. A phosphatase pre-treatment step removes intracellular dCMP and dUMP, which otherwise co-purify with DNA. UNG2 is used to deuracilate DNA as a control processed in parallel to estimate whether and how much dUrd is generated during the analysis. DNA hydrolysis to dNs by nuclease/phosphatase treatment is kept short and pH neutral. A precursory HPLC step efficiently removes dCyd from the sample. Together, it significantly improves the accuracy of the method.

recently been established [13,15]. In this sense, relative quantification of genomic uracil can be useful and in some cases preferable to absolute quantification. For instance, several assays are DNA sequence-specific and can therefore shed light on specific AID off-target effects [12,22]. Nevertheless, relative assays hamper comparison between data sets, and sequence-specific assays are biased to the sequences they target. Indeed, the wide range of reported values for genomic uracil suggests that reliable quantification of genomic uracil (as free uracil or dUrd) is technically problematic [31,33]. Therefore, all steps from cell lysis through DNA isolation and analysis should be standardized and validated. A schematic visualization of the different approaches to genomic uracil and dUrd analyses and steps at which errors may arise is presented in Fig. 4. The DNA isolation step can be a significant error source (Fig. 4A). We noticed that isopropanol precipitation steps reduced the amount of measured dUrd regardless of how DNA was isolated. Adding alkaline phosphatase prior to precipitation further decreased the dUrd signal, presumably by removing intracellular nucleotides (specifically dUMP and dCMP) co-purifying with DNA.

As an alternative to DNA hydrolysis and quantification of dUrd, uracil can be excised using uracil-DNA glycosylase and directly analyzed by MS/MS (Fig. 4B) [30,31]. Uracil is inherently more prone to background signal in MS/MS because it is a heterocyclic molecule that resonates between non-aromatic amide and aromatic imide tautomers, the chemical bonds in which require more energy to break than the *N*-glycosylic bond between U and the deoxyribose in dUrd. Consequently, the additional collision energy required to break up the uracil molecule results in a higher probability of mistaking contaminants for the analyte. In this sense, quantification of dUrd is advantageous to measuring U because the abundance of interfering components is lower. Derivatizating U abrogates this effect, but adds complexity because the degree to which U has been derivatized cannot easily be monitored. In addition, it has proven difficult to establish robust conditions for derivatization to the extent that different conditions have been required to derivatize biological samples and standards (Fig. 4C) [32]. Derivatization can be circumvented by employing hydrophilic interaction chromatography (Fig. 4D) [26]. We tested a similar method and found the sensitivity comparable to measuring dUrd by hydrolysis; however, intra-sample variability was greater. Indeed, we compared DNA samples isolated using different methods and found no variability between DNA isolation methods when employing DNA hydrolysis, whereas there was significant difference using the U excision assay. This may result from our inability to gauge the extent of the U excision under these assay conditions, as well as imprecise estimation of the DNA concentration. In contrast, the present DNA hydrolysis method normalizes samples to the amount of dNs measured spectrophotometrically during the precursory HPLC step, which both determines the extent to which DNA has undergone hydrolysis and provides very accurate determination of DNA concentration. We performed hydrolysis with 5, 10, 15, and 20 μg DNA and saw no variation in dUrd measurements (data not shown). Moreover, samples are minimally handled between precursory HPLC and analytical LC/MS/MS, resulting in better reproducibility. Uracil excision is not necessarily inferior to DNA hydrolysis as a DNA-uracil quantification method and its shortcomings may theoretically be ameliorated by meticulous standardization of sample treatment, but it is nevertheless more susceptible to intra-lab or intra-sample variations.

Employing DNA hydrolysis to measure genomic dUrd has been reported previously [27–29]; however, the methods reported are prone to overestimation of genomic dUrd content for various reasons. DNA heat denaturation causes dCyd deamination and therefore overestimates dUrd estimates several-fold (Fig. 4E) [27]. The dCyd-containing products deaminate orders of magnitude faster than double-stranded DNA during long incubation times (6–9 h at 37 °C) required for complete enzymatic hydrolysis (Fig. 4F). We optimized experimental conditions to only require 50 min incubation at 37 °C. Decreasing this incubation time further would potentially yield more accurate results. Finally, employment of a normal reverse-phase column for precursory HPLC fractionation of dNs with which dUrd elutes after dCyd results in a risk of dCyd contamination in the dUrd fraction from peak tailing because dCyd is so much more abundant than dUrd (Fig. 4G) [28]. The dCyd

contamination is problematic both because [^{13}C]-dCyd is isobaric with dUrd and because dCyd may deaminate to dUrd between steps. To avoid that problem, we employed a reverse-phase column with weak acidic ion-pairing groups with which dUrd elutes before dCyd. The precursory HPLC step may be omitted by combining the PrimeSep200 and C18 columns with a column switcher. A dual-column system would shorten the total analysis time and decrease the likelihood of dUrd contamination as a result of sample handling before LC/MS/MS analysis; however, the accuracy of the assay would not necessarily increase. An overview of our improved method is presented in Fig. 5.

We used the optimized conditions to measure dUrd in DNA isolated from *Ung*$^{+/+}$ and *Ung*$^{-/-}$ mouse embryonic fibroblasts and human lymphoblastoid cell lines derived from hyper-IgM patients carrying *UNG* mutations. The values reported for genomic uracil here were lower than those reported by other groups, approximately 400–600 dUrd per human or murine genome in repair-proficient cells [31]. Although this alone does not prove that our method is superior to those previously published, our demonstration of overestimation sources indicates that our method is probably more reliable.

Conflict of interest statement

The authors declare no conflict of interest.

Acknowledgements

This work was supported by The Norwegian Cancer Association, The Svanhild and Arne Must Fund for Medical Research, the Research Council of Norway, and Norwegian University of Science and Technology. We gratefully acknowledge Dr. Anne Durandy (Institut National de la Santé et de la Recherche Médicale, Paris, France) for providing the lymphoblastoid cell lines used in this manuscript. We would like to thank the Proteomics and Metabolomics Core Facility, PROMEC, at NTNU supported in part by the Faculty of Medicine and the Central Norway Regional Health Authority.

References

[1] H.E. Krokan, F. Drablos, G. Slupphaug, Uracil in DNA – occurrence, consequences and repair, Oncogene 21 (2002) 8935–8948.
[2] R.S. Harris, S.K. Petersen-Mahrt, M.S. Neuberger, RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators, Mol. Cell 10 (2002) 1247–1253.
[3] R. Shapiro, Damage to DNA caused by hydrolysis, in: E. Seeberg, K. Kleppe (Eds.), Chromosome Damage and Repair, Plenum Press, New York, NY, 1981, pp. 3–18.
[4] D.E. Barnes, T. Lindahl, Repair and genetic consequences of endogenous DNA base damage in mammalian cells, Annu. Rev. Genet. 38 (2004) 445–476.
[5] C.W. Carreras, D.V. Santi, The catalytic mechanism and structure of thymidylate synthase, Annu. Rev. Biochem. 64 (1995) 721–762.
[6] S. Andersen, T. Heine, R. Sneve, I. Konig, H.E. Krokan, B. Epe, H. Nilsen, 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 (2005) 547–555.
[7] H.H. el-Hajj, H. Zhang, B. Weiss, Lethality of a dut (deoxyuridine triphosphatase) mutation in *Escherichia coli*, J. Bacteriol. 170 (1988) 1069–1075.
[8] M.M. Sousa, H.E. Krokan, G. Slupphaug, DNA-uracil and human pathology, Mol. Aspects Med. 28 (2007) 276–306.
[9] A. Verri, P. Mazzarello, G. Biamonti, S. Spadari, F. Focher, The specific binding of nuclear protein(s) to the cAMP responsive element (CRE) sequence (TGACGTCA) is reduced by the misincorporation of U and increased by the deamination of C, Nucleic Acids Res. 18 (1990) 5775–5780.
[10] K. Kemmerich, F.A. Dingler, C. Rada, M.S. Neuberger, Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung−/−Msh2−/− mice, Nucleic Acids Res. 40 (2012) 6016–6025.
[11] S. Schanz, D. Castor, F. Fischer, J. Jiricny, Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 5593–5598.
[12] R.W. Maul, H. Saribasak, S.A. Martomo, R.L. McClure, W. Yang, A. Vaisman, H.S. Gramlich, D.G. Schatz, R. Woodgate, D.M. Wilson, P.J. Gearhart 3rd., Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions, Nat. Immunol. 12 (2011) 70–76.
[13] D.F. Robbiani, A. Bothmer, E. Callen, B. Reina-San-Martin, Y. Dorsett, S. Difilippantonio, D.J. Bolland, H.T. Chen, A.E. Corcoran, A. Nussenzweig, M.C. Nussenzweig, AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations, Cell 135 (2008) 1028–1038.
[14] D.F. Robbiani, S. Bunting, N. Feldhahn, A. Bothmer, J. Camps, S. Deroubaix, K.M. McBride, I.A. Klein, G. Stone, T.R. Eisenreich, T. Ried, A. Nussenzweig, M.C. Nussenzweig, AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations, Mol. Cell 36 (2009) 631–641.
[15] D.F. Robbiani, M.C. Nussenzweig, Chromosome translocation, B cell lymphoma, and activation-induced cytidine deaminase, Ann. Rev. Pathol. 8 (2013) 79–103.
[16] O. Hakim, W. Resch, A. Yamane, I. Klein, K.R. Kieffer-Kwon, M. Jankovic, T. Oliveira, A. Bothmer, T.C. Voss, C. Ansarah-Sobrinho, E. Mathe, G. Liang, J. Cobell, H. Nakahashi, D.F. Robbiani, A. Nussenzweig, G.L. Hager, M.C. Nussenzweig, R. Casellas, DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes, Nature 484 (2012) 69–74.
[17] T. Honjo, M. Kobayashi, N. Begum, A. Kotani, S. Sabouri, H. Nagaoka, The AID dilemma: infection, or cancer? Adv. Cancer Res. 113 (2012) 1–44.
[18] H. Marusawa, T. Chiba, *Helicobacter pylori*-induced activation-induced cytidine deaminase expression and carcinogenesis, Curr. Opin. Immunol. 22 (2010) 442–447.
[19] S. Morita, Y. Matsumoto, S. Okuyama, K. Ono, Y. Kitamura, A. Tomori, T. Oyama, Y. Amano, Y. Kinoshita, T. Chiba, H. Marusawa, Bile acid-induced expression of activation-induced cytidine deaminase during the development of Barrett's oesophageal adenocarcinoma, Carcinogenesis 32 (2011) 1706–1712.
[20] S. Andersen, M. Ericsson, H.Y. Dai, J. Pena-Diaz, G. Slupphaug, H. Nilsen, H. Aarset, H.E. Krokan, Monoclonal B-cell hyperplasia and leukocyte imbalance precede development of B-cell malignancies in uracil-DNA glycosylase deficient mice, DNA Repair (Amst) 4 (2005) 1432–1441.
[21] H. Atamna, I. Cheung, B.N. Ames, A method for detecting abasic sites in living cells: age-dependent changes in base excision repair, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 686–691.
[22] A. Horvath, B.G. Vertessy, A one-step method for quantitative determination of uracil in DNA by real-time PCR, Nucleic Acids Res. 38 (2010) e196.
[23] S.U. Lari, C.Y. Chen, B.G. Vertessy, J. Morre, S.E. Bennett, Quantitative determination of uracil residues in *Escherichia coli* DNA: contribution of ung, dug, and dut genes to uracil avoidance, DNA Repair (Amst) 5 (2006) 1407–1420.
[24] R.S. Lasken, D.M. Schuster, A. Rashtchian, Archaebacterial DNA polymerases tightly bind uracil-containing DNA, J. Biol. Chem. 271 (1996) 17692–17696.
[25] N. Yan, E. O'Day, L.A. Wheeler, A. Engelman, J. Lieberman, HIV DNA is heavily uracilated, which protects it from autointegration, Proc. Natl. Acad. Sci. U.S.A. 108 (2011) 9244–9249.
[26] A.D. Bulgar, L.D. Weeks, Y. Miao, S. Yang, Y. Xu, C. Guo, S. Markowitz, N. Oleinick, S.L. Gerson, L. Liu, Removal of uracil by uracil DNA glycosylase limits pemetrexed cytotoxicity: overriding the limit with methoxyamine to inhibit base excision repair, Cell Death Dis. 3 (2012) e252.
[27] A. Chango, A.M. Abdel Nour, C. Niquet, F.J. Tessier, Simultaneous determination of genomic DNA methylation and uracil misincorporation, Med. Princ. Pract. 18 (2009) 81–84.
[28] M. Dong, P.C. Dedon, Relatively small increases in the steady-state levels of nucleobase deamination products in DNA from human TK6 cells exposed to toxic levels of nitric oxide, Chem. Res. Toxicol. 19 (2006) 50–57.
[29] M. Dong, C. Wang, W.M. Deen, P.C. Dedon, Absence of 2′-deoxyoxanosine and presence of abasic sites in DNA exposed to nitric oxide at controlled physiological concentrations, Chem. Res. Toxicol. 16 (2003) 1044–1055.
[30] S.T. Mashiyama, C. Courtemanche, I. Elson-Schwab, J. Crott, B.L. Lee, C.N. Ong, M. Fenech, B.N. Ames, Uracil in DNA, determined by an improved assay, is increased when deoxynucleosides are added to folate-deficient cultured human lymphocytes, Anal. Biochem. 330 (2004) 58–69.
[31] S.T. Mashiyama, C.M. Hansen, E. Roitman, S. Sarmiento, J.E. Leklem, T.D. Shultz, B.N. Ames, An assay for uracil in human DNA at baseline: effect of marginal vitamin B6 deficiency, Anal. Biochem. 372 (2008) 21–31.
[32] J. Ren, A. Ulvik, H. Refsum, P.M. Ueland, Uracil in human DNA from subjects with normal and impaired folate status as determined by high-performance liquid chromatography-tandem mass spectrometry, Anal. Chem. 74 (2002) 295–299.
[33] R. Olinski, M. Jurgowiak, T. Zaremba, Uracil in DNA – its biological significance, Mutat. Res. 705 (2010) 239–245.
[34] G. Slupphaug, I. Eftedal, B. Kavli, S. Bharati, N.M. Helle, T. Haug, D.W. Levine, H.E. Krokan, Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase, Biochemistry 34 (1995) 128–138.
[35] H. Nilsen, I. Rosewell, P. Robins, C.F. Skjelbred, S. Andersen, G. Slupphaug, G. Daly, H.E. Krokan, T. Lindahl, D.E. Barnes, Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication, Mol. Cell 5 (2000) 1059–1065.
[36] K. Imai, G. Slupphaug, W.I. Lee, P. Revy, S. Nonoyama, N. Catalan, L. Yel, M. Forveille, B. Kavli, H.E. Krokan, H.D. Ochs, A. Fischer, A. Durandy, Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination, Nat. Immunol. 4 (2003) 1023–1028.
[37] T. Lindahl, Instability and decay of the primary structure of DNA, Nature 362 (1993) 709–715.

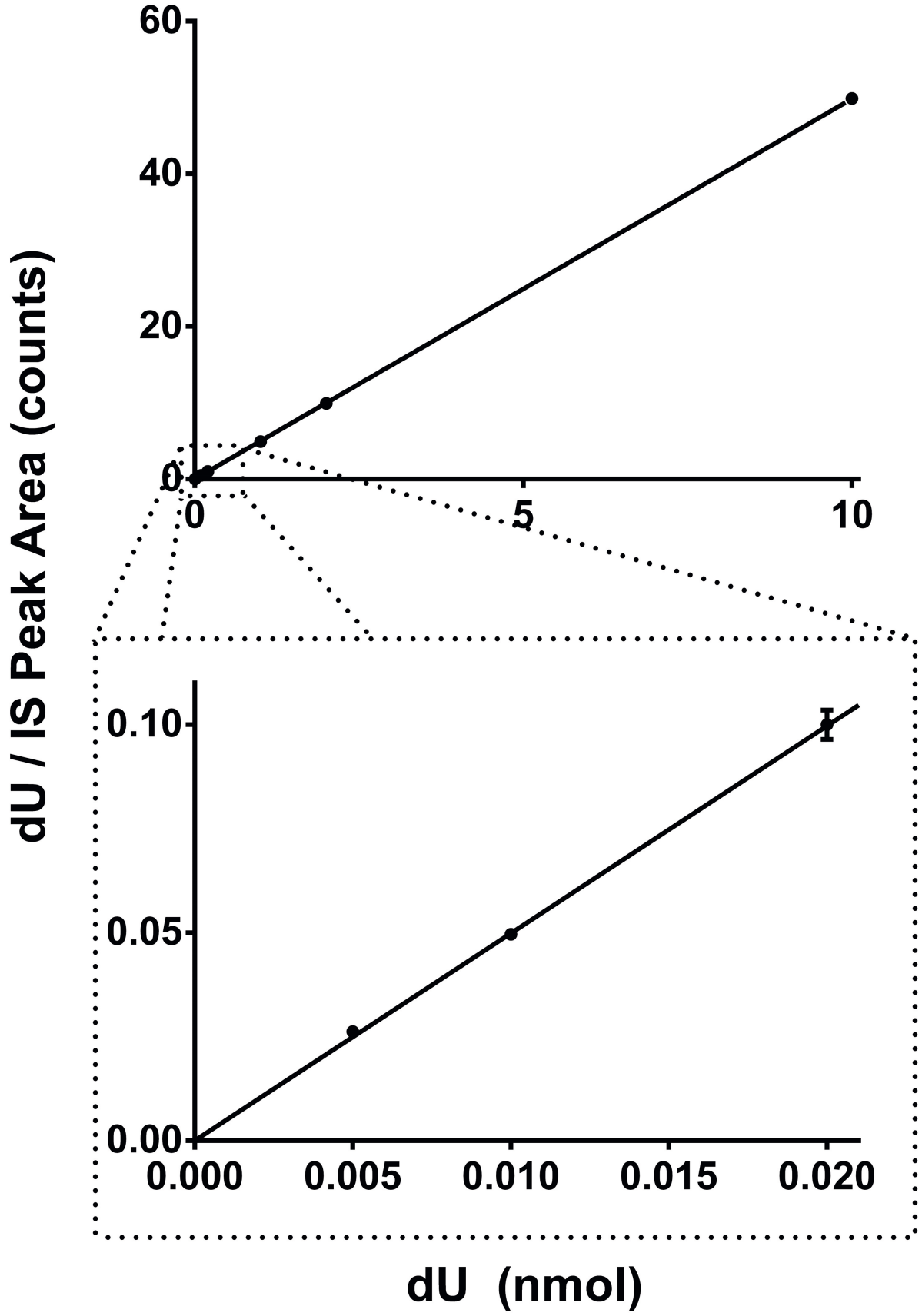
[38] T. Lindahl, B. Nyberg, Heat-induced deamination of cytosine residues in deoxyribonucleic acid, Biochemistry 13 (1974) 3405– 3410.
[39] B. Kavli, S. Andersen, M. Otterlei, N.B. Liabakk, K. Imai, A. Fischer, A. Durandy, H.E. Krokan, G. Slupphaug, B cells from hyper-IgM patients carrying UNG mutations lack ability to remove uracil from ssDNA and have elevated genomic uracil, J. Exp. Med. 201 (2005) 2011–2021.
[40] K. Torseth, B. Doseth, L. Hagen, C. Olaisen, N.B. Liabakk, H. Graesmann, A. Durandy, M. Otterlei, H.E. Krokan, B. Kavli, G. Slupphaug, The UNG2 Arg88Cys variant abrogates RPA-mediated recruitment of UNG2 to single-stranded DNA, DNA Repair (Amst) 11 (2012) 559–569.
[41] J. Cadet, T. Douki, J.L. Ravanat, Measurement of oxidatively generated base damage in cellular DNA, Mutat. Res. 711 (2011) 3–12.
[42] M. Endres, M. Ahmadi, I. Kruman, D. Biniszkiewicz, A. Meisel, K. Gertz, Folate deficiency increases postischemic brain injury, Stroke 36 (2005) 321–325.
[43] J.M. Li, M. Mogi, K. Tsukuda, H. Tomochika, J. Iwanami, L.J. Min, C. Nahmias, M. Iwai, M. Horiuchi, Angiotensin II-induced neural differentiation via angiotensin II type 2 (AT2) receptor-MMS2 cascade involving interaction between AT2 receptor-interacting protein and Src homology 2 domain-containing protein-tyrosine phosphatase 1, Mol. Endocrinol. 21 (2007) 499–511.
[44] A.M. Luke, P.D. Chastain, B.F. Pachkowski, V. Afonin, S. Takeda, D.G. Kaufman, J.A. Swenberg, J. Nakamura, Accumulation of true single strand breaks and AP sites in base excision repair deficient cells, Mutat. Res. 694 (2010) 65–71.
[45] D.R. McNeill, D.M. Wilson 3rd., A dominant-negative form of the major human abasic endonuclease enhances cellular sensitivity to laboratory and clinical DNA-damaging agents, Mol. Cancer Res. 5 (2007) 61–70.

Supplementary Figure Legends

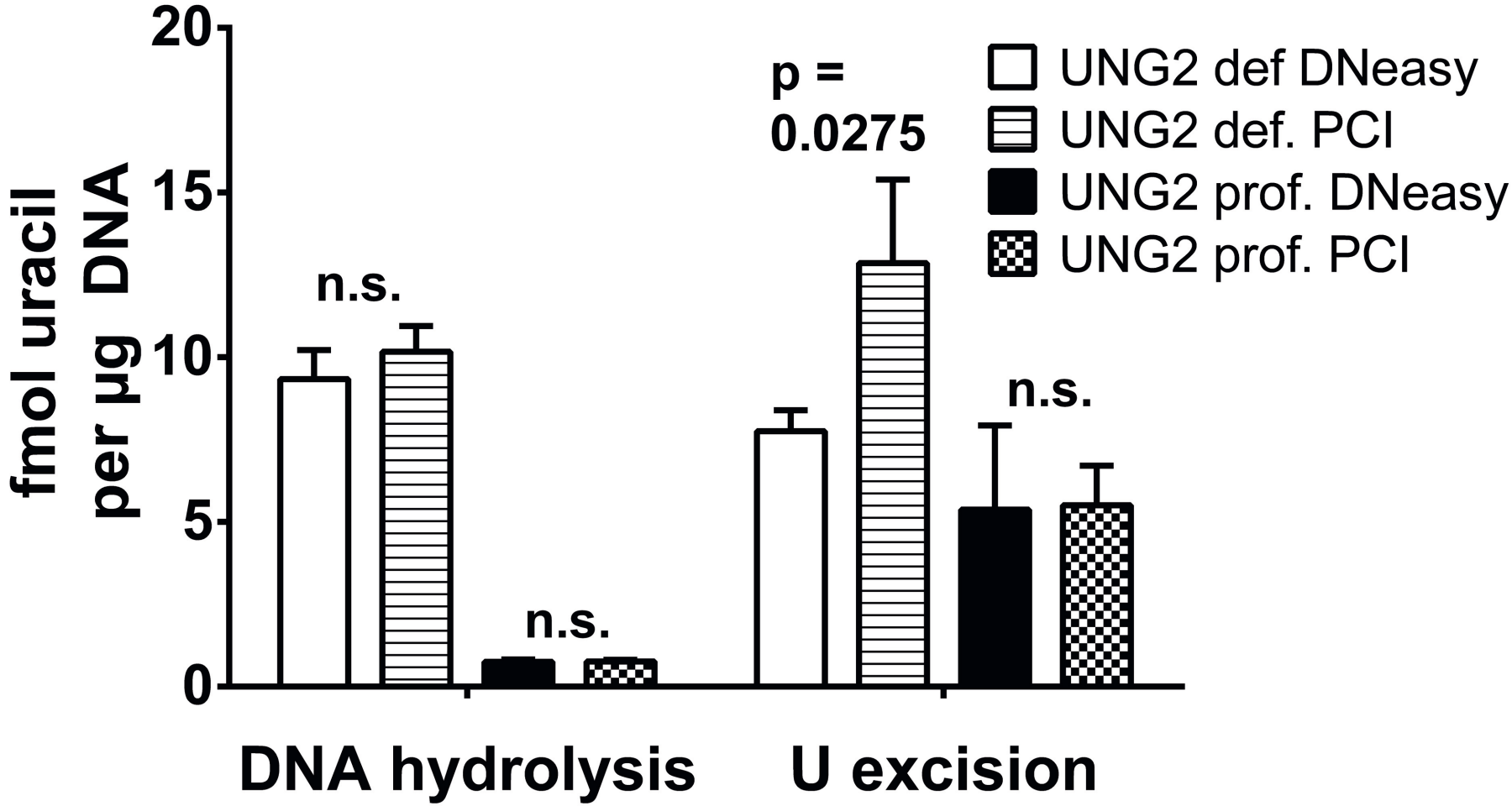
Supplementary Fig. 1. Standard curve used for genomic uracil quantification. A concentration gradient from 5 fmol to 10 nmol dU was used with 200 fmol $13^{13}C^{15}N_2$-dU as an internal standard. The R^2 value was 1.000. Note that both plots represent the same data displayed with different y-axis scales. Results represent triplicate experiments ± SD.

Supplementary Fig. 2. DNA hydrolysis is a more robust assay for genomic uracil than the uracil excision method. Both methods were employed to measure genomic uracil levels in DNA isolated from UNG2 proficient and deficient cell lines using both a DNA extraction kit from Qiagen (DNeasy) and phenol:chloroform:isoamyl alcohol DNA isolation method (PCI). Results represent triplicate experiments ± SD.

Supplementary figure 1



Supplementary figure 2



Article III

UNG AND SMUG1 EFFICIENTLY COMPLEMENT EACH OTHER IN REMOVING GENOMIC URACIL FROM MOUSE ORGANS

Antonio Sarno[a#], Lene Alsøe[b#], Anastasia Galashevskaya[a], Laure Jobert[b], Tanima SenGupta[b], Sergio Carracedo[b], Nuriye Basdag Tekin[b], Hans E. Krokan[a], Hilde Nilsen[b]

[a]Department of Cancer Research and Molecular Medicine, Department of Medicine, Norwegian University of Science and Technology, Trondheim, Norway
[b]Institute of Clinical Medicine, Department of Clinical Molecular Biology, University of Oslo, and Akershus University Hospital, Lørenskog, Norway

[*] Correspondence: Hilde Nilsen
Tel: +47-679xxxx
E-mail: **hilde.nilsen@medisin.uio.no**

[#]These authors contributed equally to this work

Key words: Uracil, 5-hydroxymethyluracil, SMUG1, UNG, Base Excision Repair, knockout mice, LC-MS/MS
Running head: SMUG1 and UNG2 prevent accumulation of genomic uracil
Character count:

ABSTRACT

Uracil in DNA is both a lesion and a necessary intermediate for antibody maturation. Uracil is primarily excised by the uracil-DNA glycosylase (UNG2) or single-strand selective monofunctional uracil DNA glycosylase (SMUG1), which initiate the error-free base excision repair pathway. The relative contributions of UNG2 and SMUG1 in limiting uracil accumulation have been difficult to resolve using biochemical approaches and remain an area of contention. Here, we used a genetic approach to clarify this issue: we generated gene targeted *Smug1*$^{-/-}$ mice and crossed them with the previously generated *Ung*$^{-/-}$ mice to generate *Smug1*$^{-/-}$*Ung*$^{-/-}$ mice. We measured uracil excision activity and genomic uracil and 5-hydroxymethyluracil (hmU) levels in mouse embryonic fibroblasts and organs from both single and double knock-out mice, as well as hmU excision activity in wild type and *Smug1*$^{-/-}$ MEFs and wild type organs. Expectedly, hmU excision activity was undetectable in *Smug1*$^{-/-}$ mice, and UNG was found to be the major contributor to uracil excision in most tissues. SMUG1 was the largest contributor to U:G, and more weakly to U:A excision in the brain, but its activity in other tissues was only apparent when UNG was inhibited. A slight 2-3-fold increase in genomic uracil was observed in UNG-deficient tissues, whereas genomic uracil levels from *Smug1*$^{-/-}$ mice were indistinguishable from isogenic wild type controls. In contrast, the *Smug1*$^{-/-}$*Ung*$^{-/-}$ organs contained 3-20-fold more uracil than their wild-type counterparts. Our data suggests that low levels of genomic uracil are maintained by either UNG or SMUG1, despite their different activity levels. Thus, we postulate that physiological variations in glycosylase activity are not a major factor in increasing genomic uracil levels.

INTRODUCTION

Uracil is a natural intermediate in thymine biosynthesis. A small dUTP pool is therefore necessary, the consequence of which is occasional dUTP misincorporation in place of dTTP because DNA polymerases cannot distinguish between the two nucleotides (Friedberg et al., 1995). Genomic uracil is also generated by spontaneous and enzymatic cytosine deamination, generating U:G mispairs. dUMP misincorporation (generating U:A) is likely the larger contributor to the genomic uracil burden in proliferating cells, with an estimated $\sim 10^4$ dUTP misincorporated per genome per division, whereas the spontaneous cytosine deamination rate is estimated to be 100 to 500 per cytosines per cell per day (Lindahl, 1993; Mosbaugh and Bennett, 1994). There is little evidence to indicate that U:A pairs arising from dUTP misincorporation are deleterious, though they have been shown to alter transcription factor binding and topoisomerase activity *in vitro* (Pourquier et al., 1997; Risse et al., 1989). In contrast, unrepaired cytosine-deamination events replicate to generate C:G to T:A mutations. Indeed, enzymatic deamination of cytosine by activation induced deaminase (AID) initiates somatic hypermutation (SHM) and class switch recombination (CSR) during immunoglobulin (Ig) diversification (Rada et al., 2002). Recent work has also shown that DNA mutational signatures associated with AID and other enzymes in the APOBEC-family are present in a wide variety of cancers, suggesting a role for enzymatic DNA cytosine deamination in tumorigenesis (Alexandrov et al., 2013; Bolli et al., 2014; Burns et al., 2013a, 2013b; Chen and Wang, 2014; Davis et al., 2014; Hoogstraat et al., 2014; Lada et al., 2012; Nik-Zainal et al., 2012).

Thus, understanding how uracil is processed and repaired is increasingly important. Uracil is primarily excised by uracil-DNA glycosylases (UDGs) that initiate the error-free base excision repair (BER) pathway. Mammalian cells express several (UDGs), including mitochondrial uracil-DNA glycosylase 1 (UNG1) and nuclear UNG2, single-strand selective monofunctional uracil DNA glycosylase 1 (SMUG1), G/T mismatch-specific thymine DNA glycosylase (TDG), and methyl-CpG-binding domain protein 4 (MBD4). All UDGs have the ability to excise U:G, though only UNG and SMUG1 efficiently remove U:A. The relative contribution of the different UDGs to uracil repair are influenced by a plethora of factors like sequence context, whether the DNA is single- or double-stranded, the base opposite the uracil, cell-cycle phase, post-translational modifications, interaction partners, and their localization in the genome (reviewed in (Krokan et al., 2014)). The substrate specificities of the UDGs and their preference for various DNA contexts partially overlap. SMUG1 also has

a particular role in removing the thymine oxidation product 5-hydroxymethyluracil (hmU) (Boorstein et al., 2001).

Ung$^{-/-}$ mice revealed that UNG2 has a role in U:A and U:G repair, as well as in Ig diversification (Nilsen et al., 2000a, 2003, 2005; Rada et al., 2002). Initial work suggested that SMUG1 serves as a backup for UNG2 in both uracil repair and Ig diversification (Dingler et al., 2014; Kemmerich et al., 2012; Nilsen et al., 2001; Rada et al., 2002); however, distinct intracellular localisation and interaction partners suggest that SMUG1 and UNG2 likely also have non-redundant functions. UNG2 interacts with RPA and PCNA and removes misincorporated uracil by postreplicative DNA repair (Otterlei et al., 1999). Divergently, SMUG1 does not localise with replicating chromatin, but is instead localised throughout the nucleus with some accumulation in nucleoli (Kavli et al., 2002). Furthermore, SMUG1 directly interacts with Dyskerin (DKC1) in Cajal bodies, and contributes to RNA quality control (Jobert et al., 2013). However, the distinct functions of UNG2 and SMUG1 in genomic-uracil repair *in vivo* remain poorly understood. Here, we used a genetic approach to clarify the relative importance of UNG2 and SMUG1 in genomic uracil repair and generated gene targeted *Smug1*$^{-/-}$ mice and crossed them with *Ung*$^{-/-}$ mice to generate *Smug1*$^{-/-}$*Ung*$^{-/-}$ mice.

MATERIALS AND METHODS

Generation of conditionally targeted Smug1 knockout mice

The *Smug1* gene (NM_027885) spans 14 kb and eight predicted exons of which only the two last exons are protein coding. Two isoforms exist that differ in their 5′-UTRs. A two-step strategy giving a final deletion of the protein coding exons 1 and 2 was used. A gene targeting vector containing *Smug1* exon 1 and exon 2 was constructed by cloning genomic *Smug1* fragments PCR amplified from C57Bl/6 genomic DNA using Accuprime Taq DNA polymerase High Fidelity (Thermo Scientific) into the pCR4-TOPO vector (Thermo Scientific). The long-homology arm of the targeting vector was built by TA-cloning of two separate PCR fragments: i) a 3371 bp fragment comprising 5´-UTR was amplified using forward primer 5′-TAGATGTGGTGGGGATAGACTAGAACCTGG-3′ and reverse primer 5′-CTACAAGCTCACTTTCCTGGTAACGAAGG-3′ and ii) a 2464 bp fragment comprising the exon 1 and exon 2 regions amplified using the primers (5′-TGACTGACAGGGTTTCTTCTGAGCCC-3′ and 5′-GAAGGGGAAGACAGCAGGAGAGCTG-3′) flanked by two LoxP sites. A positive selection neomycine gene flanked by FRT-sites was inserted. The short homology arm was generated by amplification of the 3′ untranslated region of exon 2 (5′-CCTTGAGCCTCTCACCCTTTTGTCTC-3′ and 5′-CTCCTATTGTTCCCAACAGTTGCC-3′). A Diphteria Toxin A (*DTA*) gene was used for negative selection. The absence of PCR-generated mutations was confirmed by sequencing. The targeting vector was linearized with *Pme*I and electroporated into C57Bl/6 ES cells. ES cells were selected with 200 μg/ml G418. Homologous recombination events at the 5′and 3′ arms were verified in G418 resistant ES cell clones by PCR and Southern blot analysis. Six independent clones were injected into C57BL/6J blastocysts. Chimeric mice were crossed with a C56BL/6 Cre-deleter mouse strain (GenOway) to allow germline excision of the *loxP*-flanked region thus generating heterozygous constitutive *Smug1* knock-out mice.

The generation of $Ung^{-/-}$ mice in a mixed 129SV-C57Bl/6J background was described previously and backcrossed ten generations into the C57Bl/6J background (Doseth et al., 2011; Nilsen et al., 2000b). $Smug1^{-/-}Ung^{-/-}$ double-knockout (DKO) mice were generated by crossing single-knockout mice born to heterozygous mothers. All strains were maintained as heterozygous.

Southern blotting

Genomic DNA was digested with *PciI,* blotted onto a nylon membrane and hybridized with an external 432 bp probe (ext 3′ probe) located downstream of the short homology arm of the targeting vector. The expected fragment sizes of the wild type and the recombined Cre-excised loci are 9.3 kb and 2.1 kb, respectively. The non-excised allele would give a fragment of 6.3 kb. Pre-hybridization and hybridization were performed at 65 °C for 18 h in hybridization solution (4 × SSC, 1% SDS, 0.5% skimmed milk, 20 mM EDTA, 100 µg/ml herring sperm DNA) followed by washing twice in 3 × SSC, 1% SDS at 65 °C for 15 min and twice in 0.5 × SSC, 1% SDS at 65 °C for 15 min. Bands were visualised after 3 days exposure to BioMax MS films with BioMax intensifying screens. Primers used to generate the external 3′ probe: 5′-CTCATCTGTCTCTTTAATGGTTGGTTGGATG-3′ and 5′-AGCTGGCTAGGGTCACTGTGGAGGTAT-3′.

Genotyping

Mice were genotyped by multiplex PCR. The primer set for the *Smug1* alleles were as follows: 5′-GGATGAGGGTTCAGCCAGACCTACA-3′ (forward WT), 5′-ACTGCGAATATGACTTCAGACATCCCG-3′ (reverse WT), SMUG KO 5′-TGACAGGGTCACATGTCGTACATAA-3′ (forward KO) and SMUG KO 5′-ACTGCGAATATGACTTCAGACATCC3′ (reverse KO). The primer set for the *Ung* alleles were as follows: mUNG3+: 5′-CACGGACCTAATCAAGCTCACG-3′ (forward WT), mUNG4-: 5′-GGCCCACCCTGACAAATCCCC3′ (reverse WT/KO) and Neo+: 5′-CTTGGGTGGAGAGGCTATTC-3′ (forward KO). AccuPrime *Pfx* Supermix (Invitrogen) was used for the PCR of both *Smug1* and *Ung*. The PCR program for *Smug1* was: 1 cycle of 95 °C for 5 min, 35 cycles of 95 °C for 15 s, 58 °C for 30 s and 68 °C for 1 min and then 1 cycle of 68 °C for 7 min. The PCR program for *Ung* was equal to that of *Smug1* except for the annealing temperature, which was 59 °C. The PCR products were run on 2% agarose gels. The *Smug1* PCR were expected to give a wild-type band of 271 bp and a knock-out band of 232 bp (Figure 1C). The *Ung* PCR gave a wild-type band of 550 bp and a knock-out band of 850 bp (Supplementary Figure 1).

Isolation and culture of MEFs

Timed matings were set up between either wild type, $Smug1^{-/-}$ or $Ung^{-/-}$ mice born to heterozygous parents or between $Smug1^{-/-}Ung^{-/-}$ mice born from either $Smug1^{+/-}Ung^{-/-}$ or $Smug1^{-/-}Ung^{+/-}$ parents in order to obtain wild type, $Smug1^{-/-}$ or $Smug1^{-/-}Ung^{-/-}$ MEFs. MEFs were established from 13.5 to 14.5 dpc embryos as described (Xu, 2005), and grown in

Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12, GlutaMAX (Invitrogen) supplemented with 10% fetal bovine serum (Lonza), 1x penicillin-streptomycin (Invitrogen) and 1x MEM non-essential amino acids (Invitrogen). Primary MEF cell lines (Passage 1) and cells cultured for 23 continuous passages were used.

Phenotypic assessment of *Smug1*$^{-/-}$ mice

Phenotypic assessment was performed according to the modified SHIRPA protocol (EMPRESS, eumorphia.org) at PhenoPro (http://www.phenopro.fr/). Briefly, 10 male wild type and 10 *Smug1*$^{-/-}$ mice aged 6 weeks old were housed 1 to 4 per cage and fed standard chow diet (D04, Safe) *ad libitum*. Phenotypic testing started at 11 weeks after 3 weeks acclimation in the phenotypic area. Starting from 13 weeks, mice were submitted to dysmorphology screen to assess morphological abnormalities in their general physical appearance (weight, length, and dysmorphology, including tail kinks, shape of ears, eyes, head, teeth, limbs, number and shape of digit, irregularities and variation in coat colour, hair distribution and development, irregularities in the genitals). Blood was collected from 14-week-old mice by retro-orbital puncture under isoflurane anesthesia for biochemistry and hematologic analysis. A complete blood cell count was performed in the Advia 120 workstation. Body composition and bone mineral content were evaluated by X-ray analysis.

Generation of anti-SMUG1 antibodies

Generation of an antibody against mouse SMUG1 was carried out by BioGenes. The antibody was raised in rabbit against the mouse SMUG1 peptide CLTPAELPAKQREQL-amide. The first immunisation was followed by five weekly boosts before the final bleeding of the rabbit. The antiserum was purified against the immunising peptide coupled to SulfoLink gel (Pierce). The coupling of the peptide to the SulfoLink gel and the purification were carried out according to the recommendations from the manufacturer. Eluted fractions containing the purified antibody were pooled and dialysed against 10 mM Tris HCl (pH 7.9), 20 % glycerol, 0.5 mM EDTA, 50 mM KCl.

Protein extraction for oligonucleotide nicking assays

Protein was extracted from organs and cells using two buffers. Lysis buffer I contained 10 mM Tris-HCl (pH 8.0) and 200 mM KCl and lysis buffer II contained 10 mM Tris-HCl (pH 8.0), 200 mM KCl, 40 % glycerol, and 0.5% NP-40 alternative (Merck Millipore). Buffers were freshly supplemented with 1 µM DTT, 1X Complete Protease Inhibitor Cocktail (Roche), and 1X Phosphatase Inhibitor Cocktails 2 and 3 (Sigma-Aldrich). All steps were performed with ice-cold buffer and on ice or at 4 °C. Protein from MEFs was extracted by first suspending the cells in 3 µl lysis buffer I per 10^6 cells and then adding the

same amount of lysis buffer II. The suspensions were then incubated for 2 h at 4 °C and centrifuged at 16,100 rcf. The supernatants were finally aliquoted to new tubes, snap-frozen in liquid nitrogen, and stored at -80 °C. Protein from organs were lysed by suspending organs in 1.8 µl 1:1 lysis buffers I and II per mg organ and homogenization with Dounce homogenizers. The homogenates were then incubated, centrifuged, aliquoted, and stored identically to the MEFs.

Isolation of genomic DNA

Organ samples were snap-frozen in liquid nitrogen and stored at -80 °C until isolation of genomic DNA. The genomic DNA was isolated using the DNeasy Blood & Tissue kit (Qiagen) according to the manufacturer's instructions with minor modifications. Briefly, 2 ml CK14 homogenisation tubes containing ceramic 1.4 mm zirconium oxide beads (Precellys) were prepared with 400 µl ice-cold lysis buffer consisting of: 360 µl ATL buffer, 40 µl Proteinase K and 0.1 µg/µl RNaseA. 10-25 mg tissue per tube were homogenised at 4 °C by bead beating for 30 s 3x with 30 s pauses between cycles. The lysates were incubated in a water bath for at least 1 h at 37 °C and vortexed occasionally. The lysates were recovered from the beads through two needle holes in the lid of the homogenisation tube as the tube was spin upside down inside a 15 ml tube at 250 rcf for 2 min. Each lysate was split into two spin columns. The final DNA elution step was performed with 200 µl milli-Q water.

Whole cell extract of MEFs for western analysis

Wild type and *Smug1*$^{-/-}$ primary MEFs at 80 % confluency were trypsinised with 0.25% trypsin-EDTA (Invitrogen) and washed twice in cold PBS. The cells were resuspended in 300 µl extraction buffer (20 mM Tris-HCl pH 7.5, 1 mM DTT, 20% glycerol, 400 mM KCl, 1x protease inhibitor cocktail, cOmplete Mini EDTA-free (Roche)) and incubated on ice for 30 min with occasional gentle mixing. The extracts were exposed to three freeze-thaw cycles in liquid nitrogen and spun at 16000 rcf for 20 min at 4 °C. The supernatants were collected and dialysed against 25 mM Tris-HCl pH 7.5, 5 mM MgCl2, 100 mM KCl, 10% glycerol, 1 mM DTT in a Slide-A-Lyzer Dialysis cassette 3500 MWCO (Fisher Scientific) overnight at 4 °C. The whole cell extracts were collected and added 1x protease inhibitor cocktail. Protein concentrations were determined by Bradford quantification (Bio-Rad) using BSA as standard. The extracts were snap-frozen in liquid nitrogen and stored at -80 °C until further use.

Western blot analysis

100 µg whole cell extract of wild type and *Smug1*$^{-/-}$ MEFs were run on a 15% Tris-HCl Criterion gel (Bio-Rad) and transferred to a nitrocellulose membrane (Bio-Rad). The

membrane was blocked in 1x PBS, 0.1% tween-20, 5% milk for 1 h at room temperature and incubated overnight at 4 °C in purified rabbit anti-mouse SMUG1 antibody diluted 1:500 in the blocking solution. The membrane was washed three times 10 min in 1x PBS, 0.1% tween-20 and then incubated for 1 h at room temperature in goat anti-rabbit IgG-HRP (Cell Signaling) diluted 1:2000 in the blocking solution. The membrane was washed as above before it was developed with SuperSignal West Pico Chemiluminescent kit according the instructions from the manufacturer (Thermo Scientific). Hyperfilm ECL (GE Healthcare) was exposed to the membrane for various time-points. The membrane was stripped by incubation in 0.2 M glycine, 0.1% SDS, 0.1% tween-20 (pH 2.2) two times 20 min at 37 °C. The membrane was blocked as above and incubated for 1 h at room temperature with rabbit anti-GAPDH antibody (14C10, Cell Signaling) diluted 1:2000 in blocking solution. The membrane was further processed as above.

Gene expression analyses

Kidneys from wild type, *Smug1*$^{+/-}$, and *Smug1*$^{-/-}$ mice had been stored in RNAlater at -20 °C until RNA isolation. RNA was isolated using the mirVana miRNA isolation kit (Ambion/Life Technologies). 2 ml CK14 homogenisation tubes containing 1.4 mm ceramic 1.4 mm zirconium oxide beads were prepared with 600 µl ice-cold lysis buffer and a piece of tissue of 2 x 2 x 2 mm was added per tube and covered by the lysis buffer. The tissue was homogenised and the lysate recovered as for isolation of genomic DNA. The lysate was processed further according to manufacturer's instructions. The RNA was eluted in 100 µl DEPC treated milli-Q water. cDNA was synthesised from 5 µg RNA using SuperScript II Reverse Transcriptase (Invitrogen/Life Technology) and random hexamers according to the instructions following the transcriptase. The *Smug1* transcription was measured by quantitative real-time PCR using Fast SYBR Green Master Mix (Applied Biosystems) according to the instructions following the kit. *Smug1* expression was normalised to *Gapdh* expression. 2 µl cDNA was used in the qRT-PCR for both *Smug1* and *Gapdh*. Primers for the qRT-PCR were *Smug1* (forward) 5′-TCAAGTCTTCTTCCGGCACT-3′, *Smug1* (reverse) 5′-ACTCCCACTACCAGACGCAC-3′, *Gapdh* (forward) 5′-AACTTTGGCATTGTGGAAGG-3′ and *Gapdh* (reverse) 5′-GGATGCAGGGATGATGTTCT-3′.

Oligonucleotide nicking assays

Oligonucleotide UDG assays were performed as previously described (Doseth et al., 2011; Kavli et al., 2002). 6-carboxyfluorescein-labeled uracil- or hydroxymethyluracil-containing oligonucleotides (5′-[6-FAM]-CATAAAGTG-U/5hmU-AAAGCCTG) were annealed to 1.5x of the complementary strand containing G or A opposite U/5hmU. Activity

was measured by incubating 10 µl of 20 nM substrate and 20, 2.5, or 1.25 µg of (5hm)U:A, (5hm)U:G, or ss(5hm)U protein extracts, respectively. The reactions were buffered in 20 mM Tris-HCl (pH 8.0), 60 mM NaCl, and 1 mM EDTA, freshly complemented to a final concentration of 1 mM DTT and 0.5 mg/ml BSA. The reactions were incubated at 37 °C for 60 min, 20 min, and 15 min for (5hm)U:A, (5hm)U:G, or ss(5hm)U protein extracts, respectively. UNG was inhibited by 0.4 U uracil-DNA glycosylase inhibitor (Ugi, New England Biolabs). The reactions were stopped on ice and abasic sites were cleaved by adding 50 µl 10 % ice-cold piperidine and subsequent incubation at 90 °C for 20 min. Next, the reactions were vacuum centrifuged at 60 °C for 1 h to dryness and redissolved in 30 µl 60 % formamide loading buffer containing 0.05 % bromophenol blue. The substrate and product were separated by electrophoresis on a urea-PAGE gel containing 12 % acrylamide:bis-acrylamide (19:1) and 42% urea in 0.5 x TBE and visualized using a Typhoon Trio imager (GE Healthcare). Analysis was performed using ImageQuant 7 TL (GE Healthcare).

Quantification of modified bases in genomic DNA

Modified nucleosides were quantified as previously described (Galashevskaya et al., 2013). Potentially co-purified deoxynucleotides in DNA were removed by incubating 15 µg DNA with 0.2 U alkaline phosphatase from *E. coli* per 100 µl in 100 mM NH_4HCO_3 (pH 7.6) and 10 mM $MgCl_2$ for 30 min at 37 °C, followed by isopropanol precipitation. The DNA was then hydrolyzed to nucleosides by redissolving the DNA pellet in 20 µl containing 10 mM $NH_4C_2H_3O_2$ (pH 6.0), 10 mM MgCl2, 10 mM CaCl2, 41.65 nM $^{13}C^{15}N_2$-dUrd, 0.4 U nuclease P1 from *P. citrinum* (Sigma-Aldrich), and 2 U DNase I (Roche). The solution was incubated for 30 min at 37 °C, after which 5 µl containing 0.5 NH_4HCO_3 (pH 7.6) and 0.1 U alkaline phosphatase were added and further incubated at 37 °C for 20 min. The reactions were stopped on ice and the enzymes in the reactions with 3 volumes of cold acetonitrile, after which the samples were centrifuged at 16,100 rcf for 30 min at 4 °C. The supernatants were transferred to new tubes and vacuum centrifuged at room temperature until dry.
To separate (5hm)-dUrd from dCyd, the samples were redissolved in 100 µl 90:10 water:acetonitrile and fractionated with 3x 30 µl injections on an Agilent HP1100 HPLC-UV system and a Primesep 200 column (2.1 mm x 150 mm, 5 µm, Sielc) using a flow rate of 0.4 ml/min and water and acetonitrile as mobile phase, each containing 0.1 % formic acid. The HPLC 7.5-min-long gradient was as follows: 10 % acetonitrile for 30 s, ramp to 60 % acetonitrile by 2 min to 2.5 min, and return to 10 % acetonitrile by 2.55 min. The (5hm)-dUrd-containing fractions were collected from 1.3 - 1.8 min and vacuum centrifuged until dry. The pellets were redissolved in 25 µl 95:5 water:methanol and analyzed for dUrd and

5hm-dUrd by LC/MS/MS using a reverse phase column (2.1 mm x 150 mm, 3.5 µm, Zorbax SB-C18, Agilent Technologies) on an LC-20AD HPLC (Shimadzu) coupled to an AB SCIEX 5500 triple quadrupole mass spectrometer with an electrospray ion source (AB SCIEX). The injection volume was 20 µl, the flow rate was 0.3 ml/min, and the 5-min-long HPLC gradient was as follows: 5 % methanol for 30 s, ramp to 30 % methanol by 1 min to 2 min, and return to 5 % methanol by 2.1 min. Analysis was performed in positive ionization multiple reaction monitoring mode, monitoring the mass transitions 228.994 → 113.0, 232.0 → 116.0, and 259.0 → 143.1 for dUrd, $^{13}C^{15}N_2$-dUrd, and 5hm-dUrd, respectively. Chromatogram analysis was performed using Analyst 1.5 software (AB SCIEX).

Statistical analysis

Data were analyzed using unpaired one-way- or repeated measures analysis of variance (ANOVA) with one between factor (genotype) and one within factor (time). Qualitative parameters (e.g. some of clinical observations) were analyzed using $\chi 2$ test. The level of significance was set at $p < 0.05$. Biochemical assays were evaluated using two-way t-test 95% confidence level.

RESULTS

Smug1$^{-/-}$ mice display no obvious pathological phenotypes

Gene targeted mice deleted for both coding exons of the *Smug1* gene (Figure 1A) were born at Mendelian ratios and are fertile. *Smug1*$^{-/-}$mice had no residual *Smug1* expression, as measured in total RNA isolated from MEFs by qRT-PCR (Figure 1D). Heterozygous mice expressed *Smug1* mRNA at 50% of the wild type level. An affinity purified peptide antibody generated against a mouse SMUG1 peptide revealed a specific band migrating at 29 kDa in whole cell extract prepared from wild type kidney that was absent in extracts prepared from *Smug1*$^{-/-}$ kidney (Figure 1E). Hence, we concluded that the *Smug1*$^{-/-}$ mice are true knockouts.

No obvious pathological abnormality was observed in the general appearance of the *Smug1*$^{-/-}$ mice, and they exhibited normal body shape and skeletal morphology. No major changes were observed in blood-cell counts although there was a slight reduction in the mean number of lymphocytes in the *Smug1*$^{-/-}$ mice (5.70 ± 0.15 vs 4.57 ± 0.25, $p = 0.0011$). No macroscopic or microscopic lesion or change was observed by histological assessment of the major organs. The generation and phenotypic assessment of *Ung*$^{-/-}$ mice in a mixed 129SV-C57Bl/6J background were described previously to have no obvious morphological phenotype (Nilsen et al., 2000b), they present a hyper IgM phenotype resulting from defective processing of AID-induced uracil during CSR in the *Ig* loci (Rada et al., 2002). A high incidence of B-cell lymphoma was also observed in aging, mixed background *Ung*$^{-/-}$ mice (Andersen et al., 2005; Nilsen et al., 2003). *Ung*$^{-/-}$ mice (Supplementary Figure S1) were backcrossed ten generations into C57Bl6/J and used to generate *Smug*$^{-/-}$*Ung*$^{-/-}$ mice. *Smug1*$^{-/-}$ *Ung*$^{-/-}$ mice also showed no obvious morphological abnormalities. Thus, the deletion of the *Smug1* gene and the combined loss of the two major UDGs are well tolerated in mice in the time period examined here.

Smug1$^{-/-}$ mice lose hmU excision activity and accumulate hmU in genomic DNA

SMUG1 was previously characterised as the main enzyme removing 5-hydroxymethyluracil (hmU) from DNA (Boorstein et al., 2001). To determine whether the *Smug1*$^{-/-}$ mice had any residual SMUG1 activity, we measured the excision capacity on an oligonucleotide substrate harbouring a centrally placed hmU residue. The hmU-excision activity from MEF and organ extracts was higher on hmU:G than hmU:A substrates, whereas no detectable activity was found on single-stranded substrates under our reaction conditions (Figure 2A/B). In contrast, there was no hmU-excision activity in extracts from isogenic *Smug1*$^{-/-}$ MEFs, regardless of substrate (Figure 2A). There was some variation in hmU-

excision activity between different organs, in particular on hmU:G substrate, with the highest activity in the brain and lowest in the liver extracts. (Figure 2B).

To determine whether the lack of hmU excision activity in *Smug1*$^{-/-}$ cells resulted in increased genomic hmU load, we measured the amount of hmU present in total genomic DNA by modifying our recently developed LS/MS/MS assay for measuring genomic uracil (Galashevskaya et al., 2013). Primary *Smug1*$^{-/-}$ MEFs (Passage 1) contained about 20 hmU residues per million nucleotides, which was double the genomic hmU content of isogenic primary wild types MEFs (Figure 2C). The genomic hmU-levels seemed to slightly increase upon passage in culture, as would be expected because hmU can be formed by direct oxidation of thymine; however, the increase in hmU content was similar in wild type and *Smug1*$^{-/-}$ MEFs. Similar levels of hmU were found in genomic DNA isolated from different tissues (Figure 2D). The tissue variation in hmU levels was unremarkable, ranging from about 5 to 10 hmU per million nucleotides in wild type mice (Figure 2D, top panel). Genomic DNA isolated from *Smug1*$^{+/-}$ mice had about double the amount of hmU compared to wild types, suggesting that a single expressed copy of SMUG1 is insufficient to maintain basal hmU levels (Figure 2D, middle panel). Considerably higher hmU levels were found in genomic DNA isolated from organs of *Smug1*$^{-/-}$ mice and most tissues exhibited 3- to 4-fold increased hmU levels compared to wild type organs. The largest increases relative to wild type were found in brain extracts where the levels increased from ~5 to 18 hmU per million nucleotides. Hence, we concluded that *Smug1*$^{-/-}$ mice have lost hmU excision activity and accumulate hmU in genomic DNA.

Smug1$^{-/-}$ mice have normal uracil-excision capacity

SMUG1 was previously found to the predominant UDG in *Ung*$^{-/-}$ mice (Doseth et al., 2011; Kemmerich et al., 2012; Nilsen et al., 2001). To assess whether loss of SMUG1 impacted uracil repair capacity, we performed standard uracil excision activity assays on oligonucleotide substrates containing one centrally placed uracil residue. The uracil excision activity varied more than the activity on hmU-containing substrates (Figure 3). High activity to the point of assay saturation was observed in spleen and heart extracts on U:G substrates (Figure 3, top panel). Wild-type mice showed less efficient uracil excision in splenic extracts on U:A containing substrates compared to U:G (Figure 3, middle panel). Uracil excision activity from single-stranded substrates resembled the U:G substrate, except that brain activity was reduced by 60% (Figure 3, bottom panel). SMUG1 status did not significantly impact uracil-excision activity from double-stranded substrates in tissues other than the brain, in which the activity was reduced by 40% and 60% in *Smug1*$^{+/-}$ and *Smug1*$^{-/-}$ mice,

respectively. Consistently, no statistically significant accumulation of genomic uracil was seen in *Smug1*$^{-/-}$ MEFs (Figure 4A) or any of the SMUG1-deficient organs tested (Figure 4B). Uracil levels also did not rise in MEFs upon culture *in vitro* (Figure 4A). These data are in agreement with previous findings that UNG is the dominating UDG in mouse cells, although SMUG1 contributes significantly to uracil-excision from double-stranded substrates in the brain.

The combined action of SMUG1 and UNG prevents genomic uracil accumulation

To test whether SMUG1 activity contributes to uracil repair in the absence of UNG, we generated *Smug1*$^{-/-}$*Ung*$^{-/-}$ mice. There was a complete ablation of measurable uracil excision activity in all *Smug1*$^{-/-}$*Ung*$^{-/-}$ organs regardless of substrate (Figure 5A). This demonstrated that both UNG and SMUG1 have the ability to compensate for loss of the other enzyme, but that the loss of SMUG1 in an UNG-proficient background impacts relatively little on overall uracil excision capacity (Figure 3). In contrast, SMUG1 appears not to be fully able to compensate for the loss of UNG activity, as shown by partial suppression of uracil excision activity upon inhibition of UNG by Ugi (Figure 5A). Again, these data confirm that UNG and SMUG1 account for the majority of uracil excision activity in mice, with UNG as the larger contributor.

As shown above, the loss of SMUG1 did not lead to a significant increase in the global genome uracil content (Figure 4); however, the loss of UNG activity resulted in a 2-3-fold increase in genomic uracil content in the tissues measured here (Figure 5B, middle panel). In contrast to the relatively modest increase in the UNG-deficient tissues, there was a dramatic increase in genomic uracil levels in the double knockout organs (Figure 5B, lower panel). Surprisingly, the spleen, an organ with high proliferative capacity, showed a modest 3-fold increase from ~2 to 7 uracil per million nucleotides in wild type compared to DKO extracts, whereas a 6-fold uracil accumulation was seen in the essentially post-mitotic brain extracts. Heart, skeletal muscle, kidney, and lungs all showed between ~20 and 30 uracil per million nucleotides, which represent more than 10-fold increases relative to the wild type levels. Finally, liver extracts from DKO mice showed more than 80 uracils per million nucleotides, which represents a 20-fold increase relative to wild type levels. This synergistic increase in uracil levels demonstrates that SMUG1 efficiently, albeit not fully, compensates for loss of UNG activity and *vice versa*.

In summary, SMUG1 is the sole detectable hmU excision enzyme in MEFs and hmU accumulates in the genome of *Smug1*$^{-/-}$ mice. UNG is the major UDG in mouse cells and organs. The contribution of SMUG1 to uracil removal is highly tissue specific and appears to

be either more active or abundant in brain extracts, despite the lack of uracil accumulation upon the SMUG1-deficient organs. There is extensive buffering between the two enzymes with respect to global genome uracil repair and their combined action effectively prevents accumulation of uracil in the mouse genome. However, the two enzymes are not entirely redundant as demonstrated by the synergistic increase in uracil levels in the double knockout mice.

DISCUSSION

In the present work we describe the generation of a gene targeted knockout mouse model deficient in SMUG1 and show that SMUG1 comprises the only detectable hmU-excision activity in mouse tissues. Consequently, the absence of SMUG1 leads to genomic hmU accumulation. Reduced UDG activity was detected in double-stranded substrates in *Smug*$^{-/-}$ brain extracts, but other organs were unaffected by the loss of SMUG1; however, by comparing uracil accumulation in MEF and tissues extracted from *Smug*$^{-/-}$, *Ung*$^{-/-}$, and *Smug*$^{-/-}$ *Ung*$^{-/-}$ mice, we demonstrate that SMUG1 and UNG effectively collaborate to limit uracil accumulation in genomic DNA.

A mouse model deficient in SMUG1 was previously generated by germline transmission of ES cells generated in the European Conditional Mouse Mutagenesis Consortium (EUCOMM, project ID 23057) in which suppression of SMUG1 expression was achieved by a gene trap inserted in the intron between the coding exons (Kemmerich et al., 2012). This mouse model has no detectable SMUG1 expression and has lost hmU-activity. Thus, there is no reason to suspect that this model does not represent a good *Smug1*-knockout model. We nevertheless chose to make a classical gene targeted deletion of both coding exons of the *Smug1* gene because bioinformatic analysis performed at Genoway (www.genoway.com) revealed one possible splice variant from the *Smug1* gene that harboured a cryptic ATG upstream of the gene-trap cassette coding exons (AK020817) giving a theoretical possibility of a splice event between the engrailed 2 gene encoded by the inserted gene trap and *Smug1* exon 2 giving rise to a stable mRNA. If translated, this mRNA would give rise to a 137 kDa protein with a partial UDG domain (39 out of 194 amino acids). This protein would not be predicted to have UDG activity as it would only contain 1 out of 4 substrate binding sites, but possible hypomorphic or gain-of-function effects would, nevertheless, be a theoretical possibility.

The *Smug*$^{-/-}$ mice developed here phenocopy the previously-described model and showed that loss of SMUG1 does not interfere with normal development and extensive phenotypic screening did not reveal any obvious pathophysiological changes at 10 weeks of age (Kemmerich et al., 2012). The small reduction in lymphocyte numbers found in our model might be related to the role of SMUG1, albeit minor, in CSR (Dingler et al., 2014). Stimulated B-cells undergo selection in germinal centres during the immune response, so slightly less efficient antibody generation as a result of SMUG1-deficiency may slightly increase apoptosis induction or decrease proliferation in B-cells, resulting in the observed lowered lymphocyte levels.

In agreement with the previous study, we found that *Smug1*$^{-/-}$ cells and tissues had no detectable hmU-excision activity leading to 3-4-fold higher levels of hmU in genomic DNA (Kemmerich et al., 2012). The extent of hmU accumulation was similar in all tissues analysed. A small increase in hmU, but not uracil, levels was observed upon passage of wild type and *Smug1*$^{-/-}$ MEFs in culture. This is consistent with the majority of cellular hmU being generated from direct oxidation of thymine (Pfaffeneder et al., 2014). Surprisingly, the increase was similar in *Smug1*$^{-/-}$ and wild type MEFs, which might be indicative of compensatory changes limiting the accumulation of hmU in SMUG1-deficient cells.

Importantly, the work presented here offers novel insight into the long-standing question on the relative importance of UNG2 and SMUG1 in uracil repair. The hmU excision profile (*i.e.* the differences between organs) in wild type samples was nearly identical to that of uracil excision in UNG-inhibited WT samples, suggesting that SMUG1 excises the two nucleosides with similar efficiency. Loss of SMUG1 had no effect on UDG activity in the presence of functional UNG in most organs. This supports the conclusion that UNG is the major contributor to uracil excision in most tissues in the mouse (Doseth et al., 2011). A significant reduction of UDG activity in *Smug1*$^{-/-}$ tissues was found only in brain extracts on U:G substrate and more weakly on U:A substrate. No effect was seen on uracil excision from single-stranded substrates, corroborating earlier studies indicating that SMUG1 primarily removes uracil from double-stranded substrates under our reaction conditions (Doseth et al., 2012). We cannot exclude the possibility that the contribution of SMUG1 to total UDG activity was underestimated using the present assay conditions which was optimised for UNG (Supplementary Figure 2) (Akbari and Krokan, 2012; Doseth et al., 2012). Indeed, the dramatic reduction in UDG activity in all *Smug1*$^{-/-}$*Ung*$^{-/-}$ organs tested strongly indicates that SMUG1 contributes to UDG activity and that confirms that it is the major UDG in UNG-deficient mice (Nilsen et al., 2001). *Smug1*$^{-/-}$ tissues or cells did not accumulate uracil, and *Ung*$^{-/-}$ tissues only showed a 2- to 3-fold increase in uracil levels. That the UDG activity contributed by SMUG1 is relevant *in vivo* is evident from the dramatic increase in uracil levels in the double knockout organs. Hence, both UNG2 and SMUG1 contribute to maintenance of baseline uracil levels.

As both UNG and SMUG1 efficiently remove uracil from U:A and U:G pairs, we do not yet know whether the main increase in uracil in the DKOs comes from misincorporation or deamination. The rates of spontaneous cytosine deamination are expected to be largely dependent on the degree of single-stranded DNA, and therefore depend on the transcription activity of a cell and unlikely to differ in the genetic background studied here (Lindahl,

1993). There are also other enzymes and pathways that can repair U:G mismatches in the absence of UNG and SMUG1, most notably the mismatch repair pathway and the two mismatch specific uracil-DNA glycosylases TDG and MBD4 (Cortázar et al., 2007; Kunz et al., 2009; Wong et al., 2002). For U:A the two most efficient repair enzymes are UNG2 and SMUG1 (Kavli et al., 2002). Uracil misincorporation will occur in replicating cells in direct proportion to the cellular dUTP pool, which is largely determined by the dUTPase enzyme. dUTPase expression is cell cycle regulated. A relatively low increase in uracil content in the brain, in which there is very low cellular turnover, might indicate that the bulk of genomic uracil in the *Smug1*$^{-/-}$*Ung*$^{-/-}$ organs originates from uracil misincorporation. A similarly modest increase in uracil content in the spleen was more surprising in light of the important physiological function of AID induced deamination in this tissue. AID expression is low in mouse spleen and it can be expected that the effect of AID-dependent cytosine deamination will not be seen in the absence of stimulation (Nilsen et al., 2005). Furthermore, the importance of the mismatch repair pathway in processing AID-initiated U:G lesions is supported by the severe early morbidity caused by thymic lymphomas in *Smug1*$^{-/-}$*Ung*$^{-/-}$*Msh2*$^{-/-}$ mice (Kemmerich et al., 2012; Kunz et al., 2009). Hence, whereas many alternatives exist to repair deaminated cytosine, few enzymes other than UNG and SMUG1 are known to effectively repair misincorporated uracil, which might suggest that the dramatic increase in uracil content in the *Smug*$^{-/-}$*Ung*$^{-/-}$ mice is likely dominated by U:A pairs. The absence of a tumour-prone phenotype in the *Smug*$^{-/-}$*Ung*$^{-/-}$ mice previously reported would also support this interpretation, but definite conclusions regarding the relative accumulation of U:A or U:G pairs must await assay improvements (Kemmerich et al., 2012).

In conclusion, there is extensive buffering between UNG and SMUG1 with respect to global genome uracil repair *in vivo* and their combined action effectively prevents accumulation of uracil in the mouse genome. However, the two enzymes are not entirely redundant as demonstrated by the synergistic increase in uracil levels in the double knockout mice. We therefore propose that in the absence of active uracilation by *e.g.* enzymatic deamination, physiological variations in glycosylase levels are sufficient to maintain uracil levels at near basal levels *in vivo*.

LEGENDS TO FIGURES

Figure 1 - Generation of Smug1-knockout mice

(A) Gene targeting strategy showing the endogenous *Smug1* gene (top panel) with the two coding exons (striped boxes), the 3'-UTR (blue box), and two predicted upstream exons (white boxes). The recombined *Smug1* locus (middle panel) with the floxed neomycine selection marker (grey box) and *LoxP* sites indicated. The position of the 3' external probe is indicated (red box). The *Smug1* locus after Cre-mediated excision (lower panel). The expected fragment sizes resulting from PciI digestion are indicated below each panel. (B) Southern blot confirming the size of the endogenous locus and in a heterozygous animal after Cre-mediated excision. (C) Smug1 PCR genotyping showing both wild type (wt) and knockout (ko) bands of expected sizes. (D) qRT-PCR showing *Smug1* mRNA expression (as arbitrary units AU) relative to *Gapdh* as mean +/- SD from triplicate measurements (E) Western blots of whole cell extracts prepared from mouse embryonic fibroblasts probed with anti-SMUG1 antibodies. GAPDH expression was used as loading control.

Figure 2 - *Smug1*$^{-/-}$ mice lose hmU excision activity and accumulate genomic hmU

(A) *Smug1*$^{-/-}$ MEFs exhibit no 5-hmU excision activity. (B) hmU excision activity is higher on hmU:G substrate than hmU:A substrate in all organs and undetectable on single-stranded substrate. (C) *Smug1*$^{-/-}$ MEFs accumulate 5-hmU in genomic DNA. (D) hmU incorporation in all organs is dependent on SMUG1 status. Representative gels in A and B are shown below each graph and error bars indicate SD of three biological replicates.

Figure 3 - SMUG1 status has little impact on uracil excision activity in all organs except the brain

Mice organs exhibit a wide range of uracil excision activities. The activities in brain samples were reduced by 40 % and 60 % on U:A and U:G substrates, respectively. SMUG1-status did not influence single-stranded uracil excision activity. Representative gels are shown below each graph and error bars indicate SD of three biological replicates.

Figure 4 - *Smug1*$^{-/-}$ mice do not accumulate genomic uracil

(A) Genomic uracil did not increase in either *Smug1*$^{-/-}$ mice or after 23 passages. (B) Genomic uracil did not significantly increase in any organ in either *Smug1*$^{+/-}$ or *Smug1*$^{-/-}$ mice. Error bars indicate SD of three biological replicates.

Figure 5 - *Ung*$^{-/-}$*Smug1*$^{-/-}$ mice lose all uracil excision activity and accumulate a large amount of genomic uracil.

(A) U:G and U:A excision activities is decreased upon UNG inhibition by Ugi in all organ extracts except the brain and completely ablated in all organs in *Smug1*$^{-/-}$*Ung*$^{-/-}$ mice. Single-stranded uracil excision activities are ablated both in Ugi-treated extracts and in *Smug1*$^{-/-}$*Ung*$^{-/-}$ mice. (B) Genomic uracil increased by 2-3-fold and 3-20-fold in *Ung*$^{-/-}$ and *Ung*$^{-/-}$*Smug1*$^{-/-}$ mouse organs, respectively. Representative gels in A are shown below each graph and error bars indicate SD of three biological replicates.

Supplementary Figure 1

Agarose gel electrophoresis image of *Ung* PCR genotyping showing both wild type (WT) and knockout (KO) bands of expected sizes.

Supplementary Figure 2 - Uracil excision assay optimization using liver extracts.

Uracil excision activity was measured over time in wild-type, Ugi-treated, and *Smug1*$^{-/-}$ livers on (A) U:A substrate using 20 µg extract, (B) U:G substrate using 5 µg extract, and (C) single-stranded uracil substrate using 2.5 µg extract.

REFERENCES

Akbari, M., and Krokan, H.E. (2012). Base excision repair efficiency and mechanism in nuclear extracts are influenced by the ratio between volume of nuclear extraction buffer and nuclei—Implications for comparative studies. Mutat. Res. Mol. Mech. Mutagen. *736*, 33–38.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

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.

Bolli, N., Avet-Loiseau, H., Wedge, D.C., Van Loo, P., Alexandrov, L.B., Martincorena, I., Dawson, K.J., Iorio, F., Nik-Zainal, S., Bignell, G.R., et al. (2014). Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. Nat. Commun. *5*, 2997.

Boorstein, R.J., Cummings, A., Marenstein, D.R., Chan, M.K., Ma, Y., Neubert, T.A., Brown, S.M., and Teebor, G.W. (2001). Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1. J. Biol. Chem. *276*, 41991–41997.

Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B., et al. (2013a). APOBEC3B is an enzymatic source of mutation in breast cancer. Nature *494*, 366–370.

Burns, M.B., Temiz, N.A., and Harris, R.S. (2013b). Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat. Genet. *45*, 977–983.

Chen, Z., and Wang, J.H. (2014). Generation and repair of AID-initiated DNA lesions in B lymphocytes. Front. Med. *8*, 201–216.

Cortázar, D., Kunz, C., Saito, Y., Steinacher, R., and Schär, P. (2007). The enigmatic thymine DNA glycosylase. DNA Repair *6*, 489–504.

Davis, C.F., Ricketts, C.J., Wang, M., Yang, L., Cherniack, A.D., Shen, H., Buhay, C., Kang, H., Kim, S.C., Fahey, C.C., et al. (2014). The somatic genomic landscape of chromophobe renal cell carcinoma. Cancer Cell *26*, 319–330.

Dingler, F.A., Kemmerich, K., Neuberger, M.S., and Rada, C. (2014). Uracil excision by endogenous SMUG1 glycosylase promotes efficient Ig class switching and impacts on A:T substitutions during somatic mutation. Eur. J. Immunol. *44*, 1925–1935.

Doseth, B., Visnes, T., Wallenius, A., Ericsson, I., Sarno, A., Pettersen, H.S., Flatberg, A., Catterall, T., Slupphaug, G., Krokan, H.E., et al. (2011). Uracil-DNA Glycosylase in Base Excision Repair and Adaptive Immunity. J. Biol. Chem. *286*, 16669–16680.

Doseth, B., Ekre, C., Slupphaug, G., Krokan, H.E., and Kavli, B. (2012). Strikingly different properties of uracil-DNA glycosylases UNG2 and SMUG1 may explain divergent roles in processing of genomic uracil. DNA Repair *11*, 587–593.

Friedberg, E.C., Walker, G., and Siede, W. (1995). DNA repair and mutagenesis (ASM Press).

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.

Hoogstraat, M., de Pagter, M.S., Cirkel, G.A., van Roosmalen, M.J., Harkins, T.T., Duran, K., Kreeftmeijer, J., Renkens, I., Witteveen, P.O., Lee, C.C., et al. (2014). Genomic and transcriptomic plasticity in treatment-na?ve ovarian cancer. Genome Res. *24*, 200–211.

Jobert, L., Skjeldam, H.K., Dalhus, B., Galashevskaya, A., Vågbø, C.B., Bjørås, M., and Nilsen, H. (2013). The human base excision repair enzyme SMUG1 directly interacts with DKC1 and contributes to RNA quality control. Mol. Cell *49*, 339–345.

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.

Kemmerich, K., Dingler, F.A., Rada, C., and Neuberger, M.S. (2012). Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice. Nucleic Acids Res. *40*, 6016–6025.

Krokan, H.E., Sætrom, P., Aas, P.A., Pettersen, H.S., Kavli, B., and Slupphaug, G. (2014). Error-free versus mutagenic processing of genomic uracil—Relevance to cancer. DNA Repair *19*, 38–47.

Kunz, C., Saito, Y., and Schär, P. (2009). DNA Repair in mammalian cells: Mismatched repair: variations on a theme. Cell. Mol. Life Sci. CMLS *66*, 1021–1038.

Lada, A.G., Dhar, A., Boissy, R.J., Hirano, M., Rubel, A.A., Rogozin, I.B., and Pavlov, Y.I. (2012). AID/APOBEC cytosine deaminase induces genome-wide kataegis. Biol. Direct *7*, 47; discussion 47.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709–715.

Mosbaugh, D.W., and Bennett, S.E. (1994). Uracil-excision DNA repair. Prog Nucleic Acid Res Mol Biol *48*, 315–370.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al. (2012). Mutational processes molding the genomes of 21 breast cancers. Cell *149*, 979–993.

Nilsen, H., Rosewell, I., Robins, P., Skjelbred, C.F., Andersen, S., Slupphaug, G., Daly, G., Krokan, H.E., Lindahl, T., and Barnes, D.E. (2000a). Uracil-DNA Glycosylase (UNG)-Deficient Mice Reveal a Primary Role of the Enzyme during DNA Replication. Mol. Cell *5*, 1059–1065.

Nilsen, H., Steinsbekk, K.S., Otterlei, M., Slupphaug, G., Aas, P.A., and Krokan, H.E. (2000b). Analysis of uracil-DNA glycosylases from the murine Ung gene reveals differential expression in tissues and in embryonic development and a subcellular sorting pattern that differs from the human homologues. Nucleic Acids Res. *28*, 2277–2285.

Nilsen, H., Haushalter, K.A., Robins, P., Barnes, D.E., Verdine, G.L., and Lindahl, T. (2001). Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase. EMBO J. *20*, 4278–4286.

Nilsen, H., Stamp, G., Andersen, S., Hrivnak, G., Krokan, H.E., Lindahl, T., and Barnes, D.E. (2003). Gene-targeted mice lacking the Ung uracil-DNA glycosylase develop B-cell lymphomas. Oncogene *22*, 5381–5386.

Nilsen, H., An, Q., and Lindahl, T. (2005). Mutation frequencies and AID activation state in B-cell lymphomas from Ung-deficient mice. Oncogene *24*, 3063–3066.

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.

Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S.K., Eisen, D., Truss, M., Steinbacher, J., Hackner, B., Kotljarova, O., et al. (2014). Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. Nat. Chem. Biol. *10*, 574–581.

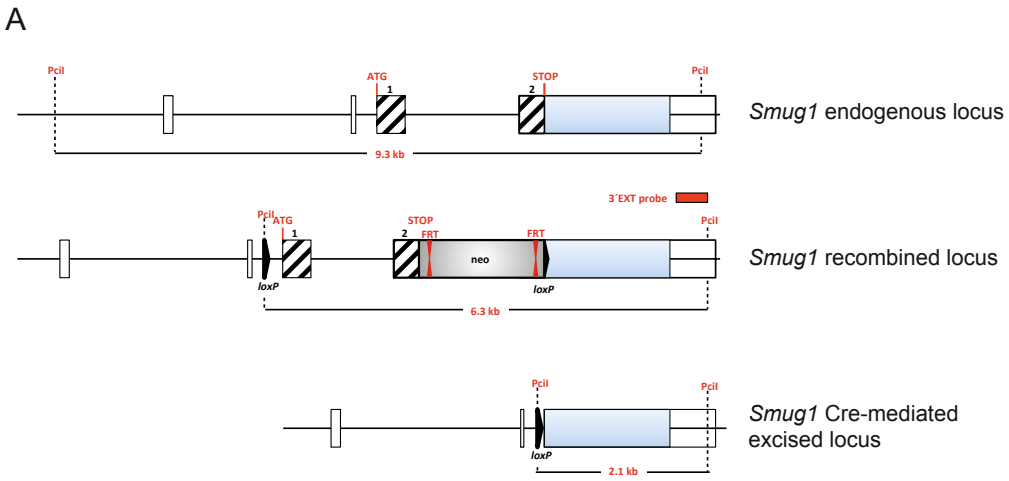
Pourquier, P., Ueng, L.M., Kohlhagen, G., Mazumder, A., Gupta, M., Kohn, K.W., and Pommier, Y. (1997). Effects of uracil incorporation, DNA mismatches, and abasic sites on cleavage and religation activities of mammalian topoisomerase I. J. Biol. Chem. *272*, 7792–7796.

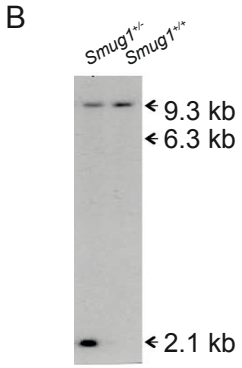
Rada, C., Williams, G.T., Nilsen, H., Barnes, D.E., Lindahl, T., and Neuberger, M.S. (2002). Immunoglobulin Isotype Switching Is Inhibited and Somatic Hypermutation Perturbed in UNG-Deficient Mice. Curr. Biol. *12*, 1748–1755.

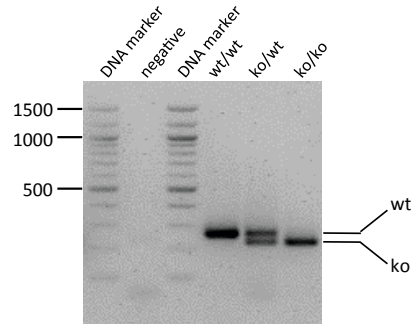
Risse, G., Jooss, K., Neuberg, M., Brüller, H.J., and Müller, R. (1989). Asymmetrical recognition of the palindromic AP1 binding site (TRE) by Fos protein complexes. EMBO J. *8*, 3825–3832.

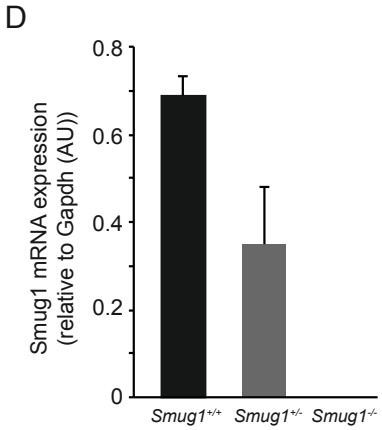
Wong, E., Yang, K., Kuraguchi, M., Werling, U., Avdievich, E., Fan, K., Fazzari, M., Jin, B., Brown, A.M.C., Lipkin, M., et al. (2002). Mbd4 inactivation increases C->T transition mutations and promotes gastrointestinal tumor formation. Proc. Natl. Acad. Sci. U. S. A. *99*, 14937–14942.

Xu, J. (2005). Preparation, culture, and immortalization of mouse embryonic fibroblasts. Curr. Protoc. Mol. Biol. Ed. Frederick M Ausubel Al *Chapter 28*, Unit 28.1.


A
PciI
ATG
1
STOP
2
PciI
9.3 kb
Smug1 endogenous locus
3´EXT probe
PciI
ATG
1
STOP
2
FRT
neo
FRT
PciI
loxP
loxP
6.3 kb
Smug1 recombined locus
PciI
PciI
loxP
2.1 kb
Smug1 Cre-mediated
excised locus


B
Smug1+/-
Smug1+/+
9.3 kb
6.3 kb
2.1 kb


C
DNA marker
negative
DNA marker
wt/wt
ko/wt
ko/ko
1500
1000
500
wt
ko


D
Smug1 mRNA expression
(relative to Gapdh (AU))
0.8
0.6
0.4
0.2
0
Smug1+/+
Smug1+/-
Smug1-/-

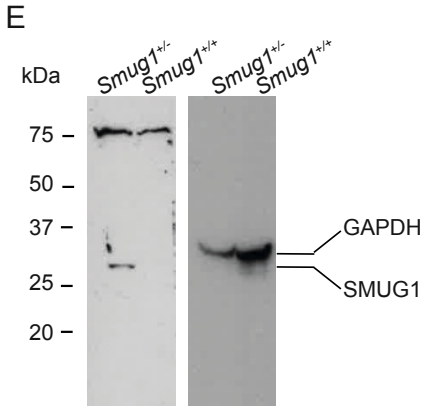

E
kDa
Smug1+/-
Smug1+/+
Smug1+/-
Smug1+/+
75
50
37
25
20
GAPDH
SMUG1

Figure 1

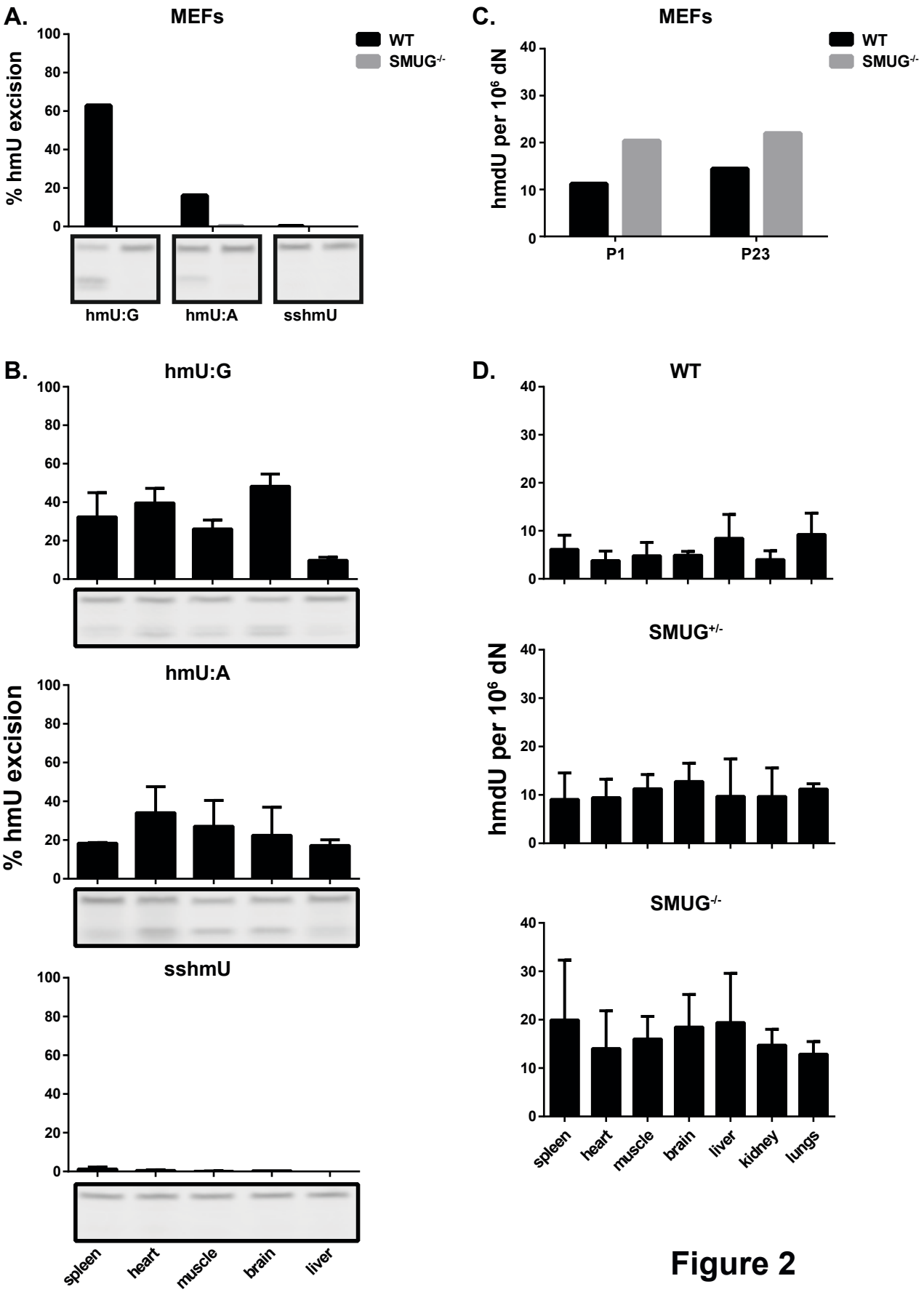

A.
MEFs
WT
SMUG-/-
% hmU excision
hmU:G
hmU:A
sshmU
B.
hmU:G
hmU:A
sshmU
% hmU excision
spleen
heart
muscle
brain
liver
C.
MEFs
WT
SMUG-/-
hmdU per 10^6 dN
P1
P23
D.
WT
SMUG+/-
SMUG-/-
hmdU per 10^6 dN
spleen
heart
muscle
brain
liver
kidney
lungs

Figure 2

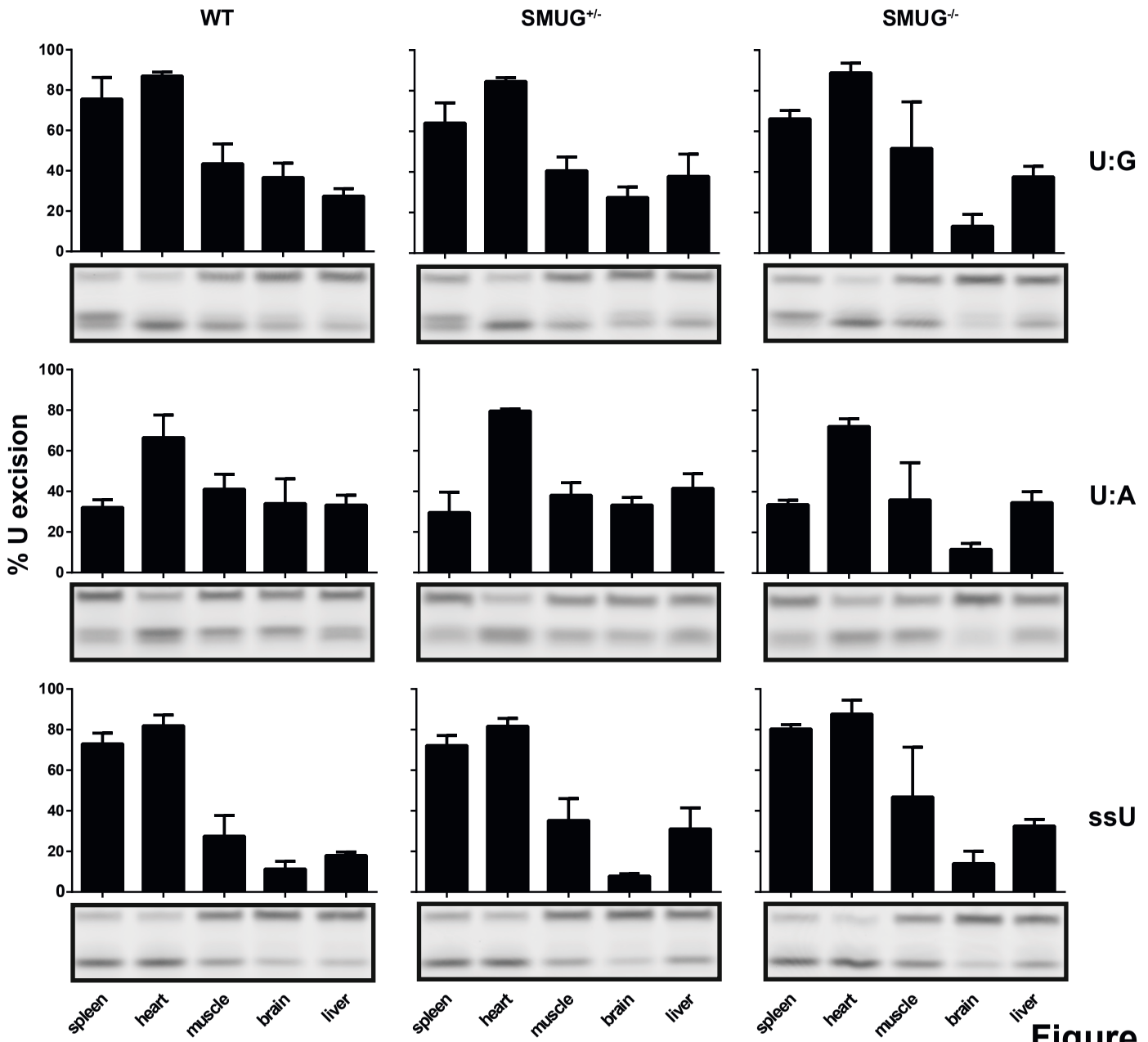

WT
SMUG+/-
SMUG-/-
% U excision
100
80
60
40
20
0
U:G
U:A
ssU
spleen
heart
muscle
brain
liver

Figure 3

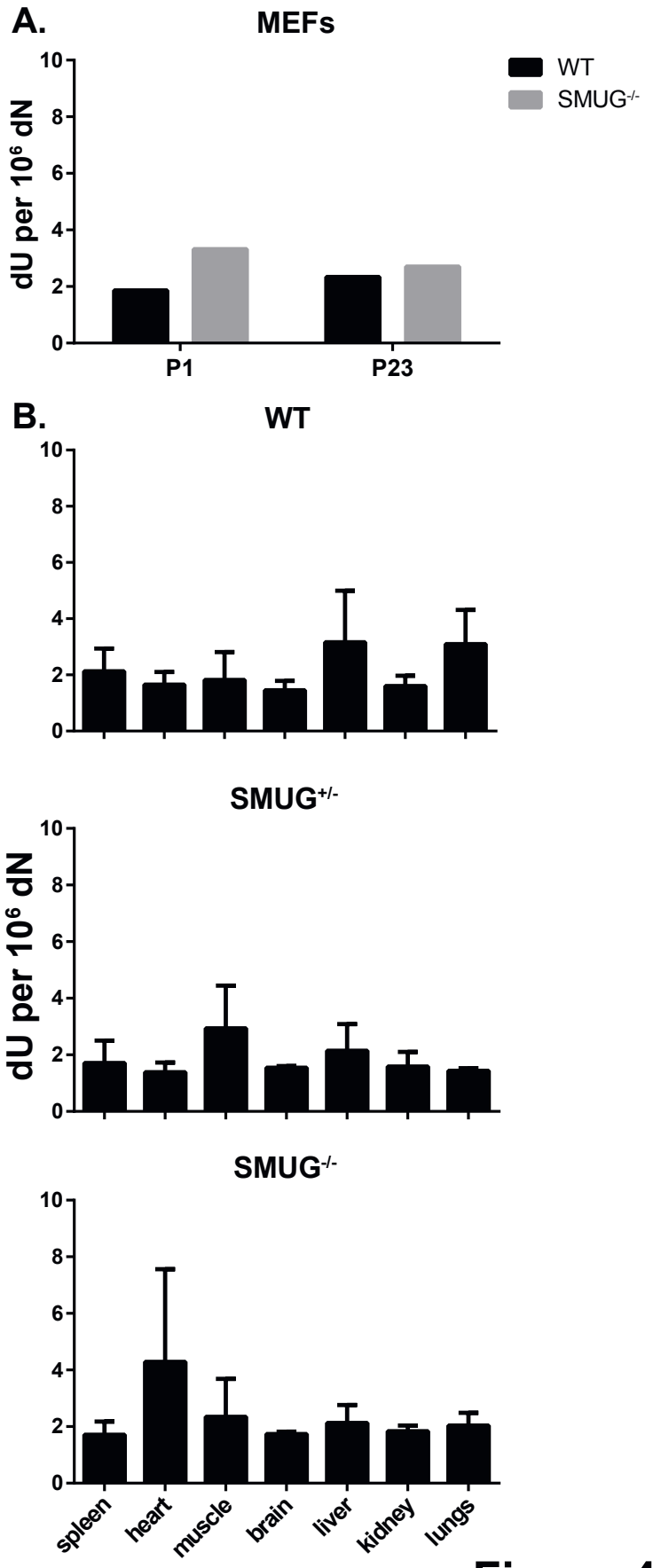

A.
MEFs
WT
SMUG-/-
dU per 10^6 dN
0
2
4
6
8
10
P1
P23
B.
WT
SMUG+/-
SMUG-/-
spleen
heart
muscle
brain
liver
kidney
lungs

Figure 4

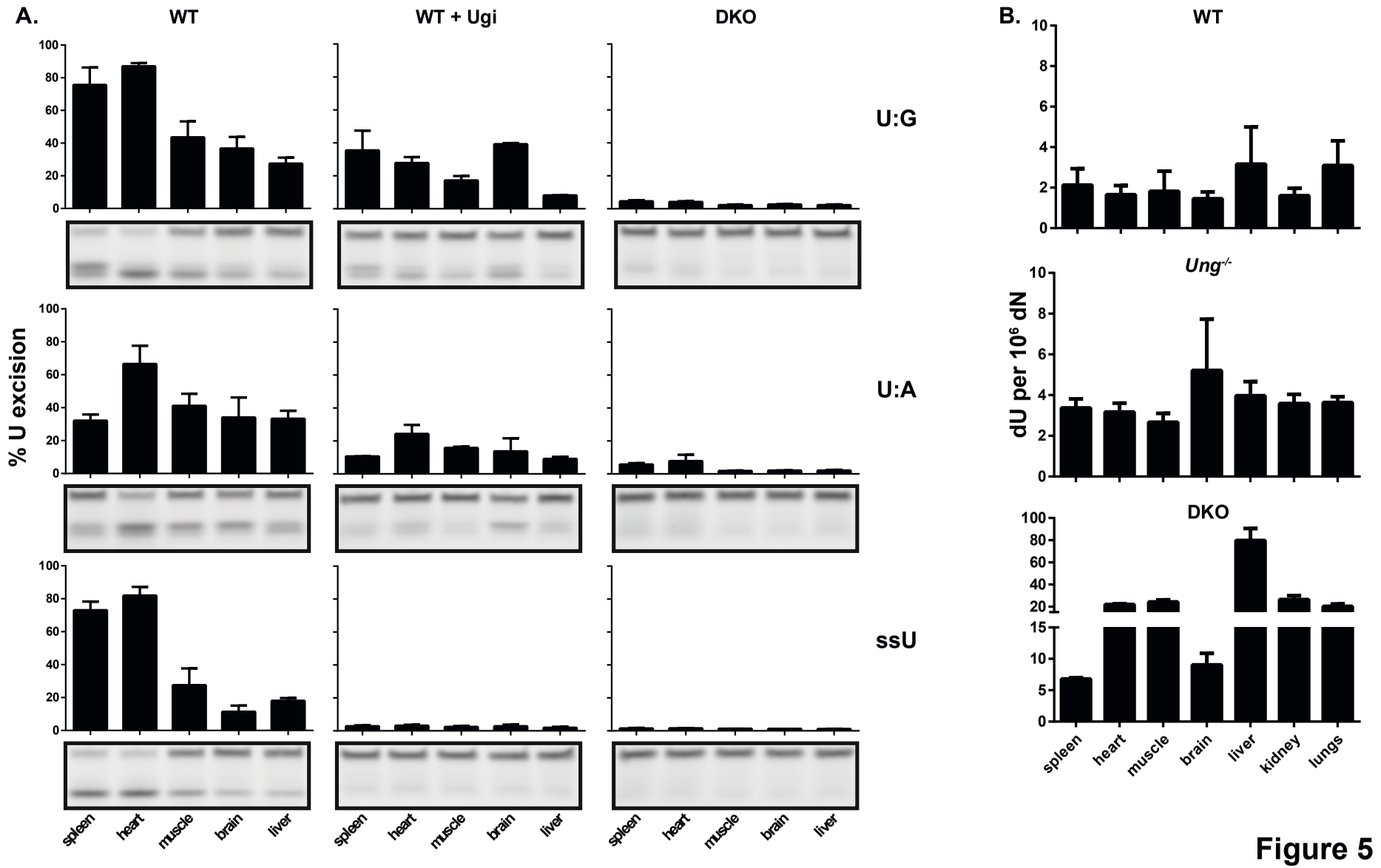
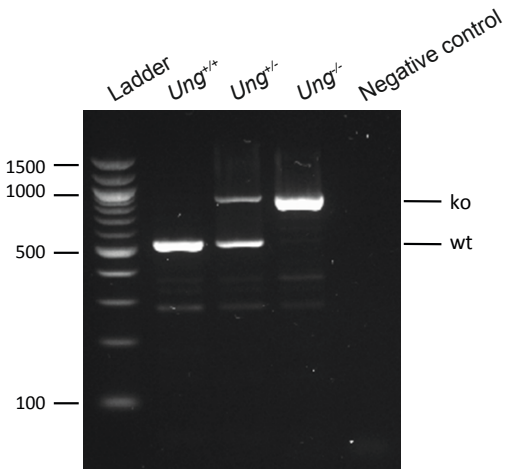
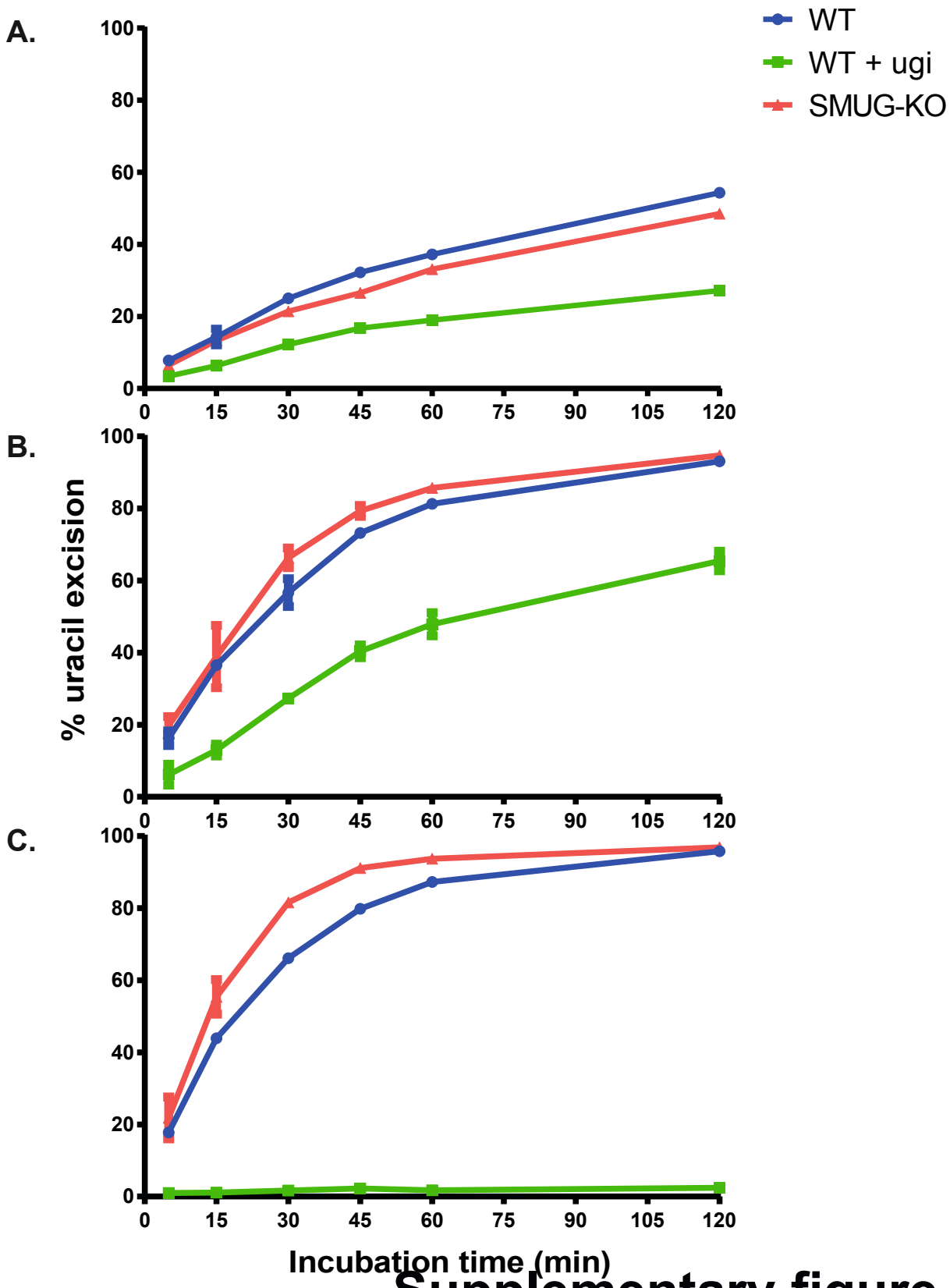

A.
WT
WT + Ugi
DKO
U:G
U:A
ssU
% U excision
spleen
heart
muscle
brain
liver
B.
WT
Ung-/-
DKO
dU per 10^6 dN
spleen
heart
muscle
brain
liver
kidney
lungs

Figure 5

Supplementary figure



Supplementary figure 1



Supplementary figure 2

Article IV

DNA Repair 25 (2015) 60–71

Contents lists available at ScienceDirect

DNA Repair

journal homepage: www.elsevier.com/locate/dnarepair

AID expression in B-cell lymphomas causes accumulation of genomic uracil and a distinct AID mutational signature

Henrik Sahlin Pettersen[a,c,d], Anastasia Galashevskaya[a], Berit Doseth[a], Mirta M.L. Sousa[a,e], Antonio Sarno[a], Torkild Visnes[a,1], Per Arne Aas[a], Nina-Beate Liabakk[a], Geir Slupphaug[a,e], Pål Sætrom[a,b], Bodil Kavli[a,*], Hans E. Krokan[a,*]

[a] Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway
[b] Department of Computer and Information Science, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway
[c] Liaison Committee between the Central Norway Regional Health Authority (RHA) and the Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[d] St. Olav's Hospital, Trondheim University Hospital, NO-7006 Trondheim, Norway
[e] The Proteomics and Metabolomics Core Facility (PROMEC) at NTNU, NO-7491 Trondheim, Norway

ARTICLE INFO

Article history:
Received 10 September 2014
Received in revised form 6 November 2014
Accepted 17 November 2014
Available online 24 November 2014

Keywords:
Activation-induced cytidine deaminase (AID)
Genomic uracil
B-cell lymphoma
Base excision repair
Kataegis
Mutational signature

ABSTRACT

The most common mutations in cancer are C to T transitions, but their origin has remained elusive. Recently, mutational signatures of APOBEC-family cytosine deaminases were identified in many common cancers, suggesting off-target deamination of cytosine to uracil as a common mutagenic mechanism. Here we present evidence from mass spectrometric quantitation of deoxyuridine in DNA that shows significantly higher genomic uracil content in B-cell lymphoma cell lines compared to non-lymphoma cancer cell lines and normal circulating lymphocytes. The genomic uracil levels were highly correlated with AID mRNA and protein expression, but not with expression of other APOBECs. Accordingly, AID knockdown significantly reduced genomic uracil content. B-cells stimulated to express endogenous AID and undergo class switch recombination displayed a several-fold increase in total genomic uracil, indicating that B cells may undergo widespread cytosine deamination after stimulation. In line with this, we found that clustered mutations (*kataegis*) in lymphoma and chronic lymphocytic leukemia predominantly carry AID-hotspot mutational signatures. Moreover, we observed an inverse correlation of genomic uracil with uracil excision activity and expression of the uracil-DNA glycosylases UNG and SMUG1. In conclusion, AID-induced mutagenic U:G mismatches in DNA may be a fundamental and common cause of mutations in B-cell malignancies.

1. Introduction

The only sources of uracil in DNA were previously thought to be misincorporation of dUMP during DNA replication and spontaneous deamination of DNA cytosine. The discovery of activation-induced cytidine deaminase (AID, also called AICDA) and several other APOBEC-family enzymes as probable DNA-cytosine deaminases introduced a third possible source (reviewed in [1]).

AID was first identified following induction of class switch recombination (CSR) in the CH12 mouse B-cell lymphoma cell line and initially thought to be an RNA-editing enzyme [2]. However, evidence that AID was a DNA mutator in *Escherichia coli* [3] and its functional interaction with uracil-DNA glycosylase UNG in adaptive immunity [4–6], indicated that AID is a DNA-cytosine deaminase. Later several of the other known APOBEC-family enzymes were also found to be DNA-cytosine deaminases *in vitro* [7,8]. DNA cytosine deamination by APOBEC-family enzymes is a natural event in both the adaptive and innate immune systems, through targeted deamination of immunoglobulin (Ig) genes by AID and deamination of viral DNA by APOBEC enzymes, respectively [7]. Despite their important physiological functions, these host defense mechanisms entail a high risk of potentially carcinogenic off-target genomic mutagenesis. Recent high-throughput sequencing of large numbers of human cancer genomes showed that mutations at cytosine residues, particularly C to T transitions, are the most prevalent

* Corresponding authors. Tel.: +47 72 57 30 74/+47 72 573221; fax: +47 72 57 64 00.
E-mail addresses: bodil.kavli@ntnu.no (B. Kavli), hans.krokan@ntnu.no (H.E. Krokan).
[1] Present address: Science for Life Laboratory, Division of Translational Medicine and Chemical Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-17121 Stockholm, Sweden.

http://dx.doi.org/10.1016/j.dnarep.2014.11.006

mutations in human cancer, highlighting enzymatic deamination of cytosine to uracil as a potential source of mutagenesis [9–11]. However, the actual uracil level in normal and various cancer genomes has remained elusive.

Here, a sensitive LC/MS/MS-based method for quantification of genomic 2′-deoxyuridine (dUrd) was applied to demonstrate that B-cell lymphoma cell lines contain several-fold increased levels of genomic uracil compared to normal human lymphocytes and non-lymphoma cell lines. Genomic uracil content correlated with AID protein expression but not with other APOBEC enzymes. In accordance with AID-generated uracil, we found that regions of clustered mutations (*kataegis*) in lymphoma and chronic lymphocytic leukemia (CLL) have a distinct AID-hotspot mutational signature. Importantly, we also show that uracil excision capacity and expression of the uracil-DNA glycosylases UNG and SMUG1 correlated negatively with genomic uracil levels and to some extent diminished the effect of AID. This study provides direct mechanistic evidence for genomic uracil accumulation due to enzymatic DNA cytosine deamination in human cancers.

2. Materials and methods

2.1. Primary cells, cell lines, cultivation, and reagents

Human cell lines HeLaS3 (ATCC CCL-2.2™), HEK293T (ATCC CRL-11268™), and U2OS (ATCC HTB-96™) were from ATCC. L428 (DSMZ ACC 197), DU145 (DSMZ ACC 261), KARPAS422 (DSMZ ACC 32), T24 (DSMZ ACC 376), DOHH2 (DSMZ ACC 47), SUDHL4 (DSMZ ACC 4956), JJN3 (DSMZ ACC 541), SUDHL5 (DSMZ ACC 571), SUDHL6 (DSMZ ACC 572 6), RAMOS (DSMZ ACC 603), RL (DSMZ ACC 613), DAUDI (DSMZ ACC 78 5), A431 (DSMZ ACC 91) were from DSMZ. OCILY3 was a gift from Dr. L.M. Staudt, Metabolism Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. Peripheral blood mononuclear cells (PMBCs) were purified from buffy coats from three healthy blood donors using the Lymphoprep™ (Progen) kit according to the manufacturer's protocol. Human B-lymphocytes were purified from buffy coats from three healthy blood donors using a negative selection kit from StemCell Technologies according to the manufacturer's protocol. HeLaS3, HEK293T, T24, A431, DU145, and U2OS cells were cultured in DMEM (4500 mg/l glucose) with 10% FCS, 0.03% L-glutamine, 0.1 mg/ml gentamicin and 2.3 μg/ml fungizone at 37 °C and 5% CO_2. DAUDI, DOHH-2, KARPAS, RAMOS, SU-DHL-4, SU-DHL-6, OCILY-3, L-428, RL, SU-DHL-5, and JJN3 cells were cultured in RPMI-1640 with 4500 mg/l glucose, 0.03% L-glutamine, Pen-Strep (1× final), 0.1 mg/ml gentamicin, and 2.3 μg/ml fungizone, and 20% heat inactivated (56 °C, 20 min) FCS at 37 °C and 5% CO_2. For quantitative rtPCR and uracil measurements cells were harvested at densities between 750 000 and 2 million cells/ml.

Cell doubling times for suspension cells were measured using a *Countess*® cell counter (Invitrogen) by two parallel daily measurements for three to five day periods from cell densities of 50 000–200 000 cells/ml to one to three million cells/ml. For adherent cells, doubling time was measured in 96 well plates (3–6 parallel wells; starting density 50 000 cells/ml) for a three day period by daily fluorescent measurement of resazurin (Sigma) metabolism according to the manufacturer's protocol. Doubling times were calculated by exponential regression.

SUDHL5 AID knockdown and control cells were made using Open Biosystem TransLenti Viral Packaging Mix, pTRIPZ AICDA shRNA (RHS4741-EG57379; vectors V2THS_58282, 58283, and 58319) or pTRIPZ non-silencing control vector according to the manufacturer's protocol. Briefly, lentiviruses were produced in HEK293T cells, and then supernatant from three consecutive days 48 h after HEK293T transfection were used to infect SUDHL5 cells. Infected SUDHL5 cells were amplified for another 48 h and then selected with 2 μg/ml puromycin for 30 days. Expression was induced with 1 μg/ml doxycycline.

CH12F3 AID-EYFP and EYFP stable transfectants, confocal microscopy, and stimulation experiments were described previously [12]. CH12F3 cells (2×10^6 cells/ml) were cultured in RPMI medium, with 10% heat-inactivated fetal calf serum, 0.03% L-glutamine, 50 μM β-mercaptoethanol, 1 mM Na-pyruvate, 0.1 mg/ml penicillin/streptomycin, 2.3 μg/ml fungizone, and 1.0 mg/ml G418. CH12F3 cells were stimulated to undergo class switch recombination by adding 10 ng/ml mouse recombinant IL-4 (Peprotech), 2 μg/ml anti-mouse CD40 monoclonal antibody (BD Biosciences) and 1 ng/ml human TGF-β1 (Peprotech) and harvested 48 h post stimulation for DNA and protein isolation. Western analysis of AID protein expression was performed using mouse anti-AID monoclonal antibody no. 39-2500, clone ZA001, 500 μg/ml (Invitrogen). Nuclear extracts from synchronized HeLa cells were prepared essentially as described [13,14].

2.2. RNA isolation and quantitative real-time PCR (qRT-PCR)

Total RNA for mRNA analysis was prepared using the mirVana miRNA isolation kit (Ambion) according to the manufacturer's instructions. RNA concentration and quality was measured on a NanoDrop ND-1000 UV–vis spectrophotometer. Total RNA (770 ng) was reverse transcribed for gene expression analysis using TaqMan reverse transcription reagents (Applied Biosystems). The following TaqMan gene expression assays (Applied Biosystems) were used: AID (Hs00757808_m1), UNG (Hs00422172_m1), SMUG1 (Hs04274951_m1), TDG (Hs00702322_s1), MBD4 (Hs00187498_m1), APOBEC1 (Hs00242340_m1), APOBEC2 (Hs00199012_m1), APOBEC3A (Hs00377444_m1), APOBEC3B (Hs00358981_m1), APOBEC3C (Hs00819353_m1), APOBEC3D (Hs00537163_m1), APOBEC3G (Hs00222415_m1), APOBEC3F (Hs01665324_m1), APOBEC3H (Hs00419665_m1), APOBEC4 (Hs00378929_m1), and GAPDH (Hs99999905_m1). Quantitative PCR was carried out on a Chromo4 (BioRad) real-time PCR detection system. Relative expression of mRNA was calculated by the ΔCt method using GAPDH as endogenous control. Regression analyses were done using GraphPad Prism where data were fitted by linear regression (log/linear(*X*) *vs.* log/linear(*Y*)) as indicated.

2.3. Quantification of uracil in DNA by LC/MS/MS

Genomic uracil was quantified as previously described [15]. Briefly, DNA was isolated by phenol:chloroform:isoamyl extraction, treated with alkaline phosphatase to remove free deoxyribonucleosides, and then enzymatically hydrolyzed to deoxyribonucleosides. Deoxyuridine (dU) was then separated from deoxycytidine (dC) by HPLC fractionation using a reverse-phase column with embedded weak acidic ion-pairing groups (2.1 mm × 150 mm, 5 μm, Primesep 200, SIELC technologies), using a water/acetonitrile gradient containing 0.1% formic acid. The dU fraction was finally analyzed by ESI-LC/MS/MS using a reverse phase column (2.1 mm × 150 mm, 3.5 μm, Zorbax SB-C18, Agilent Technologies), using a water/methanol gradient containing 0.1% formic acid on an API5000 triple quadrupole mass spectrometer (Applied Biosystems) in positive ionization mode. A small fraction of the hydrolyzed deoxyribonucleosides were quantified by LC/MS/MS in parallel and used to determine the amount of dU per 10^6 deoxyribonucleosides.

2.4. In vitro uracil DNA excision activity and complete BER assays

Standard UDG activity assay was performed as described [16]. Briefly, 20 μl reaction mixtures containing (final) 1.8 μM nick translated [^{3}H]-dUMP-labeled calf thymus DNA (U:A substrate), 1× UDG buffer (20 mM Tris–HCl, pH 7.5, 60 mM NaCl, 1 mM DTT, 1 mM EDTA, 0.5 mg/ml BSA) and 1 μg whole cell extract were incubated at 30 °C for 10 min. Acid-soluble [^{3}H] uracil was quantified by scintillation counting. Whole cell extracts was prepared as described [17]. Oligodeoxynucleotide UDG assays were performed as described [16]. Briefly, double-stranded DNA substrates were generated by annealing 6-FAM-labeled oligonucleotides containing a centrally positioned uracil in an AID-hotspot (5′-CATAAAGAGUTAAGCCTGG-3′; Eurogentec) to complementary strands containing G opposite U. Activity was measured in 10 μl assay mixtures containing (final) 20 nM substrate, 1× UDG buffer and 0.4 μg cell extract, and incubated at 37 °C for 10 min. Reactions were stopped and AP-sites were cleaved by addition of 50 μl 10% piperidine followed by incubation at 90 °C for 20 min. Product and substrate were separated on PAGE, scanned on Typhoon Trio imager and quantified using ImageQuant TL software (GE healthcare).

BER assays were carried out essentially as described [14,17]. Briefly, 10 μg nuclear extract was incubated with 250 ng cccDNA (covalently closed circular DNA) substrates in final concentrations of 40 mM HEPES-KOH, 70 mM KCl, 5 mM $MgCl_2$, 0.5 mM DTT, 2 mM ATP, 20 μM dATP, 20 μM dGTP, 20 μM dTTP, 8 μM dCTP, 4.4 mM phosphocreatine, 62.5 ng/μl creatine kinase and 50 nCi/μl [α-^{32}P]dCTP in a final volume of 40 μl. Reactions were incubated for 25 min at 37 °C and stopped by addition of EDTA (18 mM final) and 6 μg RNaseA and incubated at 37 °C for 10 min followed by the addition of SDS (0.5% final) and 12 μg proteinase K. DNA was extracted by phenol/chloroform and precipitated in ammonium acetate/ethanol and digested with XbaI and HincII (New England Biolabs). Following 12% PAGE, bands were visualized and quantified using ImageQuant software (Fujifilm). We investigated relative contribution of SMUG1, TDG and UNG2 to the initiation of uracil repair by pre-incubating extracts with neutralizing antibodies to SMUG1 (0.11 μg/μl final concentration), UNG (0.3 μg/μl final concentration), and/or neutralizing anti-serum towards TDG (1:50 dilution) on ice for 30 min prior to the reaction.

2.5. Flow cytometric analysis of cell cycle

Cells were fixed in 70% methanol, washed twice with PBS, and then treated with 50 μl RNaseA (100 μg/ml in PBS) at 37 °C for 30 min prior to DNA staining with 200 μl propidium iodide (50 μg/ml in PBS) at 37 °C for 30 min. Cell cycle analyses were performed using a FACS Canto flow cytometer (BD-Life Science).

2.6. Sample preparation and targeted mass spectrometry

Cell pellets were resuspended in 1× packed cell volume in buffer I: 10 mM Tris–HCl pH 8.0, 200 mM KCl, 1× complete protease inhibitor, and 5× phosphatase-inhibitor cocktails I and II (Sigma-Aldrich), 10 μM suberoylanilide hydroxamic acid (SAHA) (Cayman Chemicals) and 0.05 μM, Ubiquitin Aldehyde (Biomol International LP) followed by addition of an equal final volume of buffer II: 10 mM Tris–HCl pH 8.0, 200 mM KCl, 10 mM EGTA, 10 mM $MgCl_2$, 40% glycerol, 0.5% NP40, 1 mM DTT, 1× complete protease inhibitor, and 5× phosphatase-inhibitor cocktails I and II (Sigma-Aldrich), 10 μM suberoylanilide hydroxamic acid (SAHA) (Cayman chemicals) and 0.05 μM, Ubiquitin Aldehyde (Biomol International LP) containing an endonuclease cocktail of 200 U Omnicleave (Epicenter Technologies), 2 U DNase I (Roche Inc.), 250 U Benzonase (EMD), 100–300 U micrococcal nuclease (Sigma-Aldrich), and 10 μg RNase A (Sigma-Aldrich) per 1 ml of buffer II. After resuspension, the lysates were incubated for 1.5 h at 4 °C in a roller. 50 μg protein of cell lysate pools consisting of 2–4 biological replicates from each cell line were incubated with 5 mM tris (2-carboxyethyl) phosphine (TCEP) for 30 min followed by alkylation with 1 μmol/mg protein of iodoacetamide for 45 min in the dark. Proteins were precipitated using a methanol–chloroform method as described [18], including another round of reduction and alkylation prior to overnight digestion with Trypsin (Promega) at 1:40 ratio (w/w, enzyme:protein) at 37 °C. Tryptic digests were dried out, resuspended in 0.1% formic acid and analyzed on a Thermo Scientific QExactive mass spectrometer operating in Targeted-MS2 mode coupled to an Easy-nLC 1000 UHPLC system (Thermo Scientific/Proxeon). Peptides (2 μg) were injected onto a Acclaim PepMap100 C-18 column (75 μm i.d. × 2 cm, C18, 5 μm, 100 Å) (Thermo Scientific) and further separated on a Acclaim PepMap100 C-18 analytical column (75 μm i.d. × 50 cm, C18, 3 μm, 100 Å) (Thermo Scientific). A 120 min method was used and consisted of a 300 nl/min flow rate, starting with 100% buffer A (0.1% Formic acid) with an increase to 5% buffer B (100% Acetonitrile, 0.1% Formic acid) in 2 min, followed by an increase to 35% Buffer B over 98 min and a rapid increase to 100% buffer B in 6 min, where it was held for 5.5 min. The solvent composition was quickly ramped to 0% buffer B, where it was subsequently held for 8 min to allow the column to equilibrate for the next run. The peptides eluting from the column were ionized by using a nanospray ESI ion source (Proxeon, Odense) and analyzed on the QExactive operating in positive-ion mode using electrospray voltage 1.9 kV and HCD fragmentation. Each MS/MS scan was acquired at a resolution of 35 000 FWHM, normalized collision energy (NCE) 28, automatic gain control (AGC) target value of 2×10^5, maximum injection time of 120 ms and isolation window 2 *m*/*z*.

All parallel reaction monitoring (PRM)-based targeted mass spectrometry methods were designed, analyzed, and processed using Skyline software version 2.5 [19]. *In silico* selection of proteotypic peptides was performed *via* Skyline using the *Homo sapiens* reference proteome available at www.uniprot.org to exclude non-unique peptides. Frequently modified peptides, such as those containing methionine, and peptides containing continuous sequences of R and K (*e.g.*, KR, RK, KK or RR) were avoided. However, when the inclusion of non-ideal peptides was necessary both unmodified and M-oxidized peptides as well as peptides containing a missed cleavage site were analyzed. Synthetic purified peptides (JPT Peptide Technologies) and tryptic digests from recombinant proteins were analyzed in a QExactive mass spectrometer. Information on retention time and fragmentation pattern of the top 2–6 ionizing tryptic peptides (2+ or 3+ charge states) for each protein were used to build a scheduled method with a retention time window of 5 min. The method was then used for peptide quantification in the cell lysate pools. A minimum of 2 peptides per protein was used for quantitative analysis except for APOBEC3F in which only one of the unique peptides tested was detectable in the samples. The sum of the integrated peak areas of the 3–5 most intense fragments was used for peptide quantification. Peptide areas for multiple peptides of the same protein were summed to assign relative abundance to that protein. The error bars represent the standard deviation of 3 technical replicates.

2.7. Bioinformatics analysis of DNA exome sequencing data

Kataegis regions and somatic mutations for CLL, B-Cell lymphoma, ALL, lung adenocarcinoma, and breast, liver, and pancreatic cancer were downloaded from the supplementary material of a published study [11]. The *kataegis* regions within specific cancer samples were provided as genomic coordinates into the human reference genome version 19 (hg19); the somatic mutations were provided as genomic coordinates in hg19 and nucleotide alterations. We used the following procedure to create mutational

signatures for the *kataegis* regions for each cancer type. First, for each *kataegis* region, its sample ID and genomic coordinates were used to identify the corresponding somatic mutations. Second, for each somatic mutation, the five nucleotides centered on the mutated nucleotide were retrieved from the genome sequence. Third, if the middle nucleotide within the retrieved sequence was a purine, the sequence was reverse-complemented such that all the mutations were represented by pyrimidines. Fourth, for each of the six possible single nucleotide mutations, the relative occurrence of each nucleotide at each position within the retrieved sequences was computed. These position-specific relative occurrences were the mutational signatures.

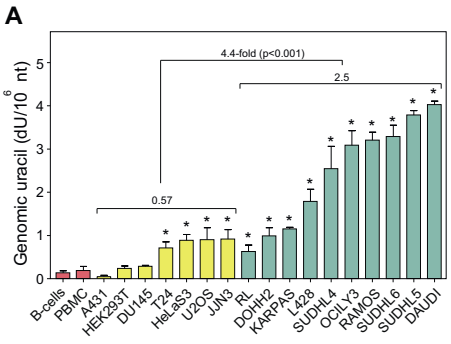
3. Results

3.1. High genomic uracil levels in B-cell lymphoma cells

To investigate whether uracil in the genome may be an important factor in lymphomagenesis, we measured genomic uracil in ten B-cell lymphoma cell lines, seven other human transformed cell lines and in lymphocytes from three healthy human blood donors (Fig. 1A). The origin and major characteristics of cell lines is displayed in Fig. 1B. We found as much as 72-fold variation in genomic uracil levels between the cell line with the highest uracil content (DAUDI, 4.03 deoxyuridines (dU) per 10^6 deoxyribonucleosides (nt)) and the cells with the lowest level of genomic uracil (A431, 0.056 dU/10^6 nt). Strikingly, all ten lymphoma cell lines and four of the other transformed cell lines had significantly ($P < 0.05$) elevated genomic uracil levels compared to genomic uracil in peripheral blood mononuclear cells (PBMC) from the mean value for three blood donors (0.19 dU/10^6 nt). We also measured genomic uracil in B-lymphocytes isolated from buffy coats from three healthy donors, using a kit for negative selection. The genomic mean uracil level in these was 0.14 dU/10^6 nt, with individual values of 0.07, 0.17 and 0.19 dU/10^6 nt, respectively. The mean value for the genomic uracil level in B-cell lymphoma cell lines (2.5 dU/10^6 nt) was 13-fold and 18-fold higher than in PBMC and primary B-cells, respectively. In addition it was significantly higher (4.4-fold, $P < 0.001$) than the mean for non-lymphoma cell lines (0.57 dU/10^6 nt). The B-cell lymphoma cell lines are likely to be exposed to enzymatic untargeted cytosine deamination by AID throughout the genome, since the total number of genomic uracils is in the range 3000–15 000 per haploid genome (this paper) and the density of genomic uracil in S_μ region of stimulated B-cells is only ~0.8 per kb [20]. Some of the non-lymphoma cancer cell lines had intermediate genomic uracil levels, clearly higher than normal peripheral blood lymphocytes, but lower than most of the B-cell lymphoma cell lines.

3.2. AID expression correlates with genomic uracil accumulation

AID has previously been shown to be expressed in several lymphoma subtypes [21–24] and AID/APOBEC family enzymes were suggested to contribute to mutational signatures in a number of cancers by deaminating cytosine to uracil in DNA [11]. We therefore investigated whether expression of AID and/or other APOBECs could explain the observed variation in genomic uracil levels in the cell line panel. We first measured mRNA expression of AID and all other APOBEC-family genes by quantitative rtPCR using GAPDH as reference gene (Fig. 2A). AID mRNA was detected in all 17 cell lines, although at highly variable levels, but not in the normal lymphocytes from blood donors. Furthermore, AID mRNA was substantially increased in lymphomas with high genomic uracil such that AID mRNA had a high positive correlation with genomic uracil ($R^2 = 0.70$, $P < 0.0001$). By contrast, APOBEC3B, -3D, -3F, and -3G mRNA content did not correlate with genomic uracil level although



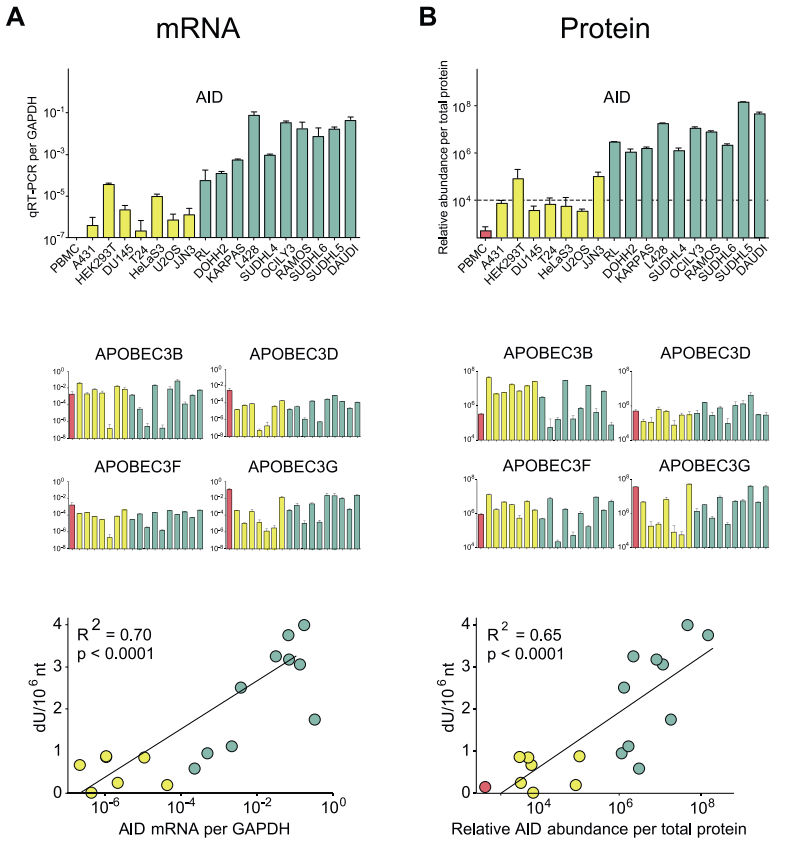
B

Name	Cell/tissue type	Disease
B-cells	B-lymphocyte/blood	Healthy blood donors
PBMC	White blood cells/blood	Healthy blood donors
A431	Epithelial/skin	Epidermoid carcinoma
HEK293T	Embryonic kidney	NA
DU145	Epithelial/prostate	Prostate cancer
T24	Epithelial/bladder	Bladder carcinoma
HeLaS3	Epithelial/cervix	Adenocarcinoma
U2OS	Osteocyte/bone	Osteosarcoma
JJN3	Plasma cell/bone marrow	Plasma cell leukemia
RL	B-cell/lymphatic	B-NHL
DOHH2	B-cell/lymphatic	B-NHL
KARPAS	B-cell/lymphatic	B-NHL
L428	B-cell/lymphatic	Hodgkin lymphoma
SUDHL4	B-cell/lymphatic	B-NHL
OCILY3	B-cell/lymphatic	B-NHL
RAMOS	B-cell/lymphatic	Burkitt lymphoma
SUDHL6	B-cell/lymphatic	B-NHL
SUDHL5	B-cell/lymphatic	B-NHL
DAUDI	B-cell/lymphatic	Burkitt lymphoma

Fig. 1. Genomic uracil levels in B-cell lymphoma-/non-lymphoma cell lines and on white blood cells from peripheral blood l. (A) Quantification of genomic uracil levels (dU/10^6 nt) by LC–MS/MS in lymphoma cell lines (green), non-lymphoma cell lines (yellow) and PBMCs or B-lymphocytes isolated from buffy coats from blood donors (red). Asterisk (*) signifies measurements significantly ($P < 0.05$) different from average genomic uracil levels in PBMC from three healthy blood donors (Student's *T*-test). Error bars represent mean and SD of at least two biological replicates. Cell lines within each group are ordered along the x-axis according to increasing genomic uracil levels. (B) Overview of cell lines, PBMCs and B-lymphocytes used in the study and their origin. B-NHL: B-cell non-Hodgkin lymphoma. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

they were expressed in all cell lines as well as in the normal lymphocytes (Fig. 2A). mRNA of the other APOBECs (APOBEC1, APOBEC2, APOBEC3A, and APOBEC) were detected only in some of the cell lines and mostly at very low levels (data not shown).

Although mRNA expression data is useful as a predictor of protein expression, it does not always correlate with the actual protein levels in the cells. Thus, we quantified AID and the APOBEC proteins by parallel reaction monitoring using a quadrupole-Orbitrap mass spectrometer (Fig. 2B). This is a highly selective method allowing quantification of many protein targets in a single sample [25,26]. In agreement with mRNA data, MS quantification revealed higher amounts of AID protein in lymphoma cells with increased genomic uracil (Fig. 2B, upper panel). Furthermore, similar to mRNA data (Fig. 2A, middle panel), APOBEC3B, -3D, -3F and-3G proteins were expressed in all cell lines (Fig. 2B, middle panel), while APOBEC1, APOBEC2, APOBEC3A, and APOBEC4 were not detectable or detected at very low levels (data not shown). In general, protein levels for AID and the APOBEC proteins (normalized to GADPH protein) correlated well with mRNA levels (Fig. 2C). As an additional control, we also quantified AID by western analysis, which yielded results similar to the MS analysis (Fig. 2D). Linear regression



C

mRNA vs. protein

	R^2	p-value
AID	0.84	<0.0001
APOBEC3B	0.47	0.003
APOBEC3D	0.47	0.002
APOBEC3F	0.44	0.004
APOBEC3G	0.74	<0.0001

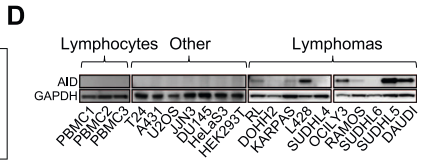


Fig. 2. Expression of AID and APOBECs, and correlation with genomic uracil. Expression of AID and APOBEC3B, 3D, 3F, and 3G mRNAs measured by qRT-PCR (A) or protein by mass spectrometric quantification. (B) Lymphoma cell lines are shown in green, non-lymphoma cell lines in yellow, and PBMC in red. Cell lines within each group are ordered along the *x*-axis according to increasing genomic uracil levels, as in Fig. 1. mRNA levels have been normalized to GAPDH mRNA, and protein levels to MS signal counts per total injected protein. Note that mRNA and protein expression data are in log-scale. Regression plots of genomic uracil (dU/10^6 nt) *vs.* AID mRNA and protein levels are presented in the lower panels in Fig. 2A and B, respectively. (C) Table of correlation coefficients between mRNA and protein expression for AID and other APOBECs. (D) Western analysis of AID protein expression with GAPDH shown as a loading control. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Regression analysis of genomic uracil levels (linear) *vs.* AID and APOBEC protein expression (log) normalized to total protein. Bold green indicates significant positive correlation (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

	All cell lines including PBMC		B-cell lymphoma cell lines		Non-lymphoma cell lines	
	R^2	*P*-value	R^2	*P*-value	R^2	*P*-value
AID	**0.65**	**<0.0001**	**0.42**	**0.04**	0.00	0.97
APOBEC3B	0.10	0.2089	0.00	0.84	0.00	0.98
APOBEC3D	0.12	0.17	0.00	0.88	0.02	0.79
APOBEC3F	0.01	0.67	0.08	0.44	0.32	0.18
APOBEC3G	0.12	0.14	0.30	0.09	0.00	0.98

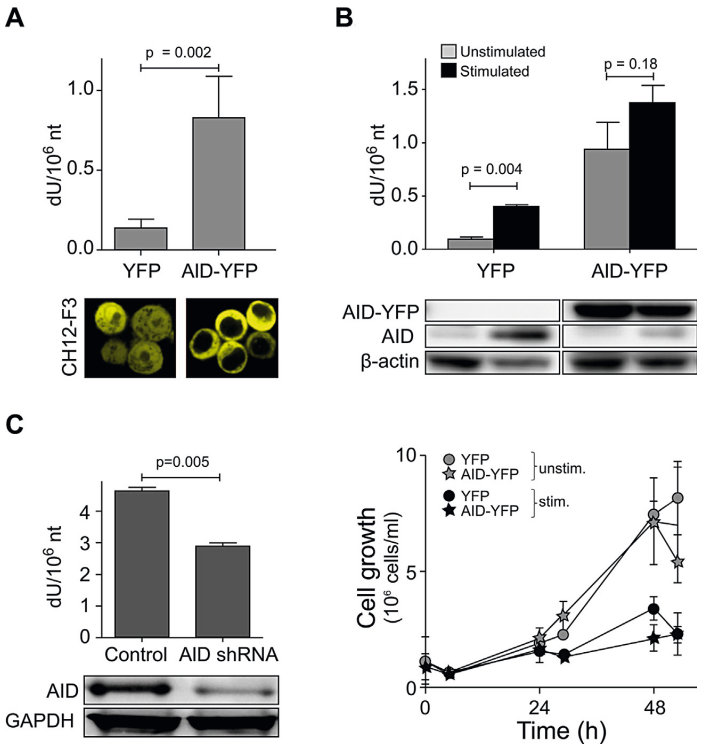


Fig. 3. Genomic uracil levels after stimulation of endogenous AID expression, AID-YFP overexpression, and AID knockdown. (A) Genomic uracil levels in DNA isolated from mouse lymphoma cells (CH12F3) stably transfected with AID-YFP or YFP, and confocal microscopy showing subcellular distribution of AID-YFP fusion protein or YFP. (B) Genomic uracil levels and cell growth of CH12F3 YFP cells and CH12F3 AID-YFP cells prior to stimulation and 48 h after being stimulated to undergo class switch recombination using mouse recombinant IL-4, CD40 monoclonal antibody and hTGF-β (upper panel) and western blots from one representative experiment showing AID protein expression levels and β-actin as loading control (middle panel). The lower panel shows cell growth of stimulated and unstimulated cells. Graphs represent mean and SD calculated from at least two biological replicates. *P*-values were calculated by a two-tailed Student's *T*-test. (C) Genomic uracil levels in SUDHL5 lymphoma cells stably transfected with AID-shRNA and control. Western blots shows AID protein expression levels with GAPDH as a loading control.

analysis of AID western signals against MS quantitation of AID protein revealed almost perfect correlation ($R^2 = 0.95$). Importantly, AID expression significantly correlated with genomic uracil also at the protein level ($R^2 = 0.65$, $P < 0.0001$) (Fig. 2B, and Table 1), and thereby seemed to account for a large part of the variation in genomic uracil between the cell lines. The correlation was still valid when including only the B-cell lymphoma cell lines in the regression analysis (Table 1). No significant correlations were observed between the other APOBEC proteins and genomic uracil (Table 1). Thus, AID was the only APOBEC-family member that correlated with genomic uracil in the human cancer cell lines examined here.

3.3. AID expression causes several-fold increases in genomic uracil

To investigate whether AID expression significantly increases the overall level of genomic uracil in an otherwise isogenic background, we used stable transfectants of the mouse B-cell lymphoma cell line CH12F3 expressing AID-YFP fusion protein, or YFP as control [27]. AID is mostly localized in the cytoplasm (Fig. 3A), but is actively imported into the nucleus where it may access the genome [12]. We found that the cells expressing AID-YFP displayed an almost six-fold higher level of genomic uracil compared to the YFP control (Fig. 3A). When appropriately stimulated, CH12F3 cells increase endogenous AID expression and have capacity to undergo CSR. Thus next, we investigated whether stimulation of these cell lines also increased the level of genomic uracil. A clear induction of AID and a four-fold increase in genomic uracil were observed in stimulated CH12F3-YFP cells already after 48 h (Fig. 3B, upper panel). An increase in genomic uracil was observed in the stimulated AID-YFP expressing cells as well, although this was not significant, probably due to the high constitutive expression AID-YFP. Importantly, the increase in genomic uracil observed after stimulation could not be ascribed to increased replicative misincorporation of dUMP due to higher proliferation rate because stimulated CH12F3 cells actually have reduced proliferation (Fig. 3B, lower panel). Finally, we examined the effect of knocking down AID using a lentiviral AID shRNA expressing vector. For this experiment, we used the human B-cell lymphoma cell line SUDHL5, which exhibited high constitutive AID expression (Fig. 2B and D). We found that a 60% knockdown of AID reduced genomic uracil level by 38% ($P = 0.005$; Fig. 3C). Taken together these results strongly support the view that enzymatic cytosine deamination is the major source of genomic uracil in AID-expressing cells.

3.4. Uracil-DNA repair capacity is inversely correlated with genomic uracil levels

Genomic uracil is predominantly repaired by base excision repair (BER), which is mainly initiated by the uracil-DNA glycosylase encoded by the *UNG* gene [16]. We have previously shown that UNG deficiency in human and mouse cells results in a several-fold

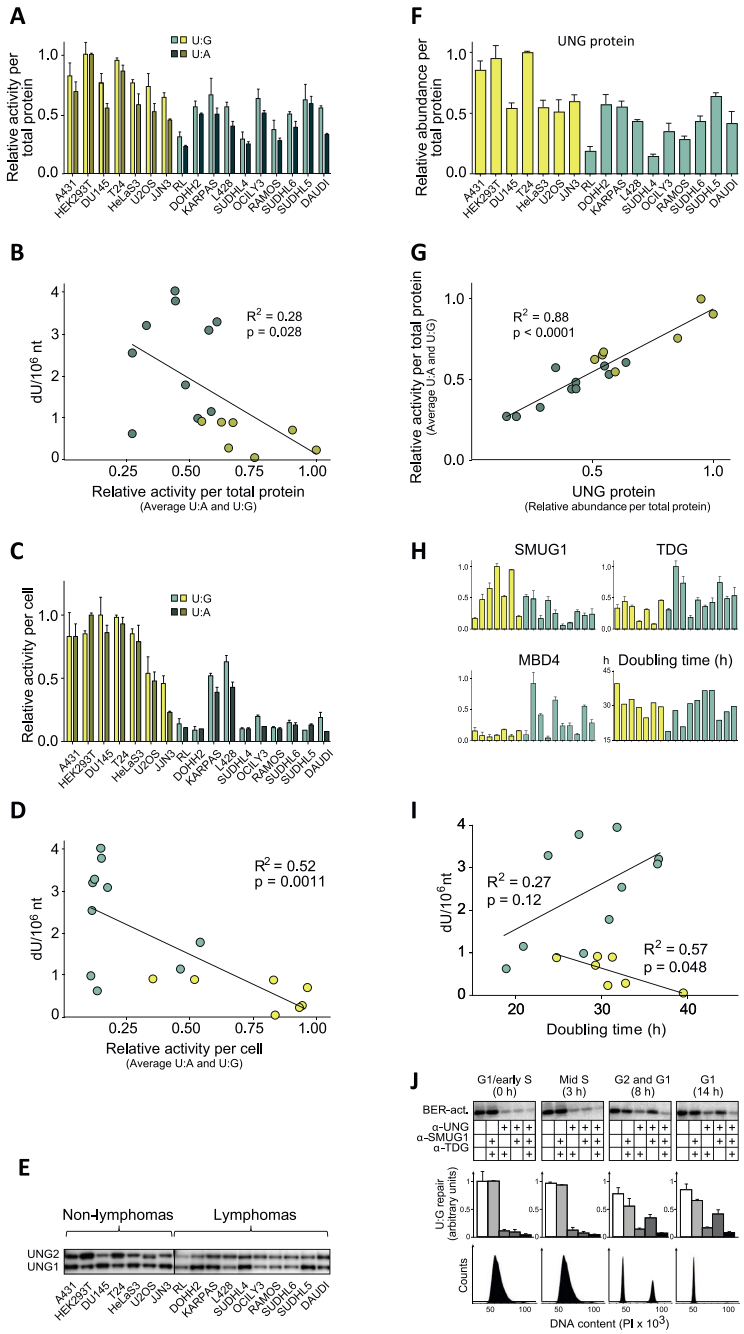


Fig. 4. Uracil excision activity, expression of uracil DNA glycosylases, and correlation with genomic uracil levels. Note that in all bar graphs cell lines are ordered according to increasing genomic uracil levels in lymphoma cell lines (green) and non-lymphoma cell lines (yellow), and *Y*-axes are normalized so that maximum activity or maximum protein abundance equals 1. Bars and error bars represent mean and SD of three biological replicates. (A) Relative uracil excision activity from an AID-hotspot sequence-oligomer containing uracil in U:G context (cleavage assay) and from a nick-translated DNA containing uracil in U:A context (^{3}H-uracil release assay), as indicated by color codes. Activity was normalized to total protein. (B) The corresponding correlation between genomic uracil and activity per total protein. (C) Relative uracil excision activity normalized to activity per cell, and (D) the corresponding correlation with genomic uracil with activity per cell. (E) Western blot of UNG2 and UNG1 in non-lymphoma and lymphoma cell lines. (F) Relative abundance of MS-quantified UNG protein per total protein; (G) Correlation plot of average uracil excision activity *vs.* relative abundance of MS quantified UNG protein. (H) Relative abundance of MS quantified DNA glycosylases SMUG1, TDG and MBD4 and cell doubling times of cell lines; (I) Correlation plot of genomic uracil content *vs.* doubling times of non-lymphoma cell lines and lymphoma cell lines. (J) Contribution of UNG, SMUG1 and TDG through the cell cycle measured by

increase in genomic uracil [15]. The other uracil-DNA glycosylases, *i.e.* SMUG1, TDG, and MBD4, are thought to be quantitatively less important contributors, at least in proliferating cells [16,28,29]. Furthermore, the DNA repair machinery has been shown to protect against AID-induced mutagenesis [30–32]. Therefore, we measured uracil excision activity of cell free extracts prepared from all cell lines against oligodeoxyribonucleotides with uracil in a U:G context. In addition, we measured [^{3}H]-uracil release from calf thymus DNA having uracil in a U:A context. The two different assays gave similar activity profiles (Fig. 4A). Regression analysis of uracil-excision activity (relative to protein content in the cell extracts) against genomic uracil content in the cells demonstrated a negative correlation (Fig. 4B), which is significant ($P < 0.05$), although weak. We also calculated relative uracil excision activity per cell since the glycosylases are predominantly nuclear enzymes and the cells tested vary in size and nucleus-to-cytoplasm ratios (Fig. 4C). Using these activity values, a stronger correlation with genomic uracil level was observed (Fig. 4D).

The *UNG* gene encodes both nuclear UNG2 and mitochondrial UNG1, having identical catalytic domains but specific N-terminal domains. These isoforms are differently regulated from two promoters [33,34]. Since activity assays measure total activity, we analyzed the isoforms by western blots. Nuclear UNG2, which is the isoform relevant for repair of genomic uracil, was expressed in all cell lines and accounted for approximately half of total UNG in most cell lines (Fig. 4E). UNG enzymes are the most active of the glycosylases, at least *in vitro*. However, each glycosylase with its specific or complementary role may exert a significant impact on the total level of genomic uracil *in vivo*. We therefore quantified all the uracil-DNA glycosylases at protein level by MS. The relative abundance of quantified UNG protein (UNG1 and UNG2) (Fig. 4F) correlated strongly with total uracil excision activity (Fig. 4G), in accordance with its presumed major role in uracil repair. Similar to the uracil excision activity, UNG protein per cell also correlated inversely with genomic uracil level when all cell lines were included in the regression analysis (Table 2). Furthermore, quantified SMUG1 protein (Fig. 4H) correlated negatively with genomic uracil, although more weakly. Surprisingly, however, SMUG1 was the only glycosylase that correlated with genomic uracil when only the B-cell lymphoma group was analyzed (Table 2). In addition, the AID/SMUG1 protein ratio displayed significantly higher correlation with genomic uracil in the B-cell lymphoma group ($R^2 = 0.65$) compared to AID alone ($R^2 = 0.42$). No significant correlations were found for TDG or MBD4 proteins and genomic uracil (Fig. 4H) when analyzed separately (Table 2) or in combination with AID or other glycosylases.

3.5. Cell doubling time, genomic uracil content and repair capacity in cell cycle phases

In cells that do not express AID, one would predict that genomic uracil from misincorporation of dUMP during replication should result in increased genomic uracil in cells with short doubling time, as suggested previously [35]. Indeed, we observed a significant inverse relationship between genomic uracil and doubling time in non-lymphoma cancer cells ($R^2 = 0.57$; $P = 0.048$; Fig. 4I). Furthermore, since AID has been shown to act in the G_1 phase of the cell cycle [36–38], one would expect that the lymphoma cell lines with long doubling times might have higher genomic uracil levels than those with shorter doubling time. However, we did not find a significant positive correlation with doubling time ($R^2 = 0.27$; $P = 0.12$), although the curve was apparently different from that of the non-lymphoma cell lines (Fig. 4I).

As mentioned above, we found an inverse correlation between genomic uracil and both total uracil excision capacity, and with SMUG1 and UNG protein levels. Nuclear UNG2 expression peaks during G_1/S-phase transition and during S-phase and is expressed at a lower level in late S-phase, G_2 and early G_1 [13,39]. In contrast, TDG is mainly expressed in the G_1 phase of the cell cycle [13,39]. Thus, TDG might have a role in counteracting untargeted generation of U:G mismatches by AID in G_1, although correlation studies did not give indications of this. SMUG1 is not cell cycle regulated [40] and may contribute in all cell cycle phases, but is a rather slow acting enzyme [16]. To explore the relative contribution of the uracil-DNA glycosylases in *in vitro* complete BER of uracil in different parts of the cell cycle, we synchronized HeLa cells by double thymidine block [13], prepared nuclear extracts from the different cell cycle phases (monitored by flow cytometry) and applied an assay for complete BER of U:G mismatches in DNA [14,17,41]. To examine UNG, SMUG1 and TDG separately, we used a combination of neutralizing antibodies against UNG, SMUG1 and TDG. UNG was found to be by far the major contributor to initiate BER in the G1/S transition and in the S phase. Total repair capacity in G2 and G1 was somewhat lower than in the S phase, but UNG remained a major contributor to the initial step in BER-process, although contributing only 1.5–1.7 more than TDG. SMUG1 contributed in all cell cycle phases, but to a minor degree (Fig. 4J). Thus, a role of TDG and SMUG1 in BER of U:G mismatches in the G1 phase, and a smaller role in the S-phase would seem likely from our *in vitro* data. The contribution of the different uracil-DNA glycosylases during the cell cycle is likely to be similar in other human cell lines, including B cell lymphoma cell lines, although this has not been formally demonstrated.

3.6. Lymphomas and CLL carry a distinct AID-hotspot mutational signature in kataegis regions

Large scale genome sequencing of cancers has produced the novel observation that several cancers carry localized hypermutation, named *kataegis*, in small regions that are also associated with genomic rearrangements. The mutational signatures observed in most cancer types with *kataegis* (acute lymphoblastic leukemia (ALL), lung adenocarcinomas, breast, pancreas, and liver cancer) suggest an association with APOBEC3 enzymes, with a strong enrichment of C to T transitions and C to G transversions at TCA/T sequence contexts [11]. As mentioned, these *kataegis* patterns might be different from those found in lymphomas and CLL [11], though this was not explored in detail in their comprehensive paper. We therefore reanalyzed these exome sequencing data from *kataegis* regions of lymphomas and CLL and compared them to *kataegis* regions in cancers with typical APOBEC signatures (Fig. 5). The preferred sequence for C to T mutation in *kataegis* regions of B-cell lymphomas and CLL revealed a target sequence that overlap with the known AID hotspot motif (WRCY W = A/T, R = purine, Y = pyrimidine). The general mutational pattern for C to T transitions in lymphomas and CLL was AG**C**T, rather than T**C**A/T for the other cancer types with *kataegis* (Fig. 5). This strongly implicates AID-induced genomic uracil formation in the development of localized hypermutation in B-cell malignancies, in accordance

an *in vitro* assay for complete BER of a single uracil in a defined U:G context. HeLa cells were synchronized by double thymidine block, and harvested after 0, 3, 8, and 14 h representing G1/early S-phase, mid S-phase, G1 and G2 phase, and G1 phase, respectively, as shown by flow cytometric confirmation of cell cycle distribution in the top row. The contribution of each uracil DNA glycosylase was measured by using neutralizing antibodies to UNG, SMUG1, or TDG as indicated. Note that the column size values in the panels are directly comparable since they are generated from the same gel using the same substrate. The data points represent mean of independent triplicate experiments. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Regression analysis of genomic uracil levels (linear) *vs.* expression of uracil-DNA repair glycosylases (linear) normalized either to total protein or to total protein per cell. Bold red indicates significant negative correlation (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

Per total protein						
	All cell lines		B-cell lymphoma cell lines		Non-lymphoma cell lines	
	R^2	*P*-value	R^2	*P*-value	R^2	*P*-value
UNG	**0.24**	**0.05**	0.01	0.82	0.23	0.28
SMUG1	**0.28**	**0.03**	**0.41**	**0.04**	0.13	0.43
TDG	0.05	0.35	0.02	0.69	0.13	0.41
MBD4	0.07	0.27	0.02	0.7	0.05	0.63
Per cell						
	All cell lines		B-cell lymphoma cell lines		Non-lymphoma cell lines	
	R^2	*P*-value	R^2	*P*-value	R^2	*P*-value
UNG	**0.42**	**0.005**	0.05	0.52	0.20	0.31
SMUG1	**0.28**	**0.03**	0.16	0.24	0.06	0.6
TDG	0.22	0.06	0.14	0.27	0.32	0.17
MBD4	0.00	0.94	0.05	0.55	0.00	0.88

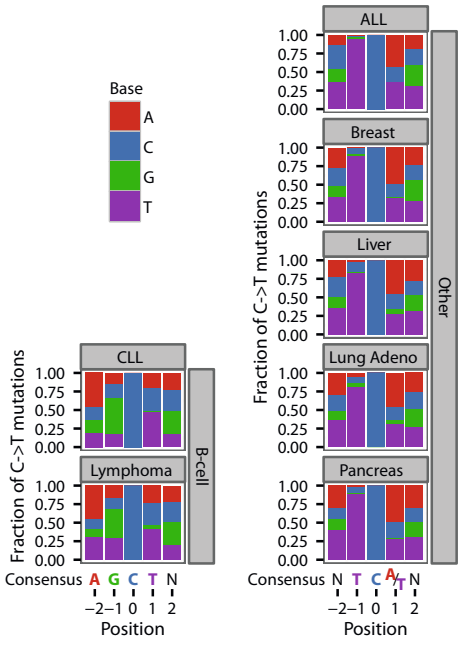


Fig. 5. Sequence context of C to T transitions in *kataegis* regions of lymphomas. Sequence analyses are based on exome sequencing data obtained from [11]. Sequence context of C to T transitions in *kataegis* regions of lymphomas ($n = 21$; 1102 single mutation sites) and CLL ($n = 15$; 290 single mutation sites) showing an AID-hotspot consensus sequence (—AG**C**TN—), where *N* represents no significant difference between A, T, C or G. Comparative analyses of cancers with known APOBEC signatures in *kataegis* regions showing an APOBEC consensus signature (—NT**C**ATN—), from ALL ($n = 1$; 153 single mutation sites), breast ($n = 67$; 5021 single mutation sites), liver ($n = 15$; 175 single mutation sites), lung adenocarcinoma ($n = 20$; 2024 single mutation sites), and pancreas ($n = 11$; 439 single mutation sites).

with our genomic uracil measurements and the published associations between AID and lymphomas [21–24,42–44] and CLL [45,46]. Moreover, these 122 lymphoma *kataegis* regions mapped to 70 distinct 100 kb blocks on 16 chromosomes, further supporting that enzymatic cytosine deamination by AID is not restricted to the S_μ region but occurs genome-wide.

4. Discussion

A major finding in our study is that AID expression is apparently a predominant source of genomic uracil in B-cell lymphoma cell lines. The LC–MS/MS method used quantifies genomic uracil as 2′-deoxyuridine in DNA [15]. The contribution of AID in this process was not solely made plausible by correlations, but also demonstrated by physiological induction of the endogenous AID gene, overexpression of recombinant AID, as well as knockdown of AID by shRNA. We feel that these results provide convincing evidence of dC to dU conversion *in vivo* by AID, which was considered missing in a recent review [47]. Furthermore; we found that mutational signatures in *kataegis* regions in human B-cell malignancies carry a distinct AID signature, strongly supporting the concept that AID is a DNA cytosine deaminase that, when mistargeted cause mutations and eventually B-cell malignancies. The increased genomic uracil is in general agreement with a recent report on relative increases in genomic uracil in B-cell lymphoma cell lines expressing AID, using an indirect genomic uracil-quantification method [48]. Evidence for targeted generation of uracil in Ig-genes has been obtained using a ligation-mediated PCR approach [20,49]. AID is normally only expressed in activated germinal center B-cells [2,50] and at low but detectable levels in early developing B-cells in the bone marrow [51]. This is apparently a risky process because AID strongly promotes the generation of germinal center-derived lymphomas [22,52,53], in which off-target activity of AID may contribute to point mutations and translocations during lymphomagenesis [31,54,55].

Recently, high-throughput sequencing of complete human cancer genomes and exomes revealed distinct mutational signatures compatible with mutagenesis by APOBEC-family enzymes in several common human cancers. This suggests that enzymatic off-target deamination of DNA-cytosine to uracil might be a major cause of mutation in human cancers [9–11]. However, direct evidence from measurements of uracil in the cancer genomes has largely been missing. Importantly, we found that endogenous AID-induction in CH12F3 mouse B-cells increases genomic uracil four-fold, from approximately 750 to 3000 uracils per genome already after 48 h. It is unlikely that this substantial increase can be confined to target regions in the Ig genes. Therefore, the increase in genomic uracil levels following endogenous AID expression indicates that even transiently induced AID expression during CSR causes widespread cytosine deamination. This is also in accordance with the mutational AID signatures found at many regions in human B-cell malignancy genomes. We did not observe correlation of genomic uracil with expression of other APOBEC-family

members. This does not rule out these as significant mutators in cancer cells, particularly since we only examined seven non-lymphoma cell lines. Low levels of enzymatic cytosine deamination may be overshadowed by dUMP misincorporation and spontaneous cytosine deamination. In addition, the strong effect of AID in B-cell lymphomas may obscure contribution of other APOBEC enzymes. A contribution from APOBECs may become significant over time and help drive transformation from normal cell to cancer cell, as indicated by mutational signatures [11,56].

Although AID expression levels correlated with variation in genomic uracil in the cells we tested, our results indicate that additional factors may modulate genomic uracil levels. The most obvious factor would be uracil repair capacity, which varies considerably between cell lines, and dUMP incorporation. We have previously shown that UNG is a rate-limiting factor in complete *in vitro* BER of genomic uracil [14] although UNG and SMUG1 may have complementary roles in uracil repair [16,29,57] and in the prevention of mutagenesis [58]. Studies on *UNG*$^{-/-}$ cells have documented an important function for UNG in keeping genomic uracil levels low [15]. However, the complete absence of any BER factor is a dramatic and rare event, whereas several-fold variation is rather common, at least in transformed cells. Earlier work demonstrated that AID-induced mutagenesis was counteracted by UNG, which initiates U:G DNA repair [31]. Our data showed that UNG and SMUG1 protein levels both correlated inversely with genomic uracil, with UNG showing the strongest correlation across all cells, while only SMUG1 correlates significantly in the lymphoma cell lines. Consequently, these results indicate that BER protein levels do affect genomic uracil. These results do not in themselves, however, necessarily reveal the relative importance of individual glycosylases for *in vivo* BER. We therefore made an effort to analyze the role of the glycosylases independently, using an assay for complete BER based on nuclear extracts from synchronized HeLa cells and a plasmid containing a single uracil. The results indicated that overall, UNG is the main contributor in initiating BER of uracil, at least in HeLa cells. However, SMUG1 and TDG may contribute significantly in G_1 (and G_2), which is also the time when AID is most active.

It is thought that U:G mismatches arising from AID in Ig genes and U:G from spontaneous deamination are processed by different mechanisms. Indeed, in order for SHM and CSR to be successfully carried out, canonical uracil DNA repair may be locally suppressed. One factor contributing to this may be transcription factor E2A, which induces AID [59,60], but represses both UNG-expression and its binding to relevant regions in the Ig genes [60]. Furthermore, p53 is actively reduced in germinal center B cells, presumably to allow mutagenic processing required for antibody maturation [61]. Although complex, the evidence that AID may drive carcinogenesis is well supported. In mice, AID expression was shown to be required for translocations between Ig loci and proto-oncogenes, a hallmark of several human B-cell lymphomas [62]. In contrast, AID knockout mice have fewer translocations [63] and accumulate fewer mutations in genes linked to B cell tumorigenesis [31]. AID expression has also been shown to confer a mutator phenotype in established lymphomas [42–44], but the role of AID in cancer progression remains unsettled [23,64,65]. Interestingly, AID expression has been reported in numerous cancers of non-B-cell origin, including breast, prostate, stomach, liver, and lung [66]. It would be interesting to investigate whether aberrant AID expression also confers high genomic uracil levels in these cancers. Interestingly, *Ung*$^{-/-}$ mice have roughly a 20-fold higher frequency of B-cell lymphoma compared with wild-type mice, but no apparent increase in other cancer types [67,68]. A straightforward explanation for this observation would be that SMUG1 and TDG together with MMR may compensate for UNG-deficiency in most tissues, but not in B-cells expressing AID, due to their increased genomic uracil levels.

A central role for AID-induced mutagenesis in lymphomas is also indicated by the AID-hotspot signature in the *kataegis* regions of a random selection of all lymphomas and CLLs (Fig. 5). We find that the *kataegis* AID-hotspot signature is not limited to lymphomas, but is also present in CLL, which overlaps with the category small lymphocytic lymphoma. Indeed, AID expression as cause of an ongoing mutator phenotype has been suggested for both lymphomas [42–44] and CLL [45,46]. Interestingly, progression of established cancers through expression of AID was also demonstrated in other blood cell cancers, such as ALL [69] and chronic myelogenous leukemia (CML), in which AID expression may lead to fatal lymphoblastoid crisis [70]. Thus, AID may be involved in development and progression of B-cell malignancies, and possibly only in late stage progression of other blood cell malignancies. This would be in agreement with the lack of an overall AID signature in ALL, as observed in our study.

In conclusion, we have provided strong evidence that AID is a DNA-cytosine deaminase that due to persistent expression causes accumulation of genomic uracil in B-cell lymphoma cell lines, as well as AID mutational signatures in human B-cell malignancies. Other factors, including expression levels for uracil-DNA glycosylases and cell doubling time, may modulate genomic uracil levels, but AID levels remain the strongest predictor.

Conflict of interest statement

There are no conflicts of interest.

Funding

This work was supported by The Norwegian Cancer Association (Grant id.: 576160), the Research Council of Norway (Grant id's: 191408; HEK) and 205316; PS), the Svanhild and Arne Must Fund for Medical Research, the Liaison Committee between the CentralNorway Regional Health Authority (RHA) and the Norwegian University of Science and Technology (NTNU) (Grant id.: 46047800; HSP) and Norwegian University of Science and Technology.

References

[1] B. Kavli, M. Otterlei, G. Slupphaug, H.E. Krokan, Uracil in DNA—general mutagen, but normal intermediate in acquired immunity, DNA Repair 6 (2007) 505–516.
[2] M. Muramatsu, V.S. Sankaranand, S. Anant, M. Sugai, K. Kinoshita, N.O. Davidson, T. Honjo, Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells, J. Biol. Chem. 274 (1999) 18470–18476.
[3] S.K. Petersen-Mahrt, R.S. Harris, M.S. Neuberger, AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification, Nature 418 (2002) 99–103.
[4] J. Di Noia, M.S. Neuberger, Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase, Nature 419 (2002) 43–48.
[5] K. Imai, G. Slupphaug, W.I. Lee, P. Revy, S. Nonoyama, N. Catalan, L. Yel, M. Forveille, B. Kavli, H.E. Krokan, H.D. Ochs, A. Fischer, A. Durandy, Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination, Nat. Immunol. 4 (2003) 1023–1028.
[6] C. Rada, G.T. Williams, H. Nilsen, D.E. Barnes, T. Lindahl, M.S. Neuberger, Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice, Curr. Biol.: CB 12 (2002) 1748–1755.
[7] S.G. Conticello, The AID/APOBEC family of nucleic acid mutators, Genome Biol. 9 (2008) 229.
[8] R.S. Harris, S.K. Petersen-Mahrt, M.S. Neuberger, RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators, Mol. Cell. 10 (2002) 1247–1253.
[9] J. Zhang, V. Grubor, C.L. Love, A. Banerjee, K.L. Richards, P.A. Mieczkowski, C. Dunphy, W. Choi, W.Y. Au, G. Srivastava, P.L. Lugar, D.A. Rizzieri, A.S. Lagoo, L. Bernal-Mizrachi, K.P. Mann, C. Flowers, K. Naresh, A. Evens, L.I. Gordon, M. Czader, J.I. Gill, E.D. Hsi, Q. Liu, A. Fan, K. Walsh, D. Jima, L.L. Smith, A.J. Johnson, J.C. Byrd, M.A. Luftig, T. Ni, J. Zhu, A. Chadburn, S. Levy, D. Dunson, S.S. Dave, Genetic heterogeneity of diffuse large B-cell lymphoma, Proc. Natl. Acad. Sci. U.S.A. 110 (2013) 1398–1403.
[10] C. Greenman, P. Stephens, R. Smith, G.L. Dalgliesh, C. Hunter, G. Bignell, H. Davies, J. Teague, A. Butler, C. Stevens, S. Edkins, S. O'Meara, I. Vastrik, E.E. Schmidt, T. Avis, S. Barthorpe, G. Bhamra, G. Buck, B. Choudhury, J. Clements, J.

Cole, E. Dicks, S. Forbes, K. Gray, K. Halliday, R. Harrison, K. Hills, J. Hinton, A. Jenkinson, D. Jones, A. Menzies, T. Mironenko, J. Perry, K. Raine, D. Richardson, R. Shepherd, A. Small, C. Tofts, J. Varian, T. Webb, S. West, S. Widaa, A. Yates, D.P. Cahill, D.N. Louis, P. Goldstraw, A.G. Nicholson, F. Brasseur, L. Looijenga, B.L. Weber, Y.E. Chiew, A. DeFazio, M.F. Greaves, A.R. Green, P. Campbell, E. Birney, D.F. Easton, G. Chenevix-Trench, M.H. Tan, S.K. Khoo, B.T. Teh, S.T. Yuen, S.Y. Leung, R. Wooster, P.A. Futreal, M.R. Stratton, Patterns of somatic mutation in human cancer genomes, Nature 446 (2007) 153–158.

[11] L.B. Alexandrov, S. Nik-Zainal, D.C. Wedge, S.A. Aparicio, S. Behjati, A.V. Biankin, G.R. Bignell, N. Bolli, A. Borg, A.L. Borresen-Dale, S. Boyault, B. Burkhardt, A.P. Butler, C. Caldas, H.R. Davies, C. Desmedt, R. Eils, J.E. Eyfjord, J.A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, N. Jager, D.T. Jones, D. Jones, S. Knappskog, M. Kool, S.R. Lakhani, C. Lopez-Otin, S. Martin, N.C. Munshi, H. Nakamura, P.A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J.V. Pearson, X.S. Puente, K. Raine, M. Ramakrishna, A.L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T.N. Schumacher, P.N. Span, J.W. Teague, Y. Totoki, A.N. Tutt, R. Valdes-Mas, M.M. van Buuren, L. van't Veer, A. Vincent-Salomon, N. Waddell, L.R. Yates, I. Australian Pancreatic Cancer Genome, I.B.C. Consortium, I.M.-S. Consortium, I. PedBrain, J. Zucman-Rossi, P.A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S.M. Grimmond, R. Siebert, E. Campo, T. Shibata, S.M. Pfister, P.J. Campbell, M.R. Stratton, Signatures of mutational processes in human cancer, Nature 500 (2013) 415–421.

[12] Y. Hu, I. Ericsson, K. Torseth, S.P. Methot, O. Sundheim, N.B. Liabakk, G. Slupphaug, J.M. Di Noia, H.E. Krokan, B. Kavli, A combined nuclear and nucleolar localization motif in activation-induced cytidine deaminase (AID) controls immunoglobulin class switching, J. Mol. Biol. 425 (2013) 424–443.

[13] L. Hagen, B. Kavli, M.M. Sousa, K. Torseth, N.B. Liabakk, O. Sundheim, J. Pena-Diaz, M. Otterlei, O. Horning, O.N. Jensen, H.E. Krokan, G. Slupphaug, Cell cycle-specific UNG2 phosphorylations regulate protein turnover, activity and association with RPA, EMBO J. 27 (2008) 51–61.

[14] T. Visnes, M. Akbari, L. Hagen, G. Slupphaug, H.E. Krokan, The rate of base excision repair of uracil is controlled by the initiating glycosylase, DNA Repair 7 (2008) 1869–1881.

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

[16] B. Kavli, O. Sundheim, M. Akbari, M. Otterlei, H. Nilsen, F. Skorpen, P.A. Aas, L. Hagen, H.E. Krokan, G. Slupphaug, 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 (2002) 39926–39936.

[17] M. Akbari, M. Otterlei, J. Pena-Diaz, P.A. Aas, B. Kavli, N.B. Liabakk, L. Hagen, K. Imai, A. Durandy, G. Slupphaug, H.E. Krokan, Repair of U/G and U/A in DNA by UNG2-associated repair complexes takes place predominantly by short-patch repair both in proliferating and growth-arrested cells, Nucleic Acids Res. 32 (2004) 5486–5498.

[18] T.S. Batth, J.D. Keasling, C.J. Petzold, Targeted proteomics for metabolic pathway optimization, Methods Mol. Biol. 944 (2012) 237–249.

[19] B. MacLean, D.M. Tomazela, N. Shulman, M. Chambers, G.L. Finney, B. Frewen, R. Kern, D.L. Tabb, D.C. Liebler, M.J. MacCoss, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments, Bioinformatics 26 (2010) 966–968.

[20] R.W. Maul, H. Saribasak, S.A. Martomo, R.L. McClure, W. Yang, A. Vaisman, H.S. Gramlich, D.G. Schatz, R. Woodgate, D.M. Wilson, P.J. 3rd, Gearhart, Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions, Nat. Immunol. 12 (2011) 70–76.

[21] J. Greeve, A. Philipsen, K. Krause, W. Klapper, K. Heidorn, B.E. Castle, J. Janda, K.B. Marcu, R. Parwaresch, Expression of activation-induced cytidine deaminase in human B-cell non-Hodgkin lymphomas, Blood 101 (2003) 3574–3580.

[22] L.A. Smit, R.J. Bende, J. Aten, J.E. Guikema, W.M. Aarts, C.J. van Noesel, Expression of activation-induced cytidine deaminase is confined to B-cell non-Hodgkin's lymphomas of germinal-center phenotype, Cancer Res. 63 (2003) 3894–3898.

[23] I.S. Lossos, R. Levy, A.A. Alizadeh, AID is expressed in germinal center B-cell-like and activated B-cell-like diffuse large-cell lymphomas and is not correlated with intraclonal heterogeneity, Leukemia 18 (2004) 1775–1779.

[24] L. Pasqualucci, R. Guglielmino, J. Houldsworth, J. Mohr, S. Aoufouchi, R. Polakiewicz, R.S. Chaganti, R. Dalla-Favera, Expression of the AID protein in normal and neoplastic B cells, Blood 104 (2004) 3318–3325.

[25] S. Gallien, E. Duriez, C. Crone, M. Kellmann, T. Moehring, B. Domon, Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer, Mol. Cell. Proteomics: MCP 11 (2012) 1709–1723.

[26] A.C. Peterson, J.D. Russell, D.J. Bailey, M.S. Westphall, J.J. Coon, Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics, Mol. Cell. Proteomics: MCP 11 (2012) 1475–1488.

[27] Y. Hu, I. Ericsson, B. Doseth, N.B. Liabakk, H.E. Krokan, B. Kavli, Activation-induced cytidine deaminase (AID) is localized to subnuclear domains enriched in splicing factors, Exp. Cell Res. 322 (2014) 178–192.

[28] H.E. Krokan, M. Bjoras, Base excision repair, Cold Spring Harbor Perspect. Biol. 5 (2013) a012583.

[29] H.S. Pettersen, O. Sundheim, K.M. Gilljam, G. Slupphaug, H.E. Krokan, B. Kavli, Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms, Nucleic Acids Res. 35 (2007) 3879–3892.

[30] A. Yamane, W. Resch, N. Kuo, S. Kuchen, Z. Li, H.W. Sun, D.F. Robbiani, K. McBride, M.C. Nussenzweig, R. Casellas, Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes, Nat. Immunol. 12 (2011) 62–69.

[31] M. Liu, J.L. Duke, D.J. Richter, C.G. Vinuesa, C.C. Goodnow, S.H. Kleinstein, D.G. Schatz, Two levels of protection for the B cell genome during somatic hypermutation, Nature 451 (2008) 841–845.

[32] M.G. Hasham, N.M. Donghia, E. Coffey, J. Maynard, K.J. Snow, J. Ames, R.Y. Wilpan, Y. He, B.L. King, K.D. Mills, Widespread genomic breaks generated by activation-induced cytidine deaminase are prevented by homologous recombination, Nat. Immunol. 11 (2010) 820–826.

[33] H. Nilsen, M. Otterlei, T. Haug, K. Solum, T.A. Nagelhus, F. Skorpen, H.E. Krokan, Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene, Nucleic Acids Res. 25 (1997) 750–755.

[34] H. Nilsen, K.S. Steinsbekk, M. Otterlei, G. Slupphaug, P.A. Aas, H.E. Krokan, Analysis of uracil-DNA glycosylases from the murine Ung gene reveals differential expression in tissues and in embryonic development and a subcellular sorting pattern that differs from the human homologues, Nucleic Acids Res. 28 (2000) 2277–2285.

[35] S. Andersen, T. Heine, R. Sneve, I. Konig, H.E. Krokan, B. Epe, H. Nilsen, 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 (2005) 547–555.

[36] E.C. Ordinario, M. Yabuki, R.P. Larson, N. Maizels, Temporal regulation of Ig gene diversification revealed by single-cell imaging, J. Immunol. 183 (2009) 4545–4553.

[37] G. Sharbeen, C.W. Yee, A.L. Smith, C.J. Jolly, Ectopic restriction of DNA repair reveals that UNG2 excises AID-induced uracils predominantly or exclusively during G1 phase, J. Exp. Med. 209 (2012) 965–974.

[38] C.E. Schrader, J.E. Guikema, E.K. Linehan, E. Selsing, J. Stavnezer, Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair, J. Immunol. 179 (2007) 6064–6071.

[39] U. Hardeland, C. Kunz, F. Focke, M. Szadkowski, P. Schar, Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2, Nucleic Acids Res. 35 (2007) 3859–3867.

[40] J. Pena-Diaz, S.A. Hegre, E. Anderssen, P.A. Aas, R. Mjelle, G.D. Gilfillan, R. Lyle, F. Drablos, H.E. Krokan, P. Saetrom, Transcription profiling during the cell cycle shows that a subset of Polycomb-targeted genes is upregulated during DNA replication, Nucleic Acids Res. 41 (2013) 2846–2856.

[41] M. Akbari, J. Pena-Diaz, S. Andersen, N.B. Liabakk, M. Otterlei, H.E. Krokan, Extracts of proliferating and non-proliferating human cells display different base excision pathways and repair fidelity, DNA Repair 8 (2009) 834–843.

[42] M.S. Hardianti, E. Tatsumi, M. Syampurnawati, K. Furuta, K. Saigo, Y. Nakamachi, S. Kumagai, H. Ohno, S. Tanabe, M. Uchida, N. Yasuda, Activation-induced cytidine deaminase expression in follicular lymphoma: association between AID expression and ongoing mutation in FL, Leukemia 18 (2004) 826–831.

[43] C. Bodor, A. Bognar, L. Reiniger, A. Szepesi, E. Toth, L. Kopper, A. Matolcsy, Aberrant somatic hypermutation and expression of activation-induced cytidine deaminase mRNA in mediastinal large B-cell lymphoma, Br. J. Haematol. 129 (2005) 373–376.

[44] A.J. Deutsch, A. Aigelsreiter, P.B. Staber, A. Beham, W. Linkesch, C. Guelly, R.I. Brezinschek, M. Fruhwirth, W. Emberger, M. Buettner, C. Beham-Schmid, P. Neumeister, MALT lymphoma and extranodal diffuse large B-cell lymphoma are targeted by aberrant somatic hypermutation, Blood 109 (2007) 3500–3504.

[45] H. McCarthy, W.G. Wierda, L.L. Barron, C.C. Cromwell, J. Wang, K.R. Coombes, R. Rangel, K.S. Elenitoba-Johnson, M.J. Keating, L.V. Abruzzo, High expression of activation-induced cytidine deaminase (AID) and splice variants is a distinctive feature of poor-prognosis chronic lymphocytic leukemia, Blood 101 (2003) 4903–4908.

[46] F. Palacios, P. Moreno, P. Morande, C. Abreu, A. Correa, V. Porro, A.I. Landoni, R. Gabus, M. Giordano, G. Dighiero, O. Pritsch, P. Oppezzo, High expression of AID and active class switch recombination might account for a more aggressive disease in unmutated CLL patients: link with an activated microenvironment in CLL disease, Blood 115 (2010) 4488–4496.

[47] A.S. Yousif, A. Stanlie, N.A. Begum, T. Honjo, Opinion: uracil DNA glycosylase (UNG) plays distinct and non-canonical roles in somatic hypermutation and class switch recombination, Int. Immunol. 26 (2014) 575–578.

[48] S. Shalhout, D. Haddad, A. Sosin, T.C. Holland, A. Al-Katib, A. Martin, A.S. Bhagwat, Genomic uracil homeostasis during normal B cell maturation and loss of this balance during B cell cancer development, Mol. Cell. Biol. 34 (2014) 4019–4032.

[49] B. Roche, A. Claes, F. Rougeon, Deoxyuridine triphosphate incorporation during somatic hypermutation of mouse VkOx genes after immunization with phenyloxazolone, J. Immunol. 185 (2010) 4777–4782.

[50] E.E. Crouch, Z. Li, M. Takizawa, S. Fichtner-Feigl, P. Gourzi, C. Montano, L. Feigenbaum, P. Wilson, S. Janz, F.N. Papavasiliou, R. Casellas, Regulation of AID expression in the immune response, J. Exp. Med. 204 (2007) 1145–1156.

[51] J.H. Han, S. Akira, K. Calame, B. Beutler, E. Selsing, T. Imanishi-Kari, Class switch recombination and somatic hypermutation in early mouse B cells are mediated by B cell and Toll-like receptors, Immunity 27 (2007) 64–75.

[52] L. Pasqualucci, G. Bhagat, M. Jankovic, M. Compagno, P. Smith, M. Muramatsu, T. Honjo, H.C. Morse, M.C. Nussenzweig III, R. Dalla-Favera, AID is required for germinal center-derived lymphomagenesis, Nat. Genet. 40 (2008) 108–112.

[53] A. Kotani, N. Kakazu, T. Tsuruyama, I.M. Okazaki, M. Muramatsu, K. Kinoshita, H. Nagaoka, D. Yabe, T. Honjo, Activation-induced cytidine deaminase (AID)

promotes B cell lymphomagenesis in Emu-cmyc transgenic mice, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 1616–1620.
[54] O. Hakim, W. Resch, A. Yamane, I. Klein, K.R. Kieffer-Kwon, M. Jankovic, T. Oliveira, A. Bothmer, T.C. Voss, C. Ansarah-Sobrinho, E. Mathe, G. Liang, J. Cobell, H. Nakahashi, D.F. Robbiani, A. Nussenzweig, G.L. Hager, M.C. Nussenzweig, R. Casellas, DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes, Nature 484 (2012) 69–74.
[55] I.A. Klein, W. Resch, M. Jankovic, T. Oliveira, A. Yamane, H. Nakahashi, M. Di Virgilio, A. Bothmer, A. Nussenzweig, D.F. Robbiani, R. Casellas, M.C. Nussenzweig, Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes, Cell 147 (2011) 95–106.
[56] M.B. Burns, L. Lackey, M.A. Carpenter, A. Rathore, A.M. Land, B. Leonard, E.W. Refsland, D. Kotandeniya, N. Tretyakova, J.B. Nikas, D. Yee, N.A. Temiz, D.E. Donohue, R.M. McDougle, W.L. Brown, E.K. Law, R.S. Harris, APOBEC3B is an enzymatic source of mutation in breast cancer, Nature 494 (2013) 366–370.
[57] H. Nilsen, K.A. Haushalter, P. Robins, D.E. Barnes, G.L. Verdine, T. Lindahl, Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase, EMBO J. 20 (2001) 4278–4286.
[58] Q. An, P. Robins, T. Lindahl, D.E. Barnes, C –> T mutagenesis and gamma-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases, EMBO J. 24 (2005) 2205–2213.
[59] J. Hauser, N. Sveshnikova, A. Wallenius, S. Baradaran, J. Saarikettu, T. Grundstrom, B-cell receptor activation inhibits AID expression through calmodulin inhibition of E-proteins, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 1267–1272.
[60] A. Wallenius, J. Hauser, P.A. Aas, A. Sarno, B. Kavli, H.E. Krokan, T. Grundstrom, Expression and recruitment of uracil-DNA glycosylase are regulated by E2A during antibody diversification, Mol. Immunol. 60 (2014) 23–31.
[61] R.T. Phan, R. Dalla-Favera, The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells, Nature 432 (2004) 635–639.
[62] R. Kuppers, R. Dalla-Favera, Mechanisms of chromosomal translocations in B cell lymphomas, Oncogene 20 (2001) 5580–5594.
[63] Y. Dorsett, D.F. Robbiani, M. Jankovic, B. Reina-San-Martin, T.R. Eisenreich, M.C. Nussenzweig, A role for AID in chromosome translocations between c-myc and the IgH variable region, J. Exp. Med. 204 (2007) 2225–2232.
[64] M. Leuenberger, S. Frigerio, P.J. Wild, F. Noetzli, D. Korol, D.R. Zimmermann, C. Gengler, N.M. Probst-Hensch, H. Moch, M. Tinguely, AID protein expression in chronic lymphocytic leukemia/small lymphocytic lymphoma is associated with poor prognosis and complex genetic alterations, Mod. Pathol. 23 (2010) 177–186 (an official journal of the United States and Canadian Academy of Pathology, Inc).
[65] K. Willenbrock, C. Renne, M. Rottenkolber, W. Klapper, M. Dreyling, M. Engelhard, R. Kuppers, M.L. Hansmann, B. Jungnickel, The expression of activation induced cytidine deaminase in follicular lymphoma is independent of prognosis and stage, Histopathology 54 (2009) 509–512.
[66] A. Orthwein, J.M. Di Noia, Activation induced deaminase: how much and where? Semin. Immunol. 24 (2012) 246–254.
[67] S. Andersen, M. Ericsson, H.Y. Dai, J. Pena-Diaz, G. Slupphaug, H. Nilsen, H. Aarset, H.E. Krokan, Monoclonal B-cell hyperplasia and leukocyte imbalance precede development of B-cell malignancies in uracil-DNA glycosylase deficient mice, DNA Repair 4 (2005) 1432–1441.
[68] H. Nilsen, G. Stamp, S. Andersen, G. Hrivnak, H.E. Krokan, T. Lindahl, D.E. Barnes, Gene-targeted mice lacking the Ung uracil-DNA glycosylase develop B-cell lymphomas, Oncogene 22 (2003) 5381–5386.
[69] T.A. Gruber, M.S. Chang, R. Sposto, M. Muschen, Activation-induced cytidine deaminase accelerates clonal evolution in BCR-ABL1-driven B-cell lineage acute lymphoblastic leukemia, Cancer Res. 70 (2010) 7411–7420.
[70] L. Klemm, C. Duy, I. Iacobucci, S. Kuchen, G. von Levetzow, N. Feldhahn, N. Henke, Z. Li, T.K. Hoffmann, Y.M. Kim, W.K. Hofmann, H. Jumaa, J. Groffen, N. Heisterkamp, G. Martinelli, M.R. Lieber, R. Casellas, M. Muschen, The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia, Cancer Cell 16 (2009) 232–245.