Anastasia Galashevskaya

Multiple Functions of Base Excision Repair Proteins: Relevance to RNA Quality Control and B-Cell Lymphomagenesis

Thesis for the degree of Philosophiae Doctor

Trondheim, November 2014

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

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

ISBN 978-82-326-0562-0 (printed ver.)
ISBN 978-82-326-0563-7 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2014:325

Printed by NTNU-trykk

NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET

DET MEDISINSKE FAKULTET

DNA i levende celler gjennomgår kontinuerlig strukturelle og kjemiske forandringer, som kan forårsake feil i replikasjon og transkripsjon, og føre til akkumulering av mutasjoner. Mutasjoner er viktige for genetisk variasjon og evolusjon, men langtids overlevelse av en art krever genetisk stabilitet. Genetisk informasjon er relativt stabil fordi et komplekst maskineri av mekanismer for reparasjon, skade-toleranse, og sjekkpunkt mekanismer opprettholder DNA integritet. Defekter i disse systemene påvirker organismens utvikling, aldringsprosessen, og kan føre til kreft.

Dette arbeidet har bidratt til bedre forståelse av genomisk uracil prosessering. De kjente kildene til uracil i DNA ble tidligere antatt å være misinkorporering av dUMP under DNA replikasjonen og spontan deaminering av cytosin. Oppdagelsen av aktiveringsindusert cytidin deaminase (AID) og andre APOBEC enzymer som kan generere uracil fra cytosin har økt interessen for dette temaet. AID-indusert DNA cytosin-deaminering i immunglobulin gener og APOBEC-indusert deaminering av viralt DNA er naturlige prosesser i immunsystemet. Til tross for sine viktige fysiologiske funksjoner, medfører disse forsvarsmekanismene i verten en høy risiko for det oppstår potensielt kreftfremkallende-mutasjoner. Høy gjennomstrømning sekvensering av et stort antall menneskelige kreft genomer viser at mutasjoner i C:G basepar, og C til T overganger i særdeleshet, er de mest utbredte. Dette tyder at enzymatisk cytosin deaminering kan være en hovedårsak.

Betydelig fremgang har blitt gjort vedrørende opprinnelsen og base-eksisjonsreparasjon av genomisk uracil, men det viktige spørsmålet om det basale nivået er fortsatt omstridt. Tallene varierer med nesten tre størrelsesordener. Vi mener at dette avviket mest sannsynlig er et resultat av metodiske feil relatert til preparasjon av DNA prøver og generering av uracil fra cytosin i løpet av analysen. I artikkel I diskuterer vi disse potensielle problemene og presenterer en forbedret LC/MS/MS-basert metode for absolutt kvantifisering av genomisk deoxyuridin (dUrd). For å teste metoden i en relevant biologisk sammenheng, har vi undersøkt murine embryonale fibroblaster og humane lymfoblastoide cellelinjer – der UNG-genet er inaktivert. De reparasjon-kompetente cellelinjene hadde omtrent 400-600 dUrd per genom. Dette tallet tyder på at det basale nivået av genomisk uracil har vært overestimert tidligere. Vi mener at vår nye metode gir det mest nøyaktige resultatet og er svært relevant for DNA-reparasjon-orienterte forskere.

I artikkel II undersøkte vi ulike genomisk uracil og uracil fjerning fra DNA i humane kreftcellelinjer. Vi fant svært høye nivåer av genomisk uracil i mange B-cell lymfomer. Nivået korrelerte med AID uttrykk og var delvis reversibelt etter AID knockdown. Uracil reparasjonskapasitet og uttrykk av uracil DNA-reparasjonsproteiner korrelerte inverst med genomisk uracil nivå. Vi fant også at regioner med grupperte mutasjoner (kataegis) i lymfom og kronisk lymfatisk leukemi har AID-hotspot mutasjons signatur. Dette er det første mekanistiske beviset på at AID kan forårsake akkumulering av genomisk uracil i kreftceller og indikerer at enzymet kan ha en viktig rolle i utvikling av B-cell lymfomer.

I tillegg til arbeidet med hovedprosjektet, studerte vi interaksjons partnere til uracil-DNA glykosylase SMUG1. I artikkel III, viser vi at SMUG1 kan samhandle med pseudouridine synthase DKC1, og at SMUG1 og DKC1 ko-lokaliserer i nucleoli og Cajal legemer, hvor rRNA biogenes og ikke-kodende RNA modning foregår, henholdsvis. Vi oppdaget at SMUG1-defekt var assosiert med redusert nivå av modne 28S, 18S, og 5.8S rRNA og økt nivå av polyadenylerte 28S rRNA. Videre observerte vi at SMUG1 har aktivitet på ssRNA som inneholder 5-hm (dUrd) *in vitro*. Vi fant også et økt 5-hm(Urd) nivå i modent rRNA i SMUG1-defekte celler. Disse funnene danner grunnlag for en ny hypotese om at SMUG1 kan være involvert i rRNA kvalitetskontroll. Det har blitt påvist at et annet DNA-reparasjon protein, apyrmidinic/apurinic endonuclease (APE1), også er assosiert med 47S, 28S, 18S rRNA, og nucleoplasmin innen nucleoli. Vi spekulerer derfor om APE1 behandler AP-seter som genereres av SMUG1. I dette tilfellet ville APE1 danne 3'-OH termini, som målrettes for videre nedbrytning i exosomer.

Funksjoner av DNA-reparasjonsproteiner er mer allsidige enn man har trodd tidligere. Det synes som en logisk mulighet at DNA glycosylaser sin evne til å gjenkjenne kjemiske modifikasjoner også kan bidra til identifisering av skadede eller feilaktig modifisert RNA og at slike proteiner i noen tilfeller kan bidre til genetisk stabilitet ikke bare gjennom DNA-reparasjon.

Kandidat: Anastasia Galashevskaya

Veiledere: Prof. Dr. Med. Hans E. Krokan og Prof. Geir Slupphaug

Institutt: Institutt for Kreftforskning og Molekylærmedisin

Finansieringskilde: Norges Forskningsråd og Svanhild og Arne Musts Fond for Medisinsk Forskning

Contents

Acknowledgements

The work presented in this thesis was carried out at the Department for Cancer Research and Molecular Medicine, Norwegian University of Science and Technology in the period from 2011 to 2014. I am grateful for the financial support provided by the Research Council of Norway and the Svanhild and Arne Must Fund for Medical Research.

First of all, I would like to express my gratitude to Professor Hans E. Krokan for introducing me to the field of DNA repair and creating such an inspiring place to work. I highly appreciate your enormous knowledge and enthusiasm in science and feel honoured for being a member of your group.

This work would not be possible without contributions from the co-authors. I especially acknowledge my friend and colleague Antonio Sarno, without whom I would never have succeeded. Thank you for scientific and non-scientific discussions, proofreading of my thesis and never ending jokes. It was a pleasure to share office and lab space with you. I gladly thank Cathrine B. Vågbø for guiding me through years of LC-MS/MS analysis. Your tremendous experience is invaluable. I appreciate your positive feedback, magic troubleshooting and all the time you have spent answering my many questions. I am grateful to Per Arne Aas for showing so much interest in my work and life in general. Your clever advices and encouraging attitude have always helped me.

I thank all my colleagues for creating a superb social and scientific atmosphere. I would like to specially mention Audun Hanssen-Bauer, Ida Ericsson, Rebekka Muller, Kamila Anna Zub, Mirta Mittelstedt Leal de Sousa, Yi Hu, Camilla Olaisen, Karin Gilljam, Karin Solvang-Garten, Siri Bachke, Lars Hagen, Berit Doseth and Henrik P. Sahlin Pettersen.

Finally, I would like to thank my friends and family for their love and support.

Trondheim, Norway Anastasia Galashevskaya

September, 2014

List of papers

Paper I

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

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

DNA repair (Amst.), 12, 699-706 (2013)

* These authors contributed equally to this work

Paper II

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

Henrik S. Pettersen, **Anastasia Galashevskaya**, Berit Doseth, Mirta M. L. Sousa, Antonio Sarno, Torkild Visnes, Per A. Aas, Nina B. Liabakk, Geir Slupphaug, Pål Sætrom, Bodil Kavli, and Hans E. Krokan

Manuscript submitted September 2014 (DNA repair)

Paper III

The human base excision repair enzyme SMUG1 directly interacts with DKC1 and contributes to RNA quality control.

Laure Jobert, Hanne K. Skjeldam, Bjørn Dalhus, **Anastasia Galashevskaya**, Cathrine B. Vågbø, Magnar Bjørås, and Hilde Nilsen

Molecular Cell 49, 339-345 (2013)

Abbreviations

3-meA	3-methyladenine
5-caC	5-carboxylcytosine
5-fC	5-formylcytosine
5-FU	5-fluorouracil
5-hmC	5-hydroxymethylcytosine
5-hmU	5-hydroxymethyluracil
5-hU	5-hydroxyuracil
5-meC	5-methylcytosine
7-meG	N7-methylguanine
8-oxoG	8-oxo-7,8-dihydroguanine
A	adenine
ACF	APOBEC1 complementation factor
AID	activation-induced cytidine deaminase
alt-NHEJ	alternative non-homologous end joining
AMP	adenosine monophosphate
AP	apurinic, apyrimidinic or abasic
APE1/2	AP endonuclease 1/2
APIM	AlkB homologue 2 PCNA-interacting motif
ApoB	apolipoprotein B
APOBEC	apolipoprotein B mRNA editing catalytic polypeptide
ATM	ataxia telangiectasia mutated kinase
ATP	adenosine triphosphate
ATR	ataxia telangiectasia and Rad3-related kinase
B cell	B lymphocyte
BAX	BCL2-associated X protein
BER	base excision repair
BRCA1/2	breast and ovarian cancer type 1/2 susceptibility protein
C	cytosine
^{13}C-dCyd	^{13}C-deoxycytidine
C region	immunoglobulin constant region
Cdc25A/B/C	cell division cycle 25 homologs A/B/C
Cdk1/2	cyclin-dependent kinases 1/2
cDNA	complementary DNA
CE-LIF	capillary electrophoresis and laser-induced fluorescence
Chk1/2	checkpoint kinases 1/2
CpG	cytosine-phosphodiester bond-guanine
CSR	class switch recombination
C-terminal	carboxyl terminal
CTNNBL1	catenin-β-like 1
CV	coefficient of variation
dCMP	deoxycytidine monophosphate
dCyd	deoxycytidine
DDT	DNA damage tolerance

DKC1	dyskerin
D-loop	displacement loop
DNA-PK	DNA-dependent protein kinase
Dnmt3a/3b	DNA methyltransferases 3a/3b
dNTP	deoxynucleotide triphosphate
dRP	deoxyribosephosphate
DSB	double-strand break
dUMP	deoxyuridine monophosphate
dUrd	deoxyuridine
E. coli	*Escherichia coli*
εC	3'N4-ethenocytosine
eEF1A	translation elongation factor 1α
ELISA	enzyme-linked immunosorbent assay
EndoIII	endonuclease III
Erg-1	early growth response protein 1
ESI	electrospray ionization
EXO1	exonuclease 1
FEN1	flap endonuclease 1
FTO	fat mass and obesity-associated protein
G	guanine
GC	gas chromatography
GGR	global genomic repair
$G_{1/2}$ phase	"growth" phase 1/2 (prior to S phase (G_1) and M phase (G_2))
H	heavy (chain)
HDAC1	high-mobility group protein B1
HIV-1	human immunodeficiency virus type 1
HMGB1	histone deacetylase 1
HPLC	High-performance liquid chromatography
HR	homologous recombination
ICLs	interstrand cross-links
Ig	immunoglobulin
kb	kilo bases
kDa	kilo Dalton
Ku70/Ku80	Ku heterodimer
L	light (chain)
LC	liquid chromatography
LDLs	low density lipoproteins
LIG1/3α	DNA ligase I/IIIα
LPO	lipid peroxidation
Lys	lysine
m^6A	N^6-methyladenosine
MBD4	methyl-CpG-binding domain 4
MCM7	minichromosome maintenance complex component 7
MEF	mouse embryonic fibroblast
miRs	microRNAs

MLH1	human homolog 1 of *E. coli* MutL
MMR	mismatch repair
MPG	N-methylpurine DNA glycosylase
M phase	mitotic phase
mRNA	messenger RNA
MRN	Mre11-Rad50-Nbs1 complex
MS	mass spectrometry
MS^n	sequential MS
MSH2/3/6	mutS homolog 2/3/6
MutLα	MLH1-PMS2 heterodimer
MutSα	MSH2-MSH6 heterodimer
MutSβ	MSH2-MSH3 heterodimer
MutM/Fpg	formamidopyrimidine DNA glycosylase
MYH	mutY homolog DNA glycosylase
m/z	mass-to-charge ratio
NAD^+	nicotinamide adenine dinucleotide
nCaRE	negative calcium response elements
NEIL	Nei endonuclease VIII-like
NES	nuclear export signal
NER	nucleotide excision repair
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B-cells
NHEJ	non-homologous end joining
NLS	Nuclear localization signal
NO	nitric oxide
NPM1	nucleoplasmin
N-terminal	amino terminal
O_2^-	superoxide anion
$^1O_2$	singlet oxygen
OH	hydroxy
OH	hydroxyl radical
O^6-meG	O^6-methylguanine
O^6-meG:C	O^6-methylguanine base paired with cytosine
O^6-meG:T	O^6-methylguanine base paired with thymine
OGG1	8-oxoguanine DNA glycosylase
PAR	poly(ADP)ribose
p53	transcription factor 53
PARG	poly(ADP)ribose-glycohydrolase
PARP1/2/3	poly(ADP-ribosyl) polymerase 1/2/3
PCNA	proliferating cell nuclear antigen
PIKK	phosphatidylinositol 3-kinase-like protein kinase
PIP	PCNA-interacting peptide
PMS2	postmeiotic segregation increased 2 protein
PNKP	polynucleotide phosphatase/kinase
Polα/β/ε/δ/λ	DNA polymerase α/β/ε/δ/λ (alpha/beta/epsilon/delta/lambda)
Polη/ι/κ/ζ/θ/ν	DNA polymerase η/ι/κ/ζ/θ/ν (eta/iota/kappa/zeta/theta/nu)

PUMA	p53 upregulated modulator of apoptosis
Q	quadruple
REV1	protein encoded by the *REV1* gene
RFC	replication factor C
RIAs	radioimmunoassays
RPA	replication protein A
RSSs	recombination signal sequences
rRNA	ribosomal RNA
S region	switch region
SFS	synchronous fluorescence spectrophotometry
siRNA	small interfering RNA
SMUG1	single-strand selective monofunctional uracil DNA glycosylase
$S_N1/2$	nucleophilic substitution mechanism 1/2
S phase	(DNA) synthesis phase
SSBR	single-strand break repair
SSBs	single-strand breaks
ssDNA	single stranded DNA
SUMO	small ubiquitine-like modifier
T	thymine
TCR	transcription-coupled repair
TDG	thymine-DNA glycosylase
TET	ten-eleven translocation
TFIIH	transcription factor IIH
T:G	thymine base mispaired with guanine
TLC	thin-layer chromatography
TLS	translesion synthesis
TTF1	thyroid transcription factor 1
U	uracil
U:A	uracil base paired with adenine
U:G	uracil base mispaired with guanine
UNG1/2	uracil-N-DNA-glycosylase protein isotype 1/2 (encoded by the *UNG* gene)
UTRs	untranslated regions
Vif	viral infectivity factor
VLDL	very low density lipoprotein
Vpr	viral protein R
XLF	Xrcc4 like factor
XRCC1/2	X-ray repair cross-complementing protein 1/2

1. Introduction

1.1 DNA damage

DNA in living cells undergoes continuous structural and chemical alterations through different types of damage, which may result in errors in replication and transcription and lead to mutation accumulation. Mutations are important for genetic variation and evolution, but survival of the individual demands genetic stability. Genetic information is stored relatively stably only because a complex machinery of repair, damage tolerance, and checkpoint pathways maintain DNA integrity. Defects in these systems affect the organism's development, aging process and predispose to cancer (reviewed in (Friedberg et al., 2006)).

Oxidation

Oxidative DNA damage can be induced by reactive oxygen species and reactive nitrogen species. Superoxide ($\cdot O_2^-$), hydroxyl radicals ($\cdot OH$), singlet oxygen ($^1O_2$), nitric oxide ($\cdot NO$) and other agents that are formed during normal cellular signaling and metabolism can result from inflammation, exposure to oxidizing chemicals, ionizing radiation, and UV light. These agents oxidize DNA bases and generate abasic sites and single- and double-strand breaks (reviewed in (Evans et al., 2004)).

The most extensively studied oxidized base is 8-oxo-7,8-dihydroguanine (8-oxoG). It originates from the attack of hydroxyl radicals on the C8 guanine. This lesion is mutagenic because it may result in GC→TA transversions during replication (reviewed in (De Bont and van Larebeke, 2004)). The steady-state level of 8-oxoG in human cells is approximately 0.3-4.2 per 10^6 guanine residues (Collins et al., 2004). Increased levels of genomic 8-oxoGua have been detected in breast cancer tissue, lung squamous cell carcinoma, colorectal adenoma, male prostate cancer, and many other human malignancies (reviewed in (Kryston et al., 2011)). In addition, 8-oxoG accumulation in nuclear and mitochondrial DNA has been linked to aging (Izzotti et al., 1999; Lu et al., 2004; Nie et al., 2013; Stevnsner et al., 2002).

DNA interaction with the hydroxyl radicals can lead to single-strand breaks (SSBs) via hydrogen abstraction from sugar moieties. SSBs can be converted to double-strand DNA breaks (DSBs) when encountered by replisomes or as a result of adjacent SSBs (reviewed in (Caldecott, 2008)). These lesions frequently induce chromosomal aberrations, including aneuploidy, deletions, and chromosomal translocations. Defective repair of DNA strand breaks is associated with congenital brain development abnormalities, immunodeficiency,

predisposition to breast, ovarian, and prostate cancer (reviewed in (Katyal and McKinnon, 2008)).

Alkylation

Alkylation damage refers to the addition of an alkyl group to DNA bases. Alkylating agents can be classified according to the number of reactive sites (monofunctional versus bifunctional), the chemical mechanism (S_N1 versus S_N2 nucleophilic substitution), and the length of alkyl chain (methyl, ethyl, etc.). The induced DNA lesions range from base modifications and strand breaks to DNA-DNA and DNA-protein cross-links (reviewed in (Fu et al., 2012)).

Methylation is the predominant type of alkylation. DNA can be methylated by endogenous agents such as S-adenosylmethionine, betaine, choline, and exogenous factors such as tobacco-specific nitrosamines and atmospheric halocarbons. Most of them generate N7-methylguanine (7-meG), 3-methyladenine (3-meA) and O^6-methylguanine (O^6-meG). 7-meG is considered to be relatively harmless because it base pairs similarly to unmodified guanine; however, it may create cytotoxic abasic sites upon excision by *N*-alkylpurine DNA-glycosylases. 3-meA is considered cytotoxic: it can cause S phase arrest, sister chromatid exchange, chromosome aberrations, and apoptosis (reviewed in (Shrivastav et al., 2010)). O^6-meG is mutagenic because it results in GC→AT transitions. Moreover, recognition of O^6-meG:T and O^6-meG:C mispairs by the mismatch repair system may lead to generation of double-strand breaks (reviewed in (Drablos et al., 2004)).

Bifunctional alkylating agents such as nitrogen mustards and aziridine compounds have two reactive groups and can form interstrand cross-links (ICLs). Cross-links may also occur on the same DNA strand or between modified DNA base and protein. This type of damage disrupts replication, transcription, and recombination (reviewed in (Kondo et al., 2010)).

Hydrolysis

Spontaneous hydrolytic cleavage of the ***N***-glycosidic bond between the base and deoxyribose in DNA results in approximately 10 000 apurinic/apyrimidinic sites per cell per day (Lindahl and Nyberg, 1972). Abasic sites stall DNA replication and transcription, induce base substitutions, and cause frameshift mutations. Furthermore, abasic sites can be converted into SSBs by AP endonucleases or DNA glycosylases/AP lyases (Loeb and Preston, 1986). Another type of hydrolytic DNA damage is deamination. Deamination of the exocyclic amino group in cytosine (C), 5-methylcytosine (5-meC), adenine (A), and guanine (G) generates

uracil (U), thymine (T), hypoxanthine, and xanthine, respectively. Subsequent DNA replication of unrepaired U:G and T:G mismatches result in GC→AT transitions, the most frequent type of mutation in human cancers. Hypoxanthine preferentially base pairs with cytosine and may lead to AT→GC transitions. Xanthine does not stably pair with any bases and may arrest DNA synthesis (reviewed in (Friedberg et al., 2006)).

Damage to DNA by other chemical species

Estrogen metabolites can damage DNA directly through the formation of bulky adducts or indirectly through redox cycling processes that generate radical species. These free radicals induce single-strand breaks and 8-hydroxylation of guanine bases and produce malondialdehyde-DNA adducts. Quinone intermediates of 4-hydroxylated estrogens have been shown to create unstable DNA adducts, which are prone to decomposition and formation of abasic sites. It has been demonstrated that estrogenic steroids raise the risk of breast, endometrial, and uterine cancers in humans (reviewed in (Liehr, 2001)).

The major aldehyde products of lipid peroxidation (LPO), malondialdehyde, 4-hydroxynonenal, acrolein, and crotonaldehyde form exocyclic DNA adducts that can compromise base pairing or produce substitution mutations. Several studies have shown that inflammatory processes and metal storage diseases cause persistent oxidative and nitrosative stress and lead to accumulation of LPO-derived DNA modifications (reviewed in (Nair et al., 2007)). Significantly increased levels of etheno-dAdo and etheno-dCyd adducts have been found in patients with primary hemochromatosis, chronic pancreatitis, Crohn's disease, ulcerative colitis, and familial adenomatous polyposis (Bartsch and Nair, 2005).

Reactive carbonyl species originate from glycation and lipid peroxidation. Glyoxal and methylglyoxal are the most studied ones. It has been reported that human keratinocytes treated with glyoxal accumulate DNA strand breaks, whereas methylglyoxal treatment generates DNA cross-links (reviewed in (Mano, 2012)). Carbonyl stress is believed to be one of the most important mechanisms of tissue damage in vascular complications of diabetes (Bourajjaj et al., 2003).

DNA replication errors

Nuclear DNA is replicated by polymerase (Pol) α, Polδ, and Polε. Although these enzymes ensure correct choice of deoxynucleotide triphosphate (dNTP) during DNA synthesis,

mistakes may occur. However, occasionally misincorporated nucleotides can be readily removed by the 3'-5' exonuclease activity of the polymerases (reviewed in (Lange et al., 2011)). The proofreading through 3'-5' exonuclease activity is crucial for preventing mutations. Polε proofreading-deficient mice die prematurely of intestinal adenomas and adenocarcinomas (Albertson et al., 2009). Mice lacking Polδ 3'-5' exonuclease activity are not viable (Uchimura et al., 2009). In addition to misincorporation, DNA polymerases can induce mutagenesis through incorporation of oxidized dNTPs during DNA synthesis (Hidaka et al., 2008).

1.2 DNA damage and replication checkpoints

DNA damage and replication checkpoints are signal transduction pathways that control a variety of cellular responses, including cell cycle arrest, activation of DNA repair, and apoptosis. These processes are believed to be primarily mediated by members of the phosphatidylinositol 3-kinase-like protein kinase (PIKK) family, Ataxia telangiectasia mutated (ATM), Ataxia telangiectasia and Rad3-related (ATR), and DNA-dependent protein kinase (DNA-PK), as well as by the Poly(ADP-ribosyl) polymerase (PARP) family members (reviewed in (Ciccia and Elledge, 2010)).

Localization of DNA damage response factors to specific DNA lesions is initiated by sensor proteins. Stretches of single strand DNA (ssDNA) resulting from replication fork stalling are rapidly coated by replication protein A (RPA). RPA-ssDNA stimulates binding and activation of a protein complex that consists of Rad17 and four replication factor C subunits (RFC_{2-5}). Rad17- RFC_{2-5} loads the PCNA-related Rad9-Hus1-Rad1 onto 5' or 3' DNA ends of RPA-ssDNA to initiate the ATR signaling pathway (Zou et al., 2003). SSBs are detected by the PARP family. Activated PARP1 and PARP2 modify themselves and other proteins by transient assembling of branched poly(ADP)-ribose chains, which enable interactions with single-strand break repair (SSBR) proteins and regulate chromatin structure at sites of DNA damage to block transcription and facilitate repair (reviewed in (Caldecott, 2008)). DSBs are rapidly bound by the Ku heterodimer (Ku70/Ku80), which activates the catalytic subunit of DNA-PK to trigger repair through canonical non-homologous end joining (NHEJ) (Mahaney et al., 2009). PARP1 may compete with Ku and probably together with DNA ligase III, participate in an alternative pathway of non-homologous end joining (alt-NHEJ) (Wang et al., 2006). In addition, PARP-3 has been shown to be activated by DNA DSBs *in vitro* and function to accelerate NHEJ (Rulten et al., 2011). Finally, the DSB lesions can be recognized

by Mre11-Rad50-Nbs1 (MRN) complex, which recruits ATM and prepares DNA for repair through homologous recombination (Williams et al., 2007). The mechanism of single base damage detection by DNA glycosylases is less clear. Recent structural and biophysical studies conclude that error-prone scanning is provided through an electrostatically guided migration of these enzymes along the DNA backbone. Once an inappropriate nucleotide is found, the excision process begins (reviewed in (Friedman and Stivers, 2010)).

The recognition of DNA damage results in initiation of a signal transduction cascade. The list of downstream targets is far from complete and is beyond the scope of this thesis. Briefly, ATM and ATR act mainly through phosphorylation of their respective substrates, checkpoint kinases (Chk) 2 and 1, which inhibit cell division cycle 25 homologs (Cdc25) A, B, and C, and activate the tumor suppression transcription factor p53. The Cdc25 homologs control dephosphorylation of cyclin-dependent kinases (Cdk) Cdk1 and Cdk2, which are required for S-phase progression, promotion of G_1/S, and G_2/M transition. p53 regulates transcription of the Cdk2 and PCNA inhibitor p21, proapoptotic BAX and PUMA proteins, which induce cell-cycle arrest, apoptosis, or senescence (reviewed in (Ciccia and Elledge, 2010)).

1.3 DNA damage tolerance

DNA damage tolerance (DDT) mechanisms, translesion synthesis (TLS), and template switching facilitate the continuation of DNA replication in the presence of polymerase-blocking lesions. In mammals TLS is governed by at least seven low-fidelity DNA polymerases: Polη, Polι, Polκ, REV1, Polζ, Polθ and Polν. These polymerases lack 3'-5' exonuclease proofreading activity and are thus error-prone (reviewed in (Stallons and McGregor, 2010)). The mechanisms by which lesions are bypassed and a particular TLS Pol is selected are not yet fully elucidated. There are currently two models: the polymerase-switching and the gap-filling. The polymerase-switching model is believed to occur when a replication fork is blocked during DNA synthesis. It is likely that RPA-ssDNA recruits Rad18, which recruits the E2 ubiquitin ligase Rad6 to the damage site. The Rad6/Rad18 heterodimer initiates proliferating cell nuclear antigen (PCNA) monoubiquitination, which is required for replacement of the stalled replicative DNA polymerase with a TLS polymerase. The switch is followed by lesion bypass and extension of the primer-template. In an alternative model, TLS polymerases fill in ssDNA gaps during late S or early G2 phase (reviewed in (Waters et al., 2009)).

Template switching is less studied. It has been suggested that DNA lesions may be avoided either by replication fork reversal and use of the nascent sister strand or invasion of the sister duplex by a single-stranded gap (reviewed in (Chang and Cimprich, 2009)).

1.4 DNA repair mechanisms

1.4.1 Base excision repair

The base excision repair (BER) pathway corrects oxidation, alkylation, and deamination of DNA bases. It is also responsible for elimination of misincorporated nucleotides, abasic sites, and single-strand breaks (reviewed in (Zharkov, 2008)). BER can be presented as a multistep process (Figure 1).

BER is initiated by lesion-specific DNA glycosylases, which detect the incorrect or damaged base and catalyze cleavage of the N-glycosidic bond (reviewed in (Dalhus et al., 2009)). DNA glycosylases can be classified according to their reaction mechanism (monofunctional versus bifunctional). Most DNA glycosylases are monofunctional and require assistance of a separate enzyme, apurinic/apyrimidinic endonuclease 1 (APE1), which cleaves the phosphodiester bond located 5' to the AP site and leaves a nick containing a 3'OH and 5'dRP moiety. Subsequently, Polβ removes the dRP group and inserts the correct nucleotide. The bifunctional glycosylases have AP lyase activity and can cleave the DNA strand by β-elimination (3' to the AP site) or by β/δ-elimination (3' and 5' to the AP site). The gap resulting from β-elimination is processed by APE1. β/δ-elimination is APE1-independent, but requires recruitment of the phosphatase/kinase PNKP to remove the 3'phosphate residue and prepare the gap for the repair synthesis step (reviewed in (Svilar et al., 2011)).

BER further proceeds by short-patch or long-patch sub-pathways. The short-patch BER is typically characterized by insertion of only one nucleotide by Polβ or Polλ and strand sealing by either DNA ligase I or a complex of DNA ligase IIIα and the scaffold protein XRCC1 (reviewed in (Kim and Wilson, 2012)). DNA synthesis during the long-patch BER is mediated by Polβ or Polδ and Polε, which in conjunction with PCNA and RFC form a 2-12 nucleotide flap. Finally, the structure-specific flap endonuclease 1 (FEN1) removes the intermediate and ligase I seals the substrate (reviewed in (Krokan and Bjoras, 2013)).

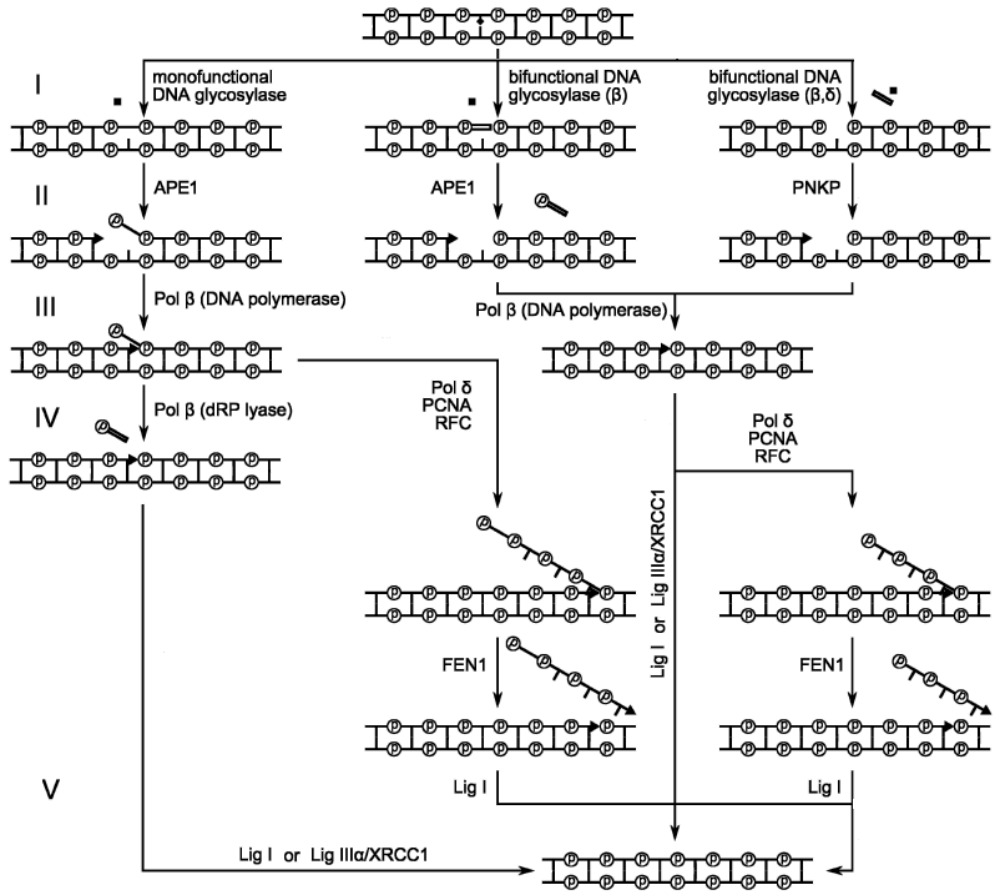


Figure 1. Schematic representation of the base excision repair pathway.

The core enzymatic steps: damage recognition and removal (I), DNA strand incision and 3'termini tailoring (II), gap filling (III), 5'termini tailoring (IV), ligation (V). The figure was modified from (Zharkov, 2008).

Recent findings suggest the existence of a distinct Polβ-dependent two-nucleotide insertion BER mechanism in extracts from non-proliferating cells, indicating that the concept of short and long-patch repair can be further extended (Akbari et al., 2009). What determines the choice of sub-pathway is not clear, but it may be influenced by the type of DNA glycosylase, concentration of ATP near the AP site, protein-protein interactions, and the cell cycle stage (reviewed in (Fortini and Dogliotti, 2007)).

A more comprehensive description of several BER core players is given below.

Uracil-DNA glycosylase

Human uracil-DNA glycosylase is present in two forms: mitochondrial (UNG1) and nuclear (UNG2). These enzymes are generated by transcription from different promoters and alternative splicing in the *UNG* gene located in chromosome 12q24.1 (Haug et al., 1996; Nilsen et al., 1997). The catalytic domains in UNG1 and UNG2 are identical, containing a conserved DNA-binding groove and a recognition pocket, which ensures selectivity for uracil by shape complementary, main chain, and Asn-204, Tyr-147, Phe-158 side chain hydrogen bonds (Mol et al., 1995). The mechanism of base excision was elucidated by mutational studies and co-crystallization of human UNG with a uracil-containing DNA duplex (Slupphaug et al., 1996). It includes stabilization of the extrahelical conformation (DNA "pinching"), creation of complementary interactions with uracil ("pull"), expelling of the nucleotide from the double helix by insertion of a hydrophobic leucine side chain into the minor groove ("push"), and cleavage of the N-glycosidic bond. Excision is achieved through the attack of a water molecule, which donates a proton to the uracil and a hydroxyl group to the deoxyribose (reviewed in (Krokan et al., 2002)). The N-terminal amino acids in UNG1 and UNG2 determine their intracellular localization (Nilsen et al., 1997). UNG1 is the only known uracil- removing enzyme involved in mitochondrial BER (Akbari et al., 2008). It also repairs 5-hydroxyuracil (5-hU), isodialuric acid, and alloxan, the cytosine-derived products of oxidative base damage (Dizdaroglu et al., 1996; Krokan et al., 1997). Notably, oxidative stress results in increased expression of UNG1 (Akbari et al., 2007). UNG2 has a central role in elimination of misincorporated dUMP at the replication forks (Otterlei et al., 1999). *Ung*$^{-/-}$ mice exhibit a 20-30-fold decreased capacity for removal of misincorporated dUMP and elevated steady-state genomic uracil (Nilsen et al., 2000). UNG2 is regulated by post-translational modifications of N-terminal residues. It has been shown that phosphorylation at Ser23 increases catalytic turnover against ssDNA and association with RPA in late G1 and early S phase, whereas phosphorylation at Ser-64 and Thr-60 throughout the S phase reduces binding to RPA and facilitates ubiquitinylation and proteosomal degradation of UNG2 in G2 phase (Hagen et al., 2008). In support of replication associated repair, co-immunoprecipitation experiments revealed BER complexes containing UNG2, APE1, POLα, POLβ, POLδ, POLε, DNA PK, XRCC1, PCNA, DNA ligase I, MCM7, and the cell cycle regulatory protein cyclin A (Akbari et al., 2004; Parlanti et al., 2007). UNG2 plays a major role in repair of uracil resulting from cytosine deamination, at least in the human genome (Kavli et al., 2002). Mutagenic processing of uracil generated by activation-induced cytidine

deaminase (AID) in antigen-stimulated B-cells during somatic hypermutation and class switch recombination is discussed in section 1.6.2.

Single-strand-selective monofunctional uracil-DNA glycosylase 1 (SMUG1)

SMUG1 is a member of the uracil DNA glycosylase superfamily (Haushalter et al., 1999). In addition to uracil, it removes 5-hU, 5-fluorouracil (5-FU), and other oxidized pyrimidines with bulky substitutions at the C5 position. SMUG1 is the main DNA repair enzyme responsible for 5-hydroxymethyluracil (5-hmU) excision (Boorstein et al., 2001; Kavli et al., 2002; Masaoka et al., 2003). Unlike UNG2, SMUG1 is constitutively expressed throughout the cell cycle, does not accumulate in replication foci and maintains contact with both DNA strands by a wedge motif. High affinity for U:G mispairs and low turnover number make SMUG1 an ideal candidate for repair of uracil resulting from cytosine deamination in non-proliferating cells and proliferating cells outside the S phase (reviewed in (Visnes et al., 2009)). SMUG1 can bind to abasic sites and inhibit their cleavage by AP-endonucleases until the next repair protein is recruited to the potentially mutagenic DNA damage (Pettersen et al., 2007). Interestingly, disruption of the *Smug1* gene in $Ung^{-/-}$ mice does not affect animal viability. They develop normally beyond one year of age despite the ablation of all detectable uracil and 5-hydroxymethyluracil-excision activity. This may be explained by the backup mismatch repair pathway because $Smug1^{-/-}Ung^{-/-}Msh2^{-/-}$ triple knockout mice are cancer-prone and die prematurely (Kemmerich et al., 2012).

Thymine-DNA glycosylase (TDG)

TDG is a monofunctional mismatch-specific DNA glycosylase that eliminates uracil and thymine resulting from deamination of cytosine and 5-methylcytosine (5-meC), respectively (Neddermann et al., 1996; Neddermann and Jiricny, 1994). Its substrates also include 5-FU, 5-hU, 5-hmU, 3'N4-ethenocytosine (εC), and oxidation products of 5-meC, such as 5-formylcytosine (5-fC), and 5-carboxylcytosine (5-caC) (Hardeland et al., 2001; Maiti and Drohat, 2011). TDG shares a common α/β fold structure with other uracil DNA glycosylases and removes U mispaired to G with almost 10-fold higher efficiency than T from T:G mismatches (reviewed in (Krokan et al., 2002)). Unlike UNG2, TDG has an extremely low turnover number and is mostly expressed during the G1 and G2/M phase, suggesting a distinct role for U:G repair outside the S phase (Hardeland et al., 2007). Importantly, dissociation of TDG from the AP site is facilitated by the ubiquitine-like proteins SUMO1 and SUMO2/3, which bind the C-terminal end of the catalytic domain (Baba et al., 2005). The

N-terminus of TDG contains a lysine-rich regulatory domain, a target for acetylation by histone acetyl-transferases CBP and p300, which promotes recruitment of APE1 (Tini et al., 2002). Interestingly, TDG has also been shown to interact with DNA methyltransferases Dnmt3a/3b and a number of transcription factors, including thyroid transcription factor 1 (TTF1), retinoic acid receptor, and retinoid X receptor. These interactions are believed to be important in coordination of DNA methylation/demethylation and transcriptional regulation of target genes (Li et al., 2007; Missero et al., 2001; Um et al., 1998). This has been recently supported by the finding that germline ablation of the *Tdg* gene is embryonic lethal in mice (Cortazar et al., 2011). BER-dependent active demethylation pathway will be discussed in section 1.6.4.

Methyl-CpG-binding protein 4 (MBD4)

MBD4 belongs to the helix-hairpin-helix superfamily of DNA glycosylases (Wu et al., 2003; Zhang et al., 2011). Its substrate specificities resemble those of TDG, although MBD4 preferentially binds to 5-mCpG:TpG mismatches and cannot remove 5-fC and 5-caC from DNA (Hendrich et al., 1999; Manvilla et al., 2012). MBD4 counteracts the mutagenic effects of DNA deamination (reviewed in (Bellacosa, 2001)). Moreover, it associates with the human homolog 1 of *E.coli* MutL (MLH1) and may thus have a role in the mismatch repair pathway (Bellacosa et al., 1999). In addition, MBD4 directly binds to the transcription repression domains of Sin3A and histone deacetylase 1 (HDAC1) and may therefore be involved in epigenetic regulation (Kondo et al., 2005). Finally, interaction of MBD4 and Fas-associated death domain protein may provide a link between genome surveillance and apoptosis (Screaton et al., 2003). Mice deficient in MBD4 show only a slight increase in C→T transition mutations. The lack of an apparent phenotype can probably be explained by TDG repair activity (Millar et al., 2002; Wong et al., 2002).

Apurinic/apyrimidinic endonuclease 1 (APE1)

APE1 is the major BER enzyme incising AP sites and generating 3'-hydroxyl and deoxyribose-5'-phosphate (5'-dRP) termini. In addition to 5' AP endonuclease activity, APE1 exhibits weak 3'-5' DNA exonuclease, 3'-phosphatase, 3'-phosphoglycolate and, RNase H-like enzymatic activities. Another important function of APE1 is the regulation of gene expression. APE1 can alter the DNA binding of several transcription factors involved in cancer promotion and progression, such as early growth response protein (Erg-1), nuclear factor kappa-light-chain-enhancer of activated B-cells (NF-κB), and p53. Furthermore, APE1

can bind to the negative calcium response elements (nCaRE) of some promoters (i.e., parathyroid hormone promoter) and thus act as a transcriptional repressor (reviewed in (Abbotts and Madhusudan, 2010; Tell et al., 2010)). It has been demonstrated that APE1 may also be involved in RNA metabolism. In particular, APE1 appears to interact with nucleoplasmin (NPM1), cleave *c-myc* mRNA *in vitro* and regulate *c-myc* mRNA levels in cells (Kim et al., 2010; Vascotto et al., 2009). Several studies have reported that the functional activity of APE1 can be modulated at the post-translational level. For instance, acetylation of APE1 by the transcriptional co-activator p300 enhances its binding affinity for nCaRE, while proteolytic cleavage of the N-terminal 33 amino acids domain reduces APE1 nuclear accumulation (Bhakat et al., 2003; Chattopadhyay et al., 2006). Finally, it has been shown that homozygous deletion of the *Ape1* gene in mice is embryonic lethal (Xanthoudakis et al., 1996). In human cells, down-regulation of *Ape1* results in cell cycle arrest and apoptosis, and has been correlated with increased AP site formation (Fung and Demple, 2005).

DNA polymerase β (Polβ)

Polβ is the major polymerase in BER, particularly in short-patch sub-pathway (reviewed in (Beard and Wilson, 2006)). It is also responsible for most of the two-nucleotide BER in non-proliferating cells (Akbari et al., 2009). Polβ interacts with APE1, XRCC1, PARP1, PARP2, and DNA ligase I (reviewed in (Almeida and Sobol, 2007)). Knockout of Polβ in mice results in embryonic lethality (Gu et al., 1994). *Polb*$^{+/-}$ mice are viable, but have an increased incidence of lymphomas (Cabelof et al., 2006). Importantly, Polβ lacks an intrinsic 3' to 5' proofreading exonuclease activity and shows an average base substitution error rate of approximately 1 per 4000 inserted nucleotides (Osheroff et al., 1999). Overexpression of Polβ has been reported in gastric, uterine, prostate, ovarian, and thyroid carcinomas (reviewed in (Lange et al., 2011)).

Flap endonuclease 1 (FEN1)

FEN1 is a member of the XPG/RAD2 endonuclease family and involved in Okazaki fragment maturation and long-patch BER. FEN1 processes 5'-flaps generated by Polβ or δ-mediated strand -displacement DNA synthesis. It also exhibits 5'-3' exonuclease and gap endonuclease activity, which may be involved in apoptotic DNA fragmentation and stalled replication fork rescue (reviewed in (Shen et al., 2005)). More than 30 proteins are known to interact with FEN1, including the long-patch BER partners APE1, PCNA, Polβ, DNA ligase I, and PARP1.

Protein-protein interactions and post-translational modifications are critical for regulation of FEN1's functions in different pathways (reviewed in (Zheng et al., 2011)). For example, it has been shown that the scaffolding factor PCNA recruits FEN1 to the replication factories and stimulates its flap endonuclease and exonuclease activities (Wu et al., 1996; Zheng et al., 2007). In contrast, phosphorylation of FEN1 at Ser-187 in late S phase reduces its activities and abolishes PCNA binding (Henneke et al., 2003). Ablation of *Fen1* leads to embryonic lethality in mice (Larsen et al., 2003; Zheng et al., 2007).

DNA ligases I (LIG1) and IIIα (LIG3α)

DNA ligases catalyze the formation of a phosphodiester bond between adjacent 3'-hydroxyl and 5'-phosphate termini. There are strong indications of LIG1 involvement in BER. The S-phase specific interaction with PCNA, APE1, and Polβ links LIG1 with both short-patch and long-patch BER sub-pathways (Dimitriadis et al., 1998; Ferrari et al., 2003; Levin et al., 2000; Ranalli et al., 2002). LIG1 is also required for the joining of Okazaki fragments during DNA synthesis (Levin et al., 1997). Consistent with this, mouse embryonic fibroblast (MEF) cell lines generated from *Lig1* mutants accumulate DNA replication intermediates and demonstrate increased genome instability (Bentley et al., 2002). LIG3α has been shown to form a stable complex with the DNA repair protein XRCC1, which interacts with Polβ and PARP1 and thus links LIG3α with BER and single-strand break repair, respectively (Caldecott et al., 1996; Caldecott et al., 1994; Leppard et al., 2003). Importantly, LIG3α can also function in the mitochondrial BER pathway (Lakshmipathy and Campbell, 2000). Knock out of the *Lig3* gene in mice results in early embryonic lethality (Puebla-Osorio et al., 2006); however, the viability of *Lig3*$^{-/-}$ MEF cells can be rescued by expressing mitochondrial but not nuclear LIG3α, revealing its critical role in maintenance of mtDNA integrity (Simsek et al., 2011). Furthermore, Lig3 inactivation does not cause nuclear DNA repair deficiency, which indicates that Lig1 is the major ligase in the short-patch BER (Gao et al., 2011; Simsek et al., 2011).

Proliferating cell nuclear antigen (PCNA)

Eukaryotic PCNA forms a homotrimeric ring-shaped complex that encircles DNA and provides a scaffold for the proteins involved in chromatin remodeling, cell cycle regulation, and DNA replication and repair. PCNA is an important docking site for DNA glycosylases (UNG2, MYH, MPG), DNA polymerases (Polβ, Polδ, Polε), APE1, FEN1, LIG1, PARP1,

and XRCC1 (reviewed in (Moldovan et al., 2007)). The majority of proteins bind to PCNA through their conserved PIP (PCNA-interacting peptide) or APIM (AlkB homologue 2 PCNA-interacting motif) region (Gilljam et al., 2009; Warbrick, 2000). PCNA is cell cycle-regulated with the highest expression level in S phase (Leonhardt et al., 2000). Coordination of its functions depends on the binding affinity of interaction partners (reviewed in (Maga and Hubscher, 2003)). Post translational modifications are an additional regulatory mechanism. For example, it has been shown that PCNA monoubiquitination facilitates replacement of high-fidelity replicative polymerases with error-prone polymerases, SUMOylation inhibits association with PIP-box proteins, while dephosphorylation triggers PCNA degradation (reviewed in (Moldovan et al., 2007)).

Poly(ADP-Ribose)Polymerase 1 (PARP1)

PARP1 is an ADP-ribosylating enzyme activated by DNA strand breaks. It binds to the site of damage, cleaves nicotinamide adenine dinucleotide (NAD^+), and adds generated ADP-ribose moieties to acceptor proteins, such as histones, chromatin regulators, transcription factors and DNA repair proteins (reviewed in (Krishnakumar and Kraus, 2010)). Up to 200 negatively charged ADP-ribose units can link to each other via glycosidic ribose-ribose bonds and serve as an interaction platform for other molecules. Poly(ADP)ribose-glycohydrolase (PARG) can easily release ADP-ribose from the acceptor proteins and thus maintain the dynamic equilibrium between polymer synthesis and degradation (reviewed in (Luo and Kraus, 2012)). PARP1 is involved in single-strand and double-strand break repair. It recruits XRCC1, Polβ, and LIG3α to the DNA damage foci (reviewed in (De Vos et al., 2012)). Recent data indicate that PARP1 may also have a 5′-dRP/AP lyase activity (Khodyreva et al., 2010). *PARP1*$^{-/-}$ mice are viable, but highly susceptible to γ-irradiation and DNA alkylating agents (Shall and de Murcia, 2000). The mild phenotype may be explained by the functional redundancy with other PARP family members. Indeed, it has been shown that double knockout of PARP1 and PARP2 in mice results in embryonic lethality (Menissier de Murcia et al., 2003).

X-ray repair cross-complementing protein 1 (XRCC1)

XRCC1 provides scaffold for several DNA glycosylases (UNG2, OGG1) and downstream BER and SSBR proteins (APE1, Polβ, LIG3α, PNK, PCNA, PARP1, PARP2) (reviewed in (Hanssen-Bauer et al., 2012)). Cells lacking functional XRCC1 are viable, but hypersensitive

to ionizing radiation, oxidative stress and alkylating agents (reviewed in (Caldecott, 2003)). Knock out of the *Xrcc1* gene in mice results in embryonic lethality (Tebbs et al., 1999).

1.4.2 Nucleotide excision repair

Nucleotide excision repair (NER) removes bulky DNA lesions, such as cyclobutane pyrimidine dimers and pyrimidine-6,4-pyrimidone photoproducts, induced by UV radiation, chemical carcinogens and chemotherapeutic agents. There are two NER sub-pathways: global genomic repair (GGR) and transcription-coupled repair (TCR), which differ only during the DNA damage recognition step. In GGR, the DNA lesion is detected by the *xeroderma pigmentosum* group C protein complex XPC-HR23B, which recruits transcription factor IIH (TFIIH). In TCR, this role is performed by CSA and CSB proteins, activated by stalled RNA polymerase. The TFIIH subunits XPB and XPD possess 3'-5' and 5'-3' helicase activity and can thus unwind DNA in both directions to ensure access for the other NER factors. The DNA strand is further incised by the structure specific endonucleases XPF and XPG, which release the damaged base in the form of 24-32 nucleotide-long oligomers. DNA polymerases δ and ε use the undamaged strand as a template to fill in the gap (reviewed in (Leibeling et al., 2006)). The final step, DNA ligation, is performed by LIG1 during the S phase and by LIG3α throughout the whole cell cycle (Moser et al., 2007). It has been shown that mutations in NER associated genes result in the rare recessively inherited human syndromes: *xeroderma pigmentosum*, Cockayne syndrome, and trichothiodystrophy (reviewed in (Cleaver et al., 2009)). Defective NER may also play a critical role in chronic neurodegenerative disorders, such as Alzheimer's and Parkinson's disease (Sepe et al., 2013).

1.4.3 Mismatch repair

Mismatch repair (MMR) corrects replication errors resulting from nucleotide misincorporation and DNA polymerase slippage. MMR proteins have also been implicated in DNA damage-induced cell cycle arrest and apoptosis, homologous recombination, inter-strand crosslink repair, trinucleotide repeat expansion, and antibody diversification. In eukaryotes, mispairs of 1 or 2 nucleotides are recognized by MSH2-MSH6 (MutSα) heterodimer. Mismatches containing up to 16 base pairs are processed by MSH2-MSH3 (MutSβ). Binding of MutS homologs to DNA triggers ATP-dependent conformational changes and facilitates recruitment of exonuclease 1 (EXO1) and endonuclease MutLα (MLH1-PMS2). EXO1 starts 5' directed mismatch excision in the presence of RPA. It may also catalyze MutLα-associated 3'-nick directed repair. After the error is removed, a new

DNA strand with correct base pairing is synthesized by Polδ. Other proteins involved in MMR are PCNA, RFC, FEN1, HMGB1 and LIG1 (reviewed in (Fukui, 2010; Li, 2008)). Defects in MMR lead to microsatellite instability and predispose to colorectal cancers (reviewed in (Boland and Goel, 2010)).

1.4.4 Double-strand break repair

DNA double-strand breaks (DSBs) in mammalian cells can be repaired by two distinctly different mechanisms, homologous recombination (HR) and nonhomologous end-joining (NHEJ) (reviewed in (Brandsma and Gent, 2012)). In "classical" HR, broken DNA ends are resected by the MRN-complex (Mre11, Rad50, Nbs1) and CtBP-interacting protein. The resulting 3'-single stranded DNA is kept unwound by RPA. Subsequently, BRCA2 initiates the replacement of RPA by RAD51 to form a nucleoprotein filament that can search for DNA sequences similar to that of the 3' overhang. When the template is found, the nucleoprotein filament invades the sister chromatid and makes a displacement loop (D-loop) in a process called strand invasion. Extension of the invading 3' strand by DNA polymerases changes the D-loop to a cross-shaped structure known as a Holliday junction. The second end of the broken DNA also forms a Holliday junction with the homologous chromosome. The junctions are further resolved in crossover or non-crossover products (reviewed in (Krejci et al., 2012)).
An alternative synthesis-dependent strand annealing HR pathway results only in non-crossover products. In this model, the extended 3' end of the invading strand anneals to the other break end through complementary base pairing. HR is restricted to S and G2 phases of the cell cycle (reviewed in (Chapman et al., 2012)). Mutations in the HR genes have been linked to the most common form of hereditary breast and ovarian cancer, ataxia-telangiectasia-like disorder, and the Nijmegen breakage syndrome (reviewed in (Thompson and Schild, 2002)).
In contrast to HR, NHEJ repairs DSBs throughout the cell cycle and more often results in small insertions, deletions, substitutions or translocations. It is initiated by the Ku70/80 heterodimer, which binds to the broken DNA ends and recruits DNA-PKcs. Further processing requires polynucleotide kinase/phosphatase to add 5'-phosphates and eliminate 3'-phosphates, tyrosyl-DNA phosphodiesterases I and II to preclude peptide fragments of topoisomerases, the Artemis and Metnase endonucleases to cut off unnecessary DNA overhangs, and Polμ or Pol λ to create compatible ends. Finally, XRCC4, DNA ligase IV and Xrcc4 like factor (XLF) rejoin the DNA strands (reviewed in (Lieber, 2010)). NHEJ is the primary mechanism during V(D)J and class switch recombination. Aberrant NHEJ is

associated with impaired embryonic development, increased radiosensitivity, severe immunodeficiency, and predisposition to lymphomas (reviewed in (Malu et al., 2012)). Recent studies suggest that a backup microhomology-mediated end joining (alt-NHEJ) route may partially rescue the deficient DSB repair in the absence of the core NHEJ components: Ku70/80, DNA-PKcs and DNA ligase IV. This pathway employs PARP1, histone H1, Polβ, LIGIII, XRCC1, and the MRN complex (reviewed in (Mladenov and Iliakis, 2011)).

1.5 AID/APOBEC cytidine deaminases

The APOBEC (apolipoprotein B mRNA-editing catalytic polypeptide) family of cytidine deaminases comprises 11 members: APOBEC1, APOBEC2, APOBEC3 (A, B, C, DE, F, G, H), APOBEC4 and AID (Figure 2). These proteins are characterized by the presence of one or two zinc-binding catalytic domains with the conserved His-X-Glu-$X_{(23-28)}$-Pro-Cys-$X_{(2-4)}$-Cys sequence (X is any amino acid), which mediate hydrolytic removal of the exocyclic amino group from cytidine or deoxycytidine to form uridine or deoxyuridine (reviewed in (Vieira and Soares, 2013)).

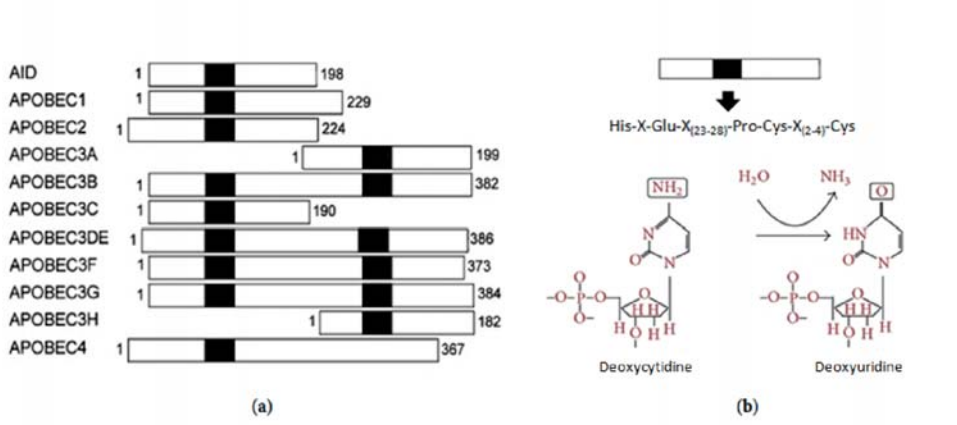


Figure 2. Schematic representation of the human APOBEC proteins. (a) The number of zinc-binding catalytic motifs (black) and amino acids for each protein. (b) The APOBEC-mediated hydrolytic deamination reaction. The figure was modified from (Vieira and Soares, 2013).

APOBEC1

APOBEC1 is the apolipoprotein B (ApoB) mRNA-editing enzyme, but can process ssDNA when overexpressed in bacterial assays (Harris et al., 2002; Petersen-Mahrt and Neuberger,

2003; Teng et al., 1993). It acts in the nucleus in the presence of the APOBEC1 complementation factor (ACF) (Lau et al., 1991; Sowden et al., 2002). APOBEC1 deaminates mRNA-cytidine at nucleotide position 6666. Conversion of a CAA glutamine codon to a translation stop UAA codon enables expression of full length (ApoB100) and truncated (ApoB48) isoforms. ApoB100 is a non-exchangeable structural component of very low density lipoproteins (VLDLs), which are produced in liver and converted to low density lipoproteins (LDLs) in the bloodstream. LDLs play a central role in atherogenesis. ApoB48 is a part of chylomicrons responsible for transport of exogenous dietary lipids in the proximal small intestine (reviewed in (Blanc and Davidson, 2010)). Recently, 32 additional mRNA targets of APOBEC1 have been identified. All of them are located in AU-rich segments of transcript 3′ untranslated regions (3′ UTRs) (Rosenberg et al., 2011).

APOBEC2

APOBEC2 is cardiac- and skeletal muscle-specific (Liao et al., 1999). Although its crystal structure has been solved, functions remain unclear (Prochnow et al., 2007). Unlike other family members, APOBEC2 has no deaminase activity on ssDNA in yeast- and bacteria-based mutator assays (Harris et al., 2002; Lada et al., 2011). However, it has been demonstrated that aberrant expression of APOBEC2 results in nucleotide alterations in the transcripts of the *Eif4g2* and *PTEN* tumor suppressor genes and contributes to the development of liver and lung malignancies in mice (Okuyama et al., 2012).

APOBEC3

The APOBEC3 subfamily comprises proteins with one (APOBEC3A, APOBEC3C, APOBEC3H) and two (APOBEC3B, APOBEC3DE, APOBEC3F, APOBEC3G) catalytically active zinc-binding deaminase domains. These domains are located at the N- and C-terminus (Harris et al., 2002; Navarro et al., 2005; Wedekind et al., 2006). These enzymes target specific ssDNA sequences. For instance, APOBEC3F edits cytosine at dTC contexts, whereas APOBEC3G prefers dCC (Armitage et al., 2008; Beale et al., 2004; Bishop et al., 2004; Harris et al., 2003). APOBEC3 proteins play important role in innate immune responses to exogenous viruses and endogenous retroelements (reviewed in (Vieira and Soares, 2013)). APOBEC3G is the best studied one. It restricts replication of human immunodeficiency virus type 1 (HIV-1). APOBEC3G deaminates the HIV-1 minus-strand cDNA during viral reverse

transcription and leads to degradation of the viral DNA or functional inactivation of the provirus (reviewed in (Smith et al., 2012)). There are two forms of APOBEC3G within the cell: protein of low molecular mass (LMM) and ribonucleoproteic complex of high molecular mass (HMM). Enzymatically inactive HMM can be converted to functional LMM through RNase digestion (Chiu et al., 2005). The presence of LMM APOBEC3G is correlated to the reduced susceptibility to HIV-1 infection (Ellery et al., 2007; Stopak et al., 2007). However, a HIV-1 accessory protein known as viral infectivity factor (Vif) can induce APOBEC3G polyubiquitination and proteasomal degradation and, thus, help to overcome this host defense mechanism (reviewed in (Smith et al., 2012)).

Recent data indicate that APOBEC3 enzymes can cause mutations in human mitochondrial and nuclear DNA (Suspene et al., 2011). This phenomenon will be discussed in section 1.7.

APOBEC4

The APOBEC4 substrate is not known. It is expressed primarily in testicles and may function as an editing enzyme for mRNAs involved in spermatogenesis (Rogozin et al., 2005).

AID

AID is an ssDNA-specific cytidine deaminase essential for the *Ig* gene diversification in mature B lymphocytes. It is a relatively small protein consisting of 198 amino acids encoded by the *AICDA* gene. AID preferentially targets WRCY (W=A or T, R=A or G, C and Y=T or C) hotspot motifs and generates uracils, the assumed base intermediates in somatic hypermutation and class-switch recombination (reviewed in (Gazumyan et al., 2012)). Its off-target activity is associated with development of B cell lymphomas (Hakim et al., 2012; Pasqualucci et al., 2008; Ramiro et al., 2006; Robbiani et al., 2008). Aberrant expression of AID in non-B cells is also linked to oncogenesis (Chiba and Marusawa, 2009; Lin et al., 2009; Morisawa et al., 2008; Okazaki et al., 2003; Pauklin et al., 2009; Robbiani et al., 2009). Clearly, AID must be tightly regulated.

The *AICDA* gene expression depends on binding of various transcription factors, including NF-κB, STAT6, Sp1, Sp3, HoxC4, Pax5, Smad3/4, the E-box proteins and estrogen (Dedeoglu et al., 2004; Gonda et al., 2003; Park et al., 2009; Pauklin et al., 2009; Sayegh et al., 2003; Tran et al., 2010; Yadav et al., 2006). For example, interaction with NF-κB is important for triggering AID expression by viral infection, toll-like receptor or TNFα stimulation (Gourzi et al., 2007; Pauklin et al., 2009; Xu et al., 2010). Inhibition of Pax5 and

the E-proteins by Id2, Id3 and Blimp-1 has the opposite effect (Gonda et al., 2003; Sayegh et al., 2003; Shaffer et al., 2002).

MicroRNAs (miRs) can down-regulate the level of protein by targeting mRNA transcript to degradation or preventing its translation (reviewed in (Pasquinelli, 2012)). For instance, activated B cells exhibit higher levels of AID in absence of miR-155 (Vigorito et al., 2007). Mutations of the miR-155 binding site on the *Aicda* 3' UTR are associated with the increased levels of AID protein and mRNA transcript (Dorsett et al., 2008; Teng et al., 2008).

Posttranslational modifications provide an additional layer of regulation. Phosphorylation of AID-S38 residue facilitates interaction with RPA. It has been suggested that RPA allows access to ssDNA bubbles and recruits factors of DNA repair and translesion synthesis during antibody diversification (Chaudhuri et al., 2007; Chaudhuri et al., 2004). Indeed, substitution of S38 with alanine (AID-S38A) did not affect the ssDNA deaminase activity, but significantly reduced somatic hypermutation and class-switch recombination in B cells stimulated *ex vivo* (Basu et al., 2005; McBride et al., 2006; McBride et al., 2008; Pasqualucci et al., 2006). Mutation of the other phosphorylation site (AID-T140A) impaired only somatic hypermutation (McBride et al., 2008). In contrast, AID-S3A mutants enhanced both class-switch recombination and C-MYC/IgH translocations, indicating that phosphorylation of S3 motif inhibits AID activity (Gazumyan et al., 2011).

Polyubiquitination has been shown to subject AID to proteosomal degradation in the nucleus (Aoufouchi et al., 2008). Interaction with heat shock protein 90 (Hsp90), Hsp40-Hsp70 system and DnaJa1 cochaperone seems to protect cytoplasmic AID from this posttranslational modification (Orthwein et al., 2010; Orthwein et al., 2012). A recent study suggests that the translation elongation factor 1α (eEF1A) also contributes to cytosolic retention and stabilization of AID (Hasler et al., 2011).

Nucleocytoplasmic shuttling plays a major role in functional regulation of AID (reviewed in (Zan and Casali, 2013)). The C terminus harbors a leucine-rich nuclear export signal (NES) that largely restricts AID to the cytoplasm (Brar et al., 2004; Ito et al., 2004; McBride et al., 2004; Patenaude et al., 2009). The nuclear import is mediated by a nuclear localization signal (NLS) (Patenaude et al., 2009). NLS directs AID to nucleoli, where it interacts with catenin-β-like 1 (CTNNBL1) and physically associates with nucleolin and nucleoplasmin. Importantly, release from nucleoli is dependent on the AID C-terminal motif. The exact mechanism is not yet clear, but it appears that subnuclear trafficking is an additional level of AID regulation (Hu et al., 2013).

1.6 Other functions of DNA repair proteins

DNA repair proteins are involved in the mechanisms of innate response to viral infection, adaptive immunity, cellular redox activities, control of DNA methylation, and epigenetic stability (reviewed in (Jacobs and Schar, 2012; Kelley et al., 2012)).

1.6.1 Innate immune response to viral infection

There is some evidence that the APOBEC3G-mediated uracilation of the viral genome is coupled to the action of BER proteins. Namely, that UNG2 removes U from the deaminated viral cDNA and that APE cleaves the resulting AP-sites. Indeed, expression of UNG2-inhibitor Ugi or UNG2- and APE-specific siRNA has been shown to suppress the antiviral activity of APOBEC3G (Yang et al., 2007). It is further supported by the fact that HIV-1 viral protein R (Vpr) targets UNG2 and SMUG1 for degradation by the ubiquitin-proteasome system (Schrofelbauer et al., 2007; Schrofelbauer et al., 2005).

1.6.2 Antibody diversification

Immunoglobulins (Ig) are heterodimeric proteins produced by B lymphocytes in response to infection. They exist both as a cell-surface B cell receptors and soluble molecules. All antibodies have the same basic structural units, two heavy (H) and two light (L) chains encoded by separate multigene families. Amino-terminal segments of both H and L chains compose variable (V) regions that bind antigens, while the carboxyl-terminal segments constitute constant (C) regions that define the biological effector functions such as complement fixation or binding to macrophages, natural killer cells, and neutrophils (reviewed in (Schroeder and Cavacini, 2010)). The number of genes required to encode the wide range of pathogen specific antibody molecules greatly exceeds the coding capacity of the inherited genome (Di Noia and Neuberger, 2007). The primary repertoire of functional B cell receptors is generated in the bone marrow by *V(D)J* recombination. The second stage of *Ig* gene diversification is antigen dependent and consists of somatic hypermutation and class switch recombination in the germinal centers of secondary lymphoid organs (Figure 3) (reviewed in (Tang and Martin, 2007)).

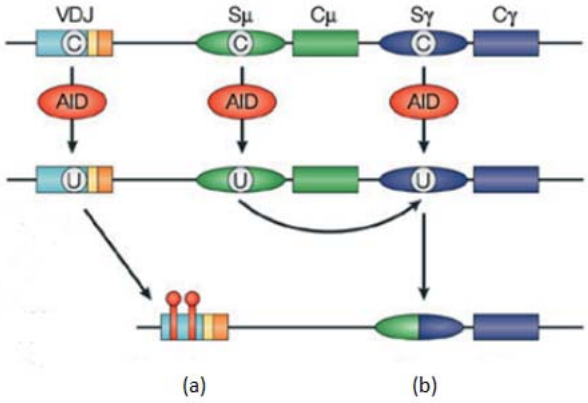


Figure 3. Schematic representation of the *Ig* gene diversification. AID deaminates cytosine to uracil to trigger somatic hypermutation (a) and class-switch recombination (b). V, D, J, S and C are the *Ig* gene variable, diversity, joining, switch, and constant regions, respectively. See the main text for further details. The figure was modified from (Harris and Liddament, 2004).

V(D)J recombination

V(D)J recombination is the rearrangement of germline variable (*V*), diversity (*D*), and joining (*J*) segments of B cell receptor genes via the introduction of site-specific DNA double-strand breaks. This process occurs between two types of recombination signal sequences (RSSs), termed the 12-RSS and the 23-RSS. In brief, the recombination activating proteins RAG1 and RAG2 assemble a pair of dissimilar RSSs into a synaptic complex and catalyze its cleavage by direct transesterification. The generated signal and coding DNA ends are reorganized and repaired by the classical NHEJ pathway, involving Ku70, Ku80, XRCC4, DNA ligase IV, and the Cernunnos/XLF protein. Genetic diversity may be further amplified by random nucleotide insertions and template-independent DNA fill-in synthesis by terminal deoxynucleotidyl transferase, Polμ and Polλ (reviewed in (Schatz and Swanson, 2011)).

Somatic hypermutation

In somatic hypermutation, the antigen binding affinity of the primary antibody repertoire is increased via the introduction of point mutations in the *IgV* gene by activation-induced cytidine deaminase. AID deaminates cytosine to uracil. Replication across the lesion results in C:G to T:A transition mutation. Uracil can also be eliminated by error-prone BER and MMR, which normally act in an error-free manner (Figure 4). In BER, AID-generated U is excised

by UNG2, which creates an AP site that can be corrected by Polβ, replicated over to produce any possible mutation, or processed by Rev1 and other TLS polymerases to yield both transversion and transition mutations at C:G base pair. Alternatively, U is recognized by the MMR proteins MSH2/MSH6 and removed by EXO1. The resulting large gap can be either faithfully repaired by Polδ, or filled in by Polη, generating mutations mainly on A:T pairs (reviewed in (Chahwan et al., 2012; Saribasak and Gearhart, 2012)).

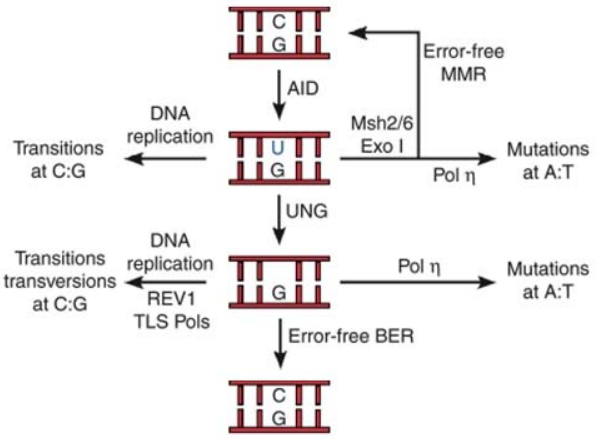


Figure 4. **Schematic representation of somatic hypermutation.** See the main text for further details. The figure was adopted from (Casellas et al., 2009).

There is no evidence to support a unique targeting of AID to the *Ig* loci. The *in vivo* hot-spot motif for AID, WRCY (W=A/T, R=A/G, Y=C/T) is widely distributed in non-*Ig* genes. That AID is associated with the transcriptional machinery cannot explain why it is recruited mainly to the IgV(D)J and S regions but not all highly transcribed genes (reviewed in (Chahwan et al., 2012; Maul and Gearhart, 2010)). Furthermore, AID has recently been reported to deaminate a large number of off-target sites, though at low levels (Yamane et al., 2011). If so, high-fidelity DNA repair mechanisms must protect non-Ig genes from AID-mediated mutagenisis. What makes BER and MMR pathways act in an erroneous manner during somatic hypermutation is not clear (reviewed in (Liu and Schatz, 2009)). There is some evidence that monoubiquitination of PCNA at Lys164 may facilitate recruitment of error-prone DNA polymerases (Arakawa et al., 2006). In line with this, mice expressing PCNA with a lysine-to-arginine substitution at residue 164 display a strong reduction in mutations at A:T base pairs (Langerak et al., 2007). It has been proposed that the cell cycle phase may also determine the fidelity of U processing. In a current model, U generated by AID during S

phase is faithfully repaired by BER, whereas deamination events outside of S phase are resolved by error-prone MMR pathway (reviewed in (Liu and Schatz, 2009)). However, a recent study indicates that UNG2 removes AID-induced U only during G1 phase (Sharbeen et al., 2012), contradicting this hypothesis.

Class switch recombination

Class switch recombination (CSR) is a unique intrachromosomal rearrangement between switch (S) regions of *IgH* genes resulting in the replacement of the heavy chain constant locus *Cμ* with *Cγ, Cε* or *Cα* loci and thus a change of the antibody isotype from IgM/IgD to IgG, IgA, or IgE (reviewed in (Stavnezer et al., 2008)). CSR is initiated by AID (Maul et al., 2011). Interaction with the RNA polymerase II-associated exosome complex enhances recruitment of AID to both DNA strands (Basu et al., 2011). Subsequent excision of U by UNG2 and cleavage of AP site by APE1/APE2 or MRN complex (MRE11/RAD50/NBS1) in two separate switch regions on both DNA strands generate DSBs, the substrate for CSR (Figure 4). The U:G mismatches can also be recognized by the MMR proteins MSH2/MSH6, which recruit EXO1 and MLH1-PMS2 to yield DSBs (reviewed in (Stavnezer et al., 2008)). CSR is completed when DSBs in Sμ and Sx regions of *IgH* genes are recombined by NHEJ or alt-NHEJ (see 1.4.4).

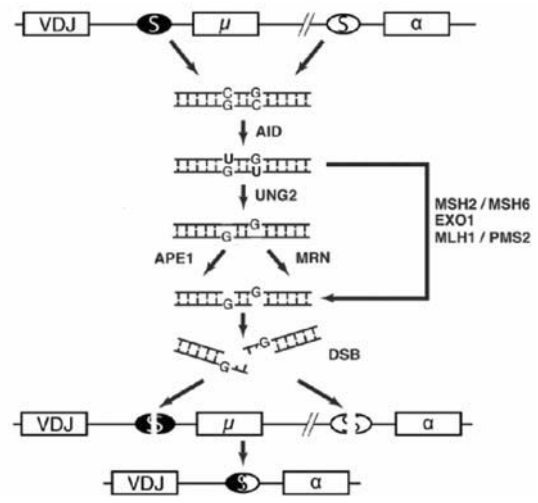


Figure 4. **Generation of DSBs in class-switch recombination.** AID deaminates cytosines to uracils in the *Ig* gene switch (S) regions during DNA transcription. UNG2 excises uracils and generates AP sites on both DNA strands. APE1 or MRN complex (MRE11/RAD50/NBS1) converts AP sites to SSBs. The MMR pathway

(MSH2/MSH6, EXO1, MLH1/PMS2) can also introduce SSBs at U:G mismatches. Opposing SSBs lead to DSBs. The figure was modified from (Offer et al., 2010).

Deficiency in UNG2 or MMR proteins (MSH2, MSH6, MLH1, PMS2 or EXO1) alone has been shown to perturb CSR (Bardwell et al., 2004; Ehrenstein and Neuberger, 1999; Imai et al., 2003; Peron et al., 2008; Rada et al., 2002; Schrader et al., 1999; Vora et al., 1999). Double knockout of UNG2 and MSH2/MSH6 in mice results in complete ablation of CSR (Rada et al., 2004; Shen et al., 2006). Notably, overexpression of SMUG1 can partially induce CSR in *Ung*$^{-/-}$ and *Ung*$^{-/-}$*Msh2*$^{-/-}$ mouse B-cells (Di Noia et al., 2006).

There is some evidence that AID can initiate DSBs in non-*Ig* genes (Robbiani et al., 2008). Aberrant CSR and chromosomal translocations, involving the *Ig* locus and proto-oncogenes (e.g. *BCL2, BCL5, C-MYC*), is a prominent hallmark of most human lymphomas derived from germinal center or post-germinal center B-cells (reviewed in (Kuppers, 2005)). It has recently been demonstrated that the HR factor XRCC2-mediated repair may protect the genome from off-target DSBs (Hasham et al., 2010).

1.6.3 DNA demethylation

DNA methylation is the key epigenetic mark associated with long-term gene silencing. It is conducted by a family of DNA methyltransferases (DNMTs) and occurs at ~70-80% of CpG sites (cytosine and guanine linked by one phosphate) distributed throughout the entire genome. The reverse of this process can restore gene expression. Passive DNA demethylation occurs in the absence of DNMT maintenance activity during DNA replication (reviewed in (Law and Jacobsen, 2010)). The mechanisms of active demethylation are only partially elucidated. It has been suggested that 5-meC is deaminated to thymine by an AID/APOBEC family member. The T:G mismatch is further processed by MBD4 (together with GADD45α) or TDG via the BER pathway. Alternatively, 5-meC can be hydroxylated to 5-hydroxymethylcytosine (5-hmC) by ten-eleven translocation TET proteins. 5-hmC is either deaminated by AID/APOBEC enzymes to 5-hmU (the substrate for SMUG1 and TDG) or subsequently oxidized to 5-fC and 5-caC, which can be removed by TDG (reviewed in (Nabel et al., 2012b)).

The mechanisms involved in active DNA demethylation are not agreed upon. Several lines of evidence support the deamination-initiated model (reviewed in (Teperek-Tkacz et al., 2011)). For example, overexpression of AID or APOBEC2 together with MBD4 glycosylase and GADD45α causes demethylation of the embryonic genome and injected plasmid DNA in

zebrafish (Rai et al., 2008). The use of siRNA against AID in heterokaryons of human fibroblasts and mouse embryonic stem cells inhibits demethylation and reactivation of the *OCT4* and *NANOG* pluripotency genes (Bhutani et al., 2010). Moreover, primordial germ cells in AID knockout mice are up to three times more methylated than wild-type controls (Popp et al., 2010). However, *Aid*$^{-/-}$, *Mbd4*$^{-/-}$ and *Gadd45α*$^{-/-}$ knockout mice are viable and have surprisingly few developmental defects, while ablation of the *Tdg* gene leads to embryonic lethality (Cortazar et al., 2011; Engel et al., 2009; Millar et al., 2002; Muramatsu et al., 2000; Wong et al., 2002), indicating that there may be more than one mechanism of active demethylation. The discovery of TET proteins responsible for converting 5-meC to 5-hmC suggested the possibility of oxidation-induced DNA demethylation pathways (reviewed in (Tan and Shi, 2012)). Recent studies have shown that 5-hmC formation is critical for embryonic development and cell differentiation (reviewed in (Pfeifer et al., 2013)). For instance, *Tet3* depletion impedes active demethylation of the paternal genome upon fertilization and leads to morphological abnormalities in mice (Gu et al., 2011). Loss of *Tet1* and *Tet2* function affects the expression of pluripotency and selected lineage marker genes in embryonic stem cells (Koh et al., 2011). Disruption of TET2 enzymatic activity is frequently observed in patients with myelodysplastic syndromes and myeloproliferative disorders (Ko et al., 2010).

It has been suggested that cytidine deaminases cooperate with TET proteins. In this model, AID/APOBECs mediate 5-hmC deamination to 5-hmU, which is removed by DNA glycosylases (Guo et al., 2011). However, SMUG1-deficient mice do not exhibit developmental deffects, though SMUG1 is the major enzyme responsible for excision of 5-hmU (Kemmerich et al., 2012). In addition, it has been recently reported that AID/APOBEC family members have no detectable 5-hmC activity (Nabel et al., 2012a). These data provide a strong argument against the proposed oxidation-deamination mechanism for active DNA demethylation.

Stepwise oxidation of 5-hmC to 5-fC and 5-caC by TET proteins followed by TDG-mediated base excision may constitute an alternative pathway (Ito et al., 2011; Maiti and Drohat, 2011). Indeed, it has been shown that depletion of TDG in mouse embryonic stem cells leads to 5-caC accumulation (He et al., 2011). There is also some evidence that 5-caC may be converted to cytosine by a putative decarboxylase without involvement of BER (Schiesser et al., 2012).

Clearly, the contribution of DNA glycosylases as well as the TET and AID/APOBEC enzymatic families to active DNA demethylation requires further investigation.

1.7 Oncogenesis

Global sequencing initiatives have made possible identification of somatic mutations from thousands of cancers (Hudson et al., 2010). Systematic computational analysis of the generated data has revealed the repertoire of mutational signatures (Alexandrov et al., 2013). These patterns result from the DNA damage and repair processes operating during the cellular lineage (reviewed in (Stratton, 2011)). Notably, a signature attributed to the APOBEC family of cytidine deaminases was observed in 16 out of 30 different primary cancer types. It is characterized by the simultaneous occurrence of C → T, C → G and in a less degree C →A mutations at TpCpN trinucleotides (Alexandrov et al., 2013). The C → T transitions arise through replication over unrepaired U:G mismatches. The C→ G transversions may originate from abasic sites processed by REV1 and Pol-ζ (Auerbach et al., 2006; Nelson et al., 1996). The mechanism of C →A substitution is not clear.

Most recently, a remarkable phenomenon of localized substitution hypermutation has been described in breast cancers. It is termed *kataegis*, which means rainfall in Greek. Foci of *kataegis* are characterized by clusters of C → T transitions and C→ G transversions at TpCpN trinucleotides on the same DNA strand (Nik-Zainal et al., 2012). These mutations are closely associated with somatic rearrangements and occur within a spatially localized genomic region on a single chromosome (Chen et al., 2012). Taking into account the substitution type and the sequence context, it has been suggested that APOBEC cytidine deaminases might be implicated in *kataegis* (Nik-Zainal et al., 2012). Indeed, breast cancer-like *kataegis* can be recapitulated in the budding yeast experimental system by expression of APOBEC3B and APOBEC3A (Taylor et al., 2013). APOBEC3B is highly expressed in most primary breast tumors and its presence correlates with increased levels of genomic uracil and C → T transitions (Burns et al., 2013a). Moreover, APOBEC3B-induced DNA deamination is responsible for mutational load in at least five other distinct cancer types – bladder, cervical, lung squamous cell carcinoma, lung adenocarcinoma, head and neck (Burns et al., 2013b).

The yeast experiments indicate that *kataegis* can be triggered by DNA breaks, which are generated through the processing of abasic sites introduced by joint action of APOBEC cytidine deaminase and UNG. Some residual *kataegis* in the UNG-deficient background might well be explained by the presence of spontaneous DNA breaks (Taylor et al., 2013). The exact mechanism is not clear and need to be elucidated.

1.8 RNA modifications and repair

Like DNA, RNA is constantly attacked by various exogenous and endogenous agents, which can cause chemical alterations, cross-linking, and strand breaks (reviewed in (Wurtmann and Wolin, 2009)). A total of 109 different RNA modifications has been identified (Cantara et al., 2011). In some cases, these structural changes ensure correct folding of the RNA molecule and have important regulatory functions (reviewed in (Agris, 2004; Helm, 2006)). For example, uridine isomerization reduces binding of protein kinase R to messenger RNA (mRNA) and enhances translation (Anderson et al., 2010). Ribose 2'-*O*-methylation provides a molecular signature for the distinction of self and non-self mRNA in host innate immune responses (Daffis et al., 2010; Zust et al., 2011). N^6-methyladenosine (m^6A), the physiological substrate of fat mass and obesity-associated protein (FTO), has recently been linked to energy homeostasis regulation (reviewed in (Jia et al., 2013)).

On the other hand, RNA modifications can be quite deleterious and result in translation of truncated proteins, ribosome dysfunction and RNA degradation. Oxidation of mRNA is thought to be one of the main contributing factors in the pathogenesis of neurodegenerative disorders, hemochromatosis, diabetes and heart failure (reviewed in (Kong and Lin, 2010; Poulsen et al., 2012)).

It remains largely unclear how the cell deals with damaged RNA. Several studies indicate that aberrant RNA molecules can be eliminated. In eukaryotes, mRNAs containing premature stop codons and lacking in-frame stop codons are degraded via the "nonsense-mediated decay" and "nonstop decay" pathways, respectively (reviewed in (Brogna and Wen, 2009; Klauer and van Hoof, 2012)). In addition, there are several nucleic acid chaperons that can sequester faulty RNA from translation. For instance, the Ro protein has been shown to act in quality control of 5S ribosomal RNA (rRNA) precursors and U2 spliceosomal small nuclear RNA (reviewed in (Wurtmann and Wolin, 2009)).

The discovery of phage-type transfer RNA repair pathway catalyzed by polynucleotide kinase-phosphatase and RNA ligase 1 demonstrated that cells may have specific mechanisms to restore RNA integrity (Amitsur et al., 1987). Indeed, it was found that the *E.coli* enzyme AlkB and its human homologue hABH3 can reverse RNA alkylation damage by demethylation of 1-methyladenine and 3-methylcytosine (Aas et al., 2003; Vagbo et al., 2013). Intriguingly, there is some evidence that another DNA repair enzyme, APE1, may have the endoribonuclease activity and play a role in rRNA quality control (Barnes et al.,

2009; Vascotto et al., 2009). Clearly, this unexpected cross-talk between the RNA post-transcriptional modification machinery and DNA repair awaits further studies.

1.9 Quantitative analysis of DNA and RNA damage

The identification and quantification of nucleic acid adducts can provide important information on the mechanism of action and biological relevance of genotoxic chemicals. A variety of techniques have been established to estimate the risk associated with the exposure to carcinogens derived from environmental and dietary sources. In general, there are four main approaches including ^{32}P-postlabeling, immunological assays, fluorimetric methods, and mass spectrometry (reviewed in (Brown, 2012)).

^{32}P-postlabeling

The ^{32}P-postlabeling method consists of four principal steps: enzymatic digestion of DNA to nucleoside 3′-monophosphates; enrichment of the adduct fraction of the DNA digest by solvent or solid-phase extraction, HPLC, or further enzymatic digestion; labeling of the 5'-position of the adducts by polynucleotide kinase-mediated transfer of ^{32}P-orthophosphate from [γ-^{32}P]ATP; separation of the labeled adducts by thin-layer chromatography (TLC) or HPLC and quantification by measuring ^{32}P-decay. The assay provides sensitivity of 1 adduct/10^{10} nucleotides and requires only 1-10μg of DNA. It can be used for detection of numerous DNA lesions resulting from exposure to a wide variety of chemical compounds (e.g. polycyclic aromatic hydrocarbons, alkenylbenzene derivatives, heterocyclic amines, etc.) (reviewed in (Phillips and Arlt, 2007)). On the other hand, this method does not provide structural information for the identification of unknown DNA adducts. The procedure is time consuming and labor intensive. It requires handling of radioactive phosphorus and lacks internal standardization. Moreover, it is known that polynucleotide kinase can mediate labeling of non-nucleic acid components and thus lead to the false-positive results (reviewed in (Klaene et al., 2013)).

Immunological assays

Radioimmunoassays (RIAs) and enzyme-linked immunosorbent assays (ELISAs) have found the most extensive application. These methods are based on the use of antibodies against certain DNA modifications. In RIAs, the concentration of antigen in a probe is determined by measuring its ability to compete with a fixed amount of radioactive antigen ("tracer") for a

limited quantity of antibody. Although highly sensitive and reproducible, this technique requires special precautions for handling of radioactive material (reviewed in (Santella, 1999)). In ELISAs, the "tracer" is substituted by immunogen coated onto a microtiter plate. After the antigen is immobilized, the secondary enzyme-linked antibody and the enzymatic substrate are added to produce a visible signal. Various detection strategies can be applied (colorimetric, fluorescent, chemiluminescent). The quantity of antigen in the probe is obtained by comparison with the standard curve made by serial dilution of either the modified denatured DNA or the monoadduct mixed with diluted antibody. This approach is inexpensive and easy to perform, but requires relatively large amounts of biological sample. In total up to 200 μg of DNA should be used to provide maximal sensitivity of 1 adduct/10^8 nucleotides (reviewed in (Phillips et al., 2000)).

The availability and selectivity of antibodies can greatly limit the application of immunological assays. It appears that antibodies may cross-react with structurally similar and unmodified DNA bases. Moreover, none of these methods permit identification of unknown nucleic acid adducts (reviewed in (Santella, 1999)).

Fluorimetric methods

The inherent or chemically induced fluorescent properties of DNA lesions can be used for their detection and quantification. Synchronous fluorescence spectrophotometry (SFS) of benzo[a]pyrene diol epoxide-DNA adducts is one of the first reported applications of this technique. In SFS, simplified spectra are generated by scanning of both excitation and emission simultaneously with a fixed wavelength difference. The method requires approximately 100 μg of DNA to provide the sensitivity of 1 adduct/10^7 nucleotides. However, its ability to distinguish between closely related compounds and determine a single lesion in a complex mixture is quite limited. Moreover, only a few DNA modifications possess intrinsic fluorescent properties (reviewed in (Chang et al., 1994)).

Chemical linkage of fluorescent dyes to DNA adducts followed by capillary electrophoresis and laser-induced fluorescence (CE-LIF) detection is an alternative strategy. The main disadvantages of this approach are the inability to control the efficiency of the labeling procedure and low sensitivity (2 adducts/10^6 nucleotides). In addition, the CE-LIF assay is vulnerable to interference by other fluorescence-emitting substances present in the biological sample (Schmitz et al., 2002).

Mass spectrometry

Mass spectrometry (MS) is one of the most sensitive analytical techniques currently available. It provides structural identification and absolute quantification of DNA lesions at the low-femtomole levels using only a few micrograms of DNA. The method is based on molecular ionization followed by separation of the generated ions and measurement of their abundance and mass-to-charge ratio (m/z). The ions can be subjected to further fragmentation. Sequential MS (MS^n) is accomplished with a triple quadrupole or ion trap mass spectrometers. Each quadrupole (Q) has a separate function: Q1 selects the ion of interest; Q2 dissociates the ions while introducing a collision gas such as argon, helium, or nitrogen; Q3 filters the resulting fragments. The ion trap mass spectrometers work on the same physical principles. The only difference is that the ions are captured and sequentially ejected. It ensures the highest level of selectivity and sensitivity (reviewed in (Kang, 2012)).

Gas chromatography (GC) and liquid chromatography (LC) systems are commonly used to separate and introduce the sample components into a mass spectrometer. The main drawback of the GC-MS approach is that it is limited to volatile/non-polar species, whereas the majority of nucleic acid adducts are non-volatile and/or polar. The analysis requires derivatization, which may be incomplete or result in formation of by-products and artifacts (reviewed in (Farmer et al., 2006)). Development of the electrospray ionization method has solved this problem. It enables transformation of the analyte in gas phase without rupturing any covalent bonds. LC coupled to ESI-MS^n is a great choice for highly polar, least volatile and thermally labile compounds (reviewed in (Singh and Farmer, 2006)).

2. Aims of the study

The main aim of this work was to gain further insight into mutagenic processing of genomic uracil. First, we wanted to address the issue concerning the basal level of uracil in DNA. Previously reported estimations vary by almost three orders of magnitude. This may be due to differences in sample type, but may also result from technical shortcomings of the employed methods. We analyzed potential methodological errors and developed an improved LC-MS/MS-based method for accurate determination and quantification of genomic uracil. This part of the study is presented in paper I.

Next, we wanted to explore whether AID expression in B-cell lymphomas was associated with increased levels of genomic uracil, presumably in the form of U:G mismatches. We aimed to measure genomic uracil in lymphoma and non-lymphoma cell lines and correlate it to APOBECs, uracil DNA glycosylases and dUTPase mRNA and protein levels. In order to provide direct mechanistic evidence of genomic uracil accumulation by AID, we assayed B-cells stably expressing AID-EYFP fusion protein and B-cells undergoing class-switch recombination. Finally, we decided to reanalyze the exome sequencing data from kataegis regions in lymphomas and chronic lymphocytic leukemia. This part of the study is presented in paper II.

In addition to the work on the main project, we collaborated with the DNA repair research group at Biotechnology Centre, University of Oslo. We aimed to reveal interaction partners of uracil-DNA glycosylase SMUG1. The results are presented in paper III.

3. Summary of results

Paper I: **A robust, sensitive assay for genomic uracil determination by LC/MS/MS reveals lower levels than previously reported**

DNA repair (Amst.), 12, 699-706 (2013)

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

* These authors contributed equally to this work

Cytosine deamination and deoxyuridine monophosphate (dUMP) misincorporation are the major sources of uracil (U) in DNA. This potentially mutagenic and cytotoxic lesion is normally eliminated by error-free BER (reviewed in (Krokan et al., 2002)). The error-prone processing of genomic uracil was found to be essential for Ig gene diversification by somatic hypermutation and class switch recombination (reviewed in (Maul et al., 2011)). Moreover, it was reported that AID- and APOBEC3B-catalyzed off-target deamination of cytosine might contribute to tumorigenesis in humans (Burns et al., 2013a; Leonard et al., 2013; Lohr et al., 2012; Robbiani and Nussenzweig, 2013). In this sense, accurate determination and quantification of genomic uracil is of fundamental importance.

A number of GC/MS and LC/MS assays have been developed, but their estimations of the basal uracil levels vary by three orders of magnitude (Mashiyama et al., 2008). We suggest that the discrepancy is a result of methodological errors related to DNA sample preparation and uracil generation during the analysis. In paper I, we discuss these potential issues and present an improved LC/MS/MS-based assay for absolute quantification of genomic deoxyuridine.

Our findings provide convincing evidence that co-purification of intracellular deoxycytidine monophosphate (dCMP) and deoxyuridine monophosphate (dUMP) during DNA isolation and deamination of deoxycytidine (dCyd) during DNA hydrolysis lead to overestimation of genomic uracil. To avoid false results, we introduced an additional clean-up step after DNA isolation, optimized reaction conditions for DNA hydrolysis and included deuracilated DNA control in a work-flow.

Further studies revealed that MS detection of deoxyuridine (dUrd) can be obfuscated by the isobaric and naturally occurring ^{13}C-deoxycytidine (^{13}C-dCyd) (~1.1% of all carbon). To circumvent this problem, we employed a precursory HPLC fractionation step with a reverse-

phase column prior to LC-MS/MS analysis. In addition to dCyd-free samples, it provided a convenient opportunity to quantify the amount of hydrolyzed DNA, thereby allowing precise calculation of dUrd per deoxynucleosides. To correct for variability in processing procedure (e.g. dilution, injection) and possible matrix effects, we utilized an internal standard 2'-deoxyuridine-2-^{13}C,1,3-$^{15}N_2$, which was added to DNA before hydrolysis.

In order to validate the method, we assayed deuracilated salmon sperm DNA spiked with dUrd in sets of six replicates. The samples demonstrated high inter-experimental accuracy (94.3%) and precision (CV 9.7%). The lower limit of quantification was found to be 5 fmol dUrd.

We compared our approach to an alternative method based on detection of U excised from DNA by UNG. The latter showed greater intra-sample variability, which may have been due to an imprecise estimation of the DNA concentration.

To test the assay in a relevant biological context, we quantified genomic dUrd in UNG-proficient and -deficient human and mouse cell lines and found several-fold higher levels in the UNG-deficient cell lines. Importantly, the repair-proficient cell lines contained approximately 400-600 dUrd moieties per genome, which is at least an order of magnitude lower than previously-reported.

We believe that the new method provides the most accurate determination of genomic uracil and is highly relevant to DNA repair-oriented researchers.

Paper II: **High AID expression in B-cell lymphomas causes accumulation of genomic uracil and a distinct AID mutational signature.**

Manuscript submitted September 2014 (DNA repair)

Henrik S. Pettersen, Anastasia Galashevskaya, Berit Doseth, Mirta M. L. Sousa, Antonio Sarno, Torkild Visnes, Per A. Aas, Nina B. Liabakk, Geir Slupphaug, Pål Sætrom, Bodil Kavli, and Hans E. Krokan

Mammalian APOBEC cytidine deaminases are the key players in the mechanisms of the host defence, though its off-target activity is potentially carcinogenic (reviewed in (Conticello, 2008)). A recent large scale genome sequencing study has revealed the APOBEC mutational signatures in 16 out of 30 primary cancer types (Alexandrov et al., 2013). APOBEC3B has been characterized as the likely major source of DNA damage in breast and ovarian cancers (Burns et al., 2013a; Leonard et al., 2013). There have also been many reports showing that AID is associated with lymphomagenesis, but the direct mechanistic evidence has not been provided (Dorsett et al., 2007; Komeno et al., 2010; Kotani et al., 2007; Liu et al., 2008; Ramiro et al., 2006; Robbiani et al., 2008; Robbiani et al., 2009; Shen et al., 1998; Takizawa et al., 2008). As a continuation of the work from paper I, we sought to employ the established LC-MS/MS method for quantitation of genomic uracil and explore whether DNA-cytosine deamination by the APOBEC-family enzyme AID could be a common mutagenic mechanism in B-cell malignancies.

We tested 17 cancer cell lines and found up to 72-fold variation in genomic uracil levels (0.056 [epidermoid carcinoma] vs. 4.03 [Burkitt lymphoma] dUrd per 10^6 deoxyribonucleosides (dNs). We noticed a clear pattern of genomic uracil accumulation in lymphoma cell lines. It averaged 2.5 (0.63-4.0) dUrd per 10^6 dNs, while in non-lymphoma cells and blood donor lymphocytes its quantity was 4.4- and 13.2-fold lower, respectively. This could best be explained by AID-induced cytosine deamination. We therefore investigated expression of all known APOBEC-family genes by means of qRT-PCR, Western analysis and mass spectrometry. Indeed, AID was by far the best predictor of genomic uracil accumulation both at the mRNA ($r^2 = 0.70$, $p < 0.0001$) and protein level ($r^2 = 0.64$, $p = 0.0001$).

Next, we assayed mouse CH12F3 lymphoma B-cells stably expressing AID-EYFP fusion protein or EYFP, and B-cells undergoing class-switch recombination. AID-EYFP expressing cells showed an almost 6-fold increase in genomic uracil levels from 0.14 to 0.83 dUrd per 10^6 dNs compared to EYFP alone. Exposure to TGFβ, IL-4, and anti-CD40 antibodies for 48

hours also led to uracil accumulation, though induction of endogenous AID was considerably lower than in CH12F3-EYFP cell line. High genomic uracil levels could not be explained by increased dUMP misincorporation during replication, because stimulation resulted in a substantial decrease in proliferation rate. Finally, we investigated the effect of AID knockdown using a lentiviral AID shRNA expressing vector in human B-cell lymphoma cell line SUDHL5. A 60% knockdown of AID reduced genomic uracil level by 38% ($p = 0.005$).

We wanted to check whether DNA repair efficacy could determine genomic uracil levels in the studied lymphoma cell lines. To this end, we measured uracil excision capacity of cell free extracts with oligonucleotide cleavage assay (U:G and U:A context) and [^{3}H]uracil-release assay ([^{3}H]U:A substrate). It was negatively correlated with genomic uracil levels ($r^2 = 0.55$, $p = 0.0007$). Furthermore, we found that UNG, SMUG1 and APE1 were negatively correlated with genomic uracil at the protein level, which apparently indicated that the BER capacity was surmounted by the high AID expression.

We reanalyzed the published exome sequencing data from lymphomas and chronic lymphocytic leukemia *kataegis* regions and revealed the target sequence, which overlapped with the known AID hotspot motif WR<u>C</u>Y (W=A or T, R=A or G, C and Y=T or C). The general mutation pattern was ApGp**C**pT, rather than Tp**C**pA/T for the other cancer types with *kataegis*.

These data provide direct mechanistic evidence for AID-induced genomic uracil formation in the development of B-cell malignancies.

Paper III: **The human base excision repair enzyme SMUG1 directly interacts with DKC1 and contributes to RNA quality control**

Molecular Cell 49, 339-345 (2013)

Laure Jobert, Hanne K. Skjeldam, Bjørn Dalhus, Anastasia Galashevskaya, Cathrine B. Vågbø, Magnar Bjørås, and Hilde Nilsen

SMUG1 excises deaminated cytosine, 5-hmU, 5-hU, 5-FU and other oxidized pyrimidines from the DNA (Boorstein et al., 2001; Haushalter et al., 1999; Kavli et al., 2002; Masaoka et al., 2003). Its biochemical properties and substrate specificities are well defined; however, little is known about its potential interaction partners. In paper III, we demonstrate that SMUG1 can interact with the main mammalian pseudouridylase Dyskerin (DKC1) and function in RNA metabolism.

Our observation that SMUG1 co-localizes with DKC1 in nucleoli and Cajal bodies was confirmed by the coimmunoprecipitation (coIP) experiments in DNase I- and RNase A-treated cell extracts. IP of overexpressed SMUG1-EYFP significantly recovered DKC1, while IP of DKC1 pulled down a fraction of endogenous SMUG1. This led us to investigate whether SMUG1 and DKC1 had a direct interaction. To this end, the DKC1 and SMUG1 sequences were synthesized as 20-mer peptides and spotted on cellulose membranes offset by 3-amino acid for further overlay with purified, recombinant glutathione-S-transferase-tagged SMUG1 and DKC1, respectively. We revealed five potential SMUG1-binding sites in DKC1 (amino acids 16-29, 112-122, 247-260, 400-410, 475-491) and two DKC1-binding sequences in SMUG1 (amino acids 25-35, 220-233). These data were used to create a structural model of the DKC1/SMUG1 complex in ZDOCK. From a series of possible solutions we selected an interaction surface involving SMUG1 Glu 231, because it did not interfere with the nucleic acid, NOP-10, GAR1 and NHP2 binding domains. Notably, DKC1 failed to recover SMUG1 E29R/E33R and SMUG1 E231R mutants in glutathione-S-transferase pull-down assay, indicating that SMUG1 amino acids 29, 33, and 231 were required for binding to DKC1.

Our next question was whether SMUG1 could act on RNA. Indeed, wild type SMUG1 excised 5-hm(dUrd) from ssRNA in standard oligonucleotide-nicking experiment, but its activity was approximately two-fold lower than on ssDNA substrate. No cleavage was seen on ssRNA containing Urd, dUrd or pseudoUrd. We established a native RNA-coIP assay to measure the recruitment of SMUG1 to RNA species processed by DKC1. Interestingly, SMUG1 associated with the 47S precursor RNA but not mature 28S, 18S and 5.8S rRNA, nor U2snRNA, nor GAPDH. Moreover, depletion of SMUG1 correlated with the reduced

expression of all three mature rRNAs and accumulation of polyadenilated 28S rRNA, which could indicate the degradation of damaged or inappropriately processed rRNA.

To get more insight into the biological role of SMUG1/DKC1 interaction, we measured the level of 5-hmUrd and pseudouridine (ψUrd) in the 28S and 18S rRNA species by LC-MS/MS. The analysis showed that SMUG1-depleted cells had increased amount of 5-hmUrd. The effect was even more pronounced in cells depleted for both SMUG1 and DKC1. However, we could not confirm that SMUG1 was required for DKC1 activity as the quantity of ψUrd in 28S and 18S rRNAs was reduced only in DKC1 knock-down cells. We therefore suggested that interaction with DKC1 could be important for proper localization of SMUG1 and transfected HeLa cells with SMUG1 E29R/E33R. Unlike SMUG1-EYFP, the mutant was diffused in the nucleoplasm and not enriched in nucleoli and Cajal bodies. The DKC1 staining pattern also changed, indicating that our hypothesis was right.

In conclusion, we propose that the DNA BER enzyme SMUG1 directly interacts with the pseudouridine synthase DKC1 and contributes to rRNA quality control *in vivo*.

4. General discussion

4.1. What is the steady state level of genomic uracil?

Uracil may arise in DNA as a result of cytosine deamination and dUTP misincorporation during replication. The level of dUTP, a necessary precursor for dTTP synthesis in mammals, is normally kept low by deoxyuridine 5'-triphosphate nucleotide-hydrolase (dUTPase) (Wist et al., 1978). The average physiological concentrations of intracellular dUTP and dTTP are 0.2 ± 0.2 and 37 ± 30 μM, respectively (reviewed in (Traut, 1994)). Based on the fact that replicative DNA polymerases do not distinguish well between dUTP and dTTP, it has been calculated that approximately one dUTP residue per 10^4 dTTPs enters DNA (reviewed in (Mosbaugh and Bennett, 1994)). The dUTP misincorporation rate is therefore thought to vary depending on the dUTPase activity and intracellular nucleotide pool imbalances (reviewed in (Olinski et al., 2010)). There are two isoforms of dUTPase in mammals: nuclear and mitochondrial. The nuclear dUTPase is proliferation-dependent, while the mitochondrial isoform is constitutive (Ladner and Caradonna, 1997). It has been shown that, independent of activation state, dUTP concentration is 6-8-fold higher in lymphocytes compared to macrophages, which are terminally differentiated non-dividing cells (reviewed in (Gavegnano et al., 2012)). The steady-state level of genomic uracil determined by alkaline elution of DNA from UNG-deficient murine embryonic fibroblasts was determined to be about 3600 residues per genome, while in hepatocytes the number was approximately 900. It was hypothesized that, in proliferating cells, genomic uracil results mostly from misincorporation during replication (Andersen et al., 2005). Intracellular nucleotide pool imbalances, for example dTTP depletion caused by nutrient folate deficiency or by treatment with anticancer drugs 5-fluorouracil, 5-fluorodeoxyuridine, or methotrexate, can also lead to aberrant uracil incorporation into DNA (reviewed in (Longley et al., 2003; Powers, 2005)).

Spontaneous deamination of genomic cytosine has been suggested to introduce 60-500 uracil residues per cell per day (reviewed in (Krokan et al., 2002)). The reaction rates determined by chemical and biological assays are in reasonable agreement with only 1.2-2.5-fold variation. It is also generally accepted that the deamination rate in dsDNA is 140-300 times lower than in ssDNA. The uncertainty concerning the number of deaminations per day in cells is mainly due to a lack of knowledge on the average size of the fraction of ssDNA in the genome. Assuming that the fraction of ssDNA in human genome at any time is 0.1%, the best estimate

would be ~70-200 spontaneous deamination events per genome per day (reviewed in (Kavli et al., 2007)).

In addition to spontaneous hydrolysis, chemicals such as sodium bisulfite ($NaHSO_3$) and nitrous anhydride (N_2O_3) can induce cytosine deamination. Sodium bisulfite is a common component of many beverages, though it is thought unlikely to be a major contributor to genomic uracil content because of the low concentrations used. Nitrous anhydride is formed upon oxidation of nitric oxide ($NO^{\cdot}$), which is an environmental pollutant, proven mutagen, and also an important physiological messenger (reviewed in (Sousa et al., 2007)). Human lymphoblastoid cells, when subjected to sub-cytotoxic levels of $NO^{\cdot}$ and O_2, display up to a 2-fold increase in genomic uracil (Dong and Dedon, 2006). In line with that, *E. coli* mutants lacking UDG activity are abnormally sensitive to nitrous acid (Daroza et al., 1977). Irradiation by ultraviolet light can also lead to the formation of genomic uracil. Deamination of cytosine in cyclopyrimidine dimers, the major UV photoproducts, has been shown to occur 7-8 times faster than in regular cytosine (Tessman et al., 1992).

Cytosine can be converted to uracil by DNA (cytosine-5)-methyltransferases and cytidine deaminases (reviewed in (Sousa et al., 2007)). Methyltransferases function at CpG islands during epigenetic DNA regulation. These enzymes transfer a methyl group from S-adenosylmethionine (SAM) to cytosine. If SAM concentration is low or if there is a mutation that impairs SAM-binding, a slow deamination reaction may take place instead (reviewed in (Hermann et al., 2004)). Several members of the APOBEC family of cytidine deaminases have recently been shown to deaminate DNA substrates *in vitro*. AID is the most studied. It primarily generates uracil at the *Ig* gene loci during somatic hypermutation and class-switch recombination in germinal center B-cells, but may also target proto-oncogenes (reviewed in (Longerich et al., 2006)).

Estimations of the genomic uracil steady-state level have initially been based on the dUTP pool size and spontaneous cytosine deamination rate (reviewed in (Kavli et al., 2007)). There have been many attempts to measure genomic uracil by indirect means (Andersen et al., 2005; Atamna et al., 2000; Bennett and Kitner, 2006; Horvath and Vertessy, 2010; Lari et al., 2006; Lasken et al., 1996; Maul et al., 2011; Yan et al., 2011). For example, the alkaline comet assay was based on AP-endonuclease treatment of abasic sites and monitoring of the DNA fragment size by alkaline elution (Andersen et al., 2005). The main issue with this approach is that DNA can be shared by high pH and even routine laboratory practices like vortexing (Karpen, 2009; Kohn et al., 1976). Another method employed an aldehyde reactive probe

reaction with the UNG-generated AP sites (Bennett and Kitner, 2006). However, it does not account for spontaneous depurination and depyrimidation, which may occur during sample handling (Lindahl, 1993). The PCR-based strategies suffered from the same drawbacks in addition to being biased to the sequences they target. Furthermore, all these methods are relative, which makes comparisons between experimental batches and laboratories quite difficult.

Direct quantification of genomic uracil can be achieved by mass spectrometry. There are two main approaches, which are in principle very straightforward: detection of uracil excised from DNA by UNG and detection of dUrd after enzymatic hydrolysis of DNA to dNs (Bulgar et al., 2012; Chango et al., 2009; Dong and Dedon, 2006; Dong et al., 2003; Mashiyama et al., 2004; Mashiyama et al., 2008; Ren et al., 2002). However, a wide variation in reported basal levels indicates that reliable quantification of genomic uracil is problematic (Mashiyama et al., 2008).

In paper I, we show that technical shortcomings of the employed methods may result in an overestimation of genomic uracil quantity. Uracil is a heterocyclic molecule, the chemical bonds in which require more energy to break than the *N*-glycosylic bond in dUrd. It results in higher background signal in MS/MS and increases the probability of mistaking contaminants for the analyte. Derivatization has been used to measure uracil, but it adds more complexity and sources for error because biological samples and standards require different reaction conditions (Ren et al., 2002). Also, the degree to which uracil has been derivatized cannot easily be monitored. Hydrophilic interaction chromatography can be a good alternative to derivatization (Bulgar et al., 2012). We tested a similar approach and found that the U excision assay has a comparable sensitivity to measuring dUrd by hydrolysis, but the intra-sample variability was greater. This may be due to an imprecise estimation of the DNA concentration. In this sense, the DNA hydrolysis method is advantageous to uracil excision strategy because dUrd can be normalized to the amount of dNs. However, the existing assays that employ this strategy are not optimal. We noticed that intracellular deoxycytidine monophosphate (dCMP) and/or deoxyuridine monophosphate (dUMP) can co-purify with DNA and be converted to dUrd during hydrolysis. Preliminary treatment with alkaline phosphatase and clean-up with isopropanol precipitation steps reduce the amount of measured dUrd. Deamination of dCyd during sample work-up also represents a major complication with this method. High temperature, acidic/alkali buffer conditions, and DNA denaturation have been shown to accelerate dCyd deamination (Lindahl, 1993; Lindahl and Nyberg, 1974; Shapiro, 1981). We performed DNA hydrolysis at pH 6-6.7 and 37 °C for 50 min using

DNase I, nuclease P1, and alkaline phosphatase. In order to control dCyd deamination we included a deuracilated DNA control. We measured the deamination rate at 37 °C to be 4.805 x 10^{-3} ± 5.9 x 10^{-5} dUrd/10^6bp/min (R^2 = 0.9964, n = 12), which corresponded to 1.059 x 10^{-2} dUrd/10^6 dCyd/min. This is in line with previously reported values of 2.6 x 10^{-2}, 1.2 x 10^{-3}, and 4.8 x 10^{-5} dUrd/10^6 dCyd/min for dNs, single-stranded DNA, and double-stranded DNA, respectively (Shapiro, 1981). MS detection of dUrd can be further complicated by the isobaric and naturally occurring ^{13}C-deoxycytidine (^{13}C-dCyd), which is four orders of magnitude more abundant than dUrd (Dong and Dedon, 2006). To circumvent this problem, we employed a precursory HPLC fractionation step with a reverse-phase column containing weak acidic ion-pairing groups to separate dUrd from dCyd. In addition to dCyd-free samples, it provided a convenient opportunity to quantify the amount of hydrolyzed DNA, thereby allowing the precise calculation of dUrd per dNs. To sum up, we believe that our method provides the most accurate determination of genomic uracil.

We quantified 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 repair-proficient cells contained approximately 400-600 dUrd moieties per genome, which is lower than the quantities reported by other groups. The determined steady state level of genomic uracil is compared to the estimated rates of spontaneous cytosine deamination (~70-200 per mammalian genome per day) (Kavli et al., 2007). It does not indicate that dUTP misincorporation or limited repair efficiency is less abundant. We believe that contribution from different sources of genomic uracil is dynamic and depends on the cell type, proliferation state, fidelity of the DNA replication machinery, mutagen exposure, and DNA repair efficacy. Spontaneous cytosine deamination should be rather a rare event in physiological conditions because it generates mutagenic U:G mismatches. Enzymatic deamination of cytosine is thought to be restricted to variable and switch regions of *Ig*-genes in antigen-stimulated B-cells. However, the APOBEC-induced mutational signatures with U:G as a likely key intermediate have been observed in 16 different primary cancer types (Alexandrov et al., 2013). It may be possible that cytidine deaminases act widely on non-*Ig* loci, many of which are protected by high-fidelity DNA repair (Liu and Schatz, 2009). What mechanisms lead to error-prone repair of U:G mismatches remain to be elucidated.

We found that the steady state level of genomic uracil in human lymphoblastoid cell line derived from hyper-IgM patient carrying *UNG* mutation is almost 11-fold higher than in repair-proficient cells. This indicates that uracil-DNA glycosylases SMUG1, TDG, MBD4,

and the mismatch repair pathway do not provide efficient backup for the lost UNG activity in these cells. UNG has a central role in U:A repair (Otterlei et al., 1999). It is therefore reasonable to conclude that dUMP misincorporation is the main source of genomic uracil in proliferating cells. U:A pairs are not directly mutagenic and may be well tolerated as it has been shown in *Ung*$^{-/-}$ mice (Nilsen et al., 2000). However, abasic sites generated during their repair are potentially mutagenic due to errors in gap-filling in BER (Akbari et al., 2009), or due to translesion synthesis (Auerbach et al., 2006). If the number of U:A pairs and abasic sites is high, they may contribute significantly to mutagenesis.

We have made efforts to improve quantification of genomic uracil by mass spectrometry. Although we cannot discriminate between U:A and U:G contexts, the accurate determination of the global steady state level may be important to better understand the mechanisms of adaptive immunity and oncogenesis.

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

Normally, AID expression is found in activated germinal center B-cells (Muto et al., 2000) and B-cells developing in the bone marrow (Kuraoka et al., 2009). Aberrant AID expression can be induced by cytokines, chemokines, reactive nitrogen and oxygen species, and several types of prostaglandins (Chiba and Marusawa, 2009). It is associated with elevated p53 mutation rate (Takai et al., 2009) and high incidence of inflammation-linked cancers (Endo et al., 2007; Komori et al., 2008; Matsumoto et al., 2007). Mice constitutively expressing AID develop malignant T cell lymphomas and lung micro-adenomas. Tumorogenesis has been characterized by the massive introduction of point mutations in the expressed T cell receptor (*tcr*), *c-myc, pim1, cd4,* and *cd5* genes (Kotani et al., 2005; Okazaki et al., 2003). Further studies of the mouse bone marrow transplantation model have shown that AID overexpression can promote B-cell lymphomagenesis, albeit less frequently (Komeno et al., 2010). The mechanism remains unclear. It is known that AID-induced DNA strand breaks may result in translocations between the *Ig*-locus and proto-oncogenes (Chiarle et al., 2011; Dorsett et al., 2007; Klein et al., 2011; Ramiro et al., 2004; Robbiani et al., 2008; Robbiani et al., 2009; Takizawa et al., 2008). However, a direct mechanistic link between AID and B-cell lymphomagenesis has not been demonstrated.

In paper II, we report that AID mRNA and protein expression levels confer a several-fold accumulation of genomic uracil in lymphoma cell lines. We show that an increase in genomic uracil can be induced by stimulation of class switch recombination and overexpression of

AID-EYFP fusion protein in CH12F3 cell line. AID knockdown results in the opposite effect. High AID expression was observed in a large proportion of B-cell lymphomas. As it has already been discussed, the *Ig*-gene U:G mismatches are the key intermediates in the mechanisms of antibody diversification in the germinal center B-cells. Non-*Ig*-gene U:G base pairs can be processed to DNA double-strand breaks or result in C-to-T transition mutations, which represent a high risk for genomic stability. We found that endogenous AID-induction can increase genomic uracil four-fold, from approximately 750 to 3000 residues per genome. We did not reveal any significant correlation between genomic uracil level and APOBECs other than AID, though a varying expressions of APOBEC3B, APOBEC3D, APOBEC3F, and APOBEC 3G were detected in all examined cancer cell lines. These results suggest that high AID expression is the main cause of genomic uracil accumulation in B-cell lymphomas.

The correlation between AID expression and genomic uracil content is not necessarily straightforward. We noticed two lymphoma cell lines with relatively high uracil levels despite low AID expression (RL and KARPAS-422) and one cell line with moderately increased genomic uracil despite high AID expression (L428). This may be explained by different dUMP misincorporation rate, spontaneous cytosine deamination, uracil repair capacity, and other regulatory mechanisms that affect AID nuclear translocation and targeting to DNA. Furthermore, we suppose that AID can obscure low level contribution of other APOBECs, which may become significant over time.

There was no positive correlation between genomic uracil and cell doubling time in the lymphoma group. Since AID acts during the G_1 phase of the cell cycle, we expected to see higher genomic uracil levels in lymphomas with longer doubling time. In contrast, we found it in non-lymphoma cell lines that did not express AID ($r^2 = 0.57$; $p = 0.048$), which could indicate that dUMP misincorporation is the main source of genomic uracil in rapidly growing non-lymphoma cells. These results are in good agreement with the previously published data (Andersen et al., 2005).

AID-generated uracil is removed by BER and MMR (reviewed in (Maul and Gearhart, 2014)). We focused on BER. In order to examine uracil repair capacity in the studied cell lines, we performed both a uracil-release assay (U:A context) and an oligonucleotide cleavage assay (U:A and U:G context). Notably, the repair capacity was negatively correlated with genomic uracil levels ($r^2 = 0.55$; $p = 0.0007$). UNG was responsible for over 90% of uracil-excision activity. An inverse correlation between UNG, SMUG1, and genomic uracil content could indicate that BER may counteract AID-induced cytosine deamination, but was apparently overridden by the high AID expression levels. It has been shown that *Ung*$^{-/-}$ mice

develop B-cell lymphomas 20 times more frequently than normal controls, but they do not show an increase in other cancer types (Nilsen et al., 2003). We think that BER and MMR compensate for UNG-deficiency in most tissues, but not in B-cells expressing AID.

We studied independent contribution of UNG, SMUG1, and TDG to the *complete repair* of a defined U:G context in a plasmid. To this end, we used nuclear extracts from synchronized HeLa cells and antibodies to uracil-DNA glycosylases. Again, UNG was the most efficient uracil-repair enzyme during the whole cell cycle. SMUG was also active in all cell cycle phases, but contributed less. TDG has previously been shown to act during G_1 phase (Hardeland et al., 2007). Although, we did not reveal significant correlation between TDG and genomic uracil levels, we confirmed that TDG excises uracil in the G_1 phase in the *in vitro* system.

U:G mismatches resulting from enzymatic and spontaneous cytosine deamination may be processed in different ways. It has been reported that transcription factor E2A can regulate the balance between AID and UNG2 at expression and *Ig* targeting levels, stimulate Ig diversification and suppress canonical DNA repair (Wallenius et al., 2014). Furthermore, it has been demonstrated that p53 expression in germinal center B cells is under active repression by BCL6, which helps to tolerate the physiological DNA breaks required for antibody diversification (Phan and Dalla-Favera, 2004). Nevertheless, excision of genomic uracil and generation of AP sites at clustered U:G mispairs may result in double strand breaks, whereas replication over unrepaired U:G mismatches leads in C-to-T transition mutations. The main finding of our study is that AID expression is a predominant source of genomic uracil in B-cell lymphoma cell lines. This is in accordance with the exome sequencing data from lymphomas and chronic lymphocytic leukemia *kataegis* regions, which carry AID-hotspot mutational signatures.

4.3. The human base excision repair enzyme SMUG1 contributes to RNA quality control *in vivo*

In paper III, we show that SMUG1 directly interacts with pseudouridine synthase DKC1, and that SMUG1 and DKC1 co-localize in nucleoli and Cajal bodies, where rRNA biogenesis and non-coding RNA maturation take place, respectively. Our data indicate that interaction with DKC1 could be important for targeting SMUG1 to a select group of RNA substrates. We demonstrate that SMUG1 associates with the 47S precursor rRNA, and that SMUG1 depletion is accompanied by reduced levels of mature 28S, 18S, and 5.8S rRNA and an increased level of polyadenylated 28S rRNA. Hence, aberrant rRNA species accumulate in

absence of SMUG1. These findings suggest that SMUG1 is involved in rRNA quality control. Our observation that SMUG1 has activity on 5-hm(dUrd)-containing ssRNA *in vitro*, and that there is an increased level of 5-hm(Urd) in mature rRNAs isolated from SMUG1-depleted cells further strengthens the possibility of this involvement.

The source of 5-hmUrd in RNA is not known. We hypothesize that it results from hydrolytic deamination of 5-hmC, which is a natural modification in 18S and 28S rRNA in eukaryotes (Racz et al., 1978). We also think that hmU can be misincorporated in RNA, as previously suggested (Pettersen et al., 2011).

The currently available data indicate that hmU-excision by SMUG1 may initiate RNA degradation. It has been recently demonstrated that DNA repair protein APE1 associates with 47S, 28S, and 18S rRNA, and the ribosome processing protein nucleoplasmin (NPM1) within nucleoli (Vascotto et al., 2009). We therefore speculate whether APE1 could process the AP site generated by SMUG1. In this case, APE1 would create a 3'-OH terminus, which may then be targeted by Ccr4-Not, TRAMP, or the exosome for further degradation.

Functions of DNA repair proteins are more versatile than we originally thought. Several studies point to their role in RNA metabolism and RNA surveillance pathways (reviewed in (Jobert and Nilsen, 2014)). It seems reasonable that the ability of DNA glycosylases to recognize subtle chemical modifications contributes to identification of damaged or inappropriately processed RNA species and preserves genetic stability not only through the DNA repair. However, RNA degradation can be triggered by many different types of aberrations. It warrants further studies into the mechanisms of RNA quality control.

References

Aas, P.A., Otterlei, M., Falnes, P.O., Vagbo, C.B., Skorpen, F., Akbari, M., Sundheim, O., Bjoras, M., Slupphaug, G., Seeberg, E., *et al.* (2003). Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. Nature *421*, 859-863.

Abbotts, R., and Madhusudan, S. (2010). Human AP endonuclease 1 (APE1): from mechanistic insights to druggable target in cancer. Cancer Treat Rev *36*, 425-435.

Agris, P.F. (2004). Decoding the genome: a modified view. Nucleic Acids Res *32*, 223-238.

Akbari, M., Otterlei, M., Pena-Diaz, J., Aas, P.A., Kavli, B., Liabakk, N.B., Hagen, L., Imai, K., Durandy, A., Slupphaug, G., *et al.* (2004). 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*, 5486-5498.

Akbari, M., Otterlei, M., Pena-Diaz, J., and Krokan, H.E. (2007). Different organization of base excision repair of uracil in DNA in nuclei and mitochondria and selective upregulation of mitochondrial uracil-DNA glycosylase after oxidative stress. Neuroscience *145*, 1201-1212.

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

Akbari, M., Visnes, T., Krokan, H.E., and Otterlei, M. (2008). Mitochondrial base excision repair of uracil and AP sites takes place by single-nucleotide insertion and long-patch DNA synthesis. DNA Repair (Amst) *7*, 605-616.

Albertson, T.M., Ogawa, M., Bugni, J.M., Hays, L.E., Chen, Y., Wang, Y., Treuting, P.M., Heddle, J.A., Goldsby, R.E., and Preston, B.D. (2009). DNA polymerase epsilon and delta proofreading suppress discrete mutator and cancer phenotypes in mice. Proc Natl Acad Sci U S A *106*, 17101-17104.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J., and Stratton, M.R. (2013). Deciphering Signatures of Mutational Processes Operative in Human Cancer. Cell Rep *3*, 246-259.

Almeida, K.H., and Sobol, R.W. (2007). A unified view of base excision repair: lesion-dependent protein complexes regulated by post-translational modification. DNA Repair (Amst) *6*, 695-711.

Amitsur, M., Levitz, R., and Kaufmann, G. (1987). Bacteriophage T4 anticodon nuclease, polynucleotide kinase and RNA ligase reprocess the host lysine tRNA. EMBO J *6*, 2499-2503.

Andersen, S., Heine, T., Sneve, R., Konig, I., Krokan, H.E., Epe, B., and Nilsen, H. (2005). Incorporation of dUMP into DNA is a major source of spontaneous DNA damage, while excision of uracil is not required for cytotoxicity of fluoropyrimidines in mouse embryonic fibroblasts. Carcinogenesis *26*, 547-555.

Anderson, B.R., Muramatsu, H., Nallagatla, S.R., Bevilacqua, P.C., Sansing, L.H., Weissman, D., and Kariko, K. (2010). Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation. Nucleic Acids Res *38*, 5884-5892.

Aoufouchi, S., Faili, A., Zober, C., D'Orlando, O., Weller, S., Weill, J.C., and Reynaud, C.A. (2008). Proteasomal degradation restricts the nuclear lifespan of AID. J Exp Med *205*, 1357-1368.

Arakawa, H., Moldovan, G.L., Saribasak, H., Saribasak, N.N., Jentsch, S., and Buerstedde, J.M. (2006). A role for PCNA ubiquitination in immunoglobulin hypermutation. PLoS Biol *4*, e366.

Armitage, A.E., Katzourakis, A., de Oliveira, T., Welch, J.J., Belshaw, R., Bishop, K.N., Kramer, B., McMichael, A.J., Rambaut, A., and Iversen, A.K. (2008). Conserved footprints of APOBEC3G on

Hypermutated human immunodeficiency virus type 1 and human endogenous retrovirus HERV-K(HML2) sequences. J Virol *82*, 8743-8761.

Atamna, H., Cheung, I., and Ames, B.N. (2000). A method for detecting abasic sites in living cells: age-dependent changes in base excision repair. Proc Natl Acad Sci U S A *97*, 686-691.

Auerbach, P., Bennett, R.A.O., Bailey, E.A., Krokan, H.E., and Demple, B. (2006). Mutagenic specificity of endogenously generated abasic sites in Saccharomyces cerevisiae chromosomal DNA (vol 102, pg 17711, 2005). P Natl Acad Sci USA *103*, 4328-4328.

Baba, D., Maita, N., Jee, J.G., Uchimura, Y., Saitoh, H., Sugasawa, K., Hanaoka, F., Tochio, H., Hiroaki, H., and Shirakawa, M. (2005). Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. Nature *435*, 979-982.

Bardwell, P.D., Woo, C.J., Wei, K., Li, Z., Martin, A., Sack, S.Z., Parris, T., Edelmann, W., and Scharff, M.D. (2004). Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. Nat Immunol *5*, 224-229.

Barnes, T., Kim, W.C., Mantha, A.K., Kim, S.E., Izumi, T., Mitra, S., and Lee, C.H. (2009). Identification of Apurinic/apyrimidinic endonuclease 1 (APE1) as the endoribonuclease that cleaves c-myc mRNA. Nucleic Acids Res *37*, 3946-3958.

Bartsch, H., and Nair, J. (2005). Accumulation of lipid peroxidation-derived DNA lesions: potential lead markers for chemoprevention of inflammation-driven malignancies. Mutat Res *591*, 34-44.

Basu, U., Chaudhuri, J., Alpert, C., Dutt, S., Ranganath, S., Li, G., Schrum, J.P., Manis, J.P., and Alt, F.W. (2005). The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. Nature *438*, 508-511.

Basu, U., Meng, F.L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C.R., Wesemann, D.R.*, et al.* (2011). The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. Cell *144*, 353-363.

Beale, R.C., Petersen-Mahrt, S.K., Watt, I.N., Harris, R.S., Rada, C., and Neuberger, M.S. (2004). Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. J Mol Biol *337*, 585-596.

Beard, W.A., and Wilson, S.H. (2006). Structure and mechanism of DNA polymerase Beta. Chem Rev *106*, 361-382.

Bedlinskiy, I., Kubarovsky, V., Niccolai, S., Stoler, P., Adhikari, K.P., Aghasyan, M., Amaryan, M.J., Anghinolfi, M., Avakian, H., Baghdasaryan, H.*, et al.* (2012). Measurement of exclusive pi(0) electroproduction structure functions and their relationship to transverse generalized parton distributions. Phys Rev Lett *109*, 112001.

Bellacosa, A. (2001). Role of MED1 (MBD4) Gene in DNA repair and human cancer. J Cell Physiol *187*, 137-144.

Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A.T., Matsumoto, Y., Golemis, E.A., Genuardi, M., and Neri, G. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. Proc Natl Acad Sci U S A *96*, 3969-3974.

Bennett, S.E., and Kitner, J. (2006). Characterization of the aldehyde reactive probe reaction with AP-sites in DNA: influence of AP-lyase on adduct stability. Nucleosides Nucleotides Nucleic Acids *25*, 823-842.

Bentley, D.J., Harrison, C., Ketchen, A.M., Redhead, N.J., Samuel, K., Waterfall, M., Ansell, J.D., and Melton, D.W. (2002). DNA ligase I null mouse cells show normal DNA repair activity but altered DNA replication and reduced genome stability. J Cell Sci *115*, 1551-1561.

Bhakat, K.K., Izumi, T., Yang, S.H., Hazra, T.K., and Mitra, S. (2003). Role of acetylated human AP-endonuclease (APE1/Ref-1) in regulation of the parathyroid hormone gene. EMBO J *22*, 6299-6309.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042-1047.

Bishop, K.N., Holmes, R.K., Sheehy, A.M., Davidson, N.O., Cho, S.J., and Malim, M.H. (2004). Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Curr Biol *14*, 1392-1396.

Blanc, V., and Davidson, N.O. (2010). APOBEC-1-mediated RNA editing. Wiley Interdiscip Rev Syst Biol Med *2*, 594-602.

Boland, C.R., and Goel, A. (2010). Microsatellite instability in colorectal cancer. Gastroenterology *138*, 2073-2087 e2073.

Boorstein, R.J., Cummings, A., Jr., 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.

Botling, J., Edlund, K., Lohr, M., Hellwig, B., Holmberg, L., Lambe, M., Berglund, A., Ekman, S., Bergqvist, M., Ponten, F.*, et al.* (2013). Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. Clin Cancer Res *19*, 194-204.

Bourajjaj, M., Stehouwer, C.D., van Hinsbergh, V.W., and Schalkwijk, C.G. (2003). Role of methylglyoxal adducts in the development of vascular complications in diabetes mellitus. Biochem Soc Trans *31*, 1400-1402.

Brandsma, I., and Gent, D.C. (2012). Pathway choice in DNA double strand break repair: observations of a balancing act. Genome Integr *3*, 9.

Brar, S.S., Watson, M., and Diaz, M. (2004). Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. J Biol Chem *279*, 26395-26401.

Brogna, S., and Wen, J. (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. Nat Struct Mol Biol *16*, 107-113.

Brooks, S.C., Adhikary, S., Rubinson, E.H., and Eichman, B.F. (2013). Recent advances in the structural mechanisms of DNA glycosylases. Bba-Proteins Proteom *1834*, 247-271.

Brown, K. (2012). Methods for the detection of DNA adducts. Methods Mol Biol *817*, 207-230.

Bulgar, A.D., Weeks, L.D., Miao, Y., Yang, S., Xu, Y., Guo, C., Markowitz, S., Oleinick, N., Gerson, S.L., and Liu, L. (2012). Removal of uracil by uracil DNA glycosylase limits pemetrexed cytotoxicity: overriding the limit with methoxyamine to inhibit base excision repair. Cell Death Dis *3*, e252.

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.

Cabelof, D.C., Ikeno, Y., Nyska, A., Busuttil, R.A., Anyangwe, N., Vijg, J., Matherly, L.H., Tucker, J.D., Wilson, S.H., Richardson, A.*, et al.* (2006). Haploinsufficiency in DNA polymerase beta increases cancer risk with age and alters mortality rate. Cancer Res *66*, 7460-7465.

Caldecott, K.W. (2003). XRCC1 and DNA strand break repair. DNA Repair (Amst) *2*, 955-969.

Caldecott, K.W. (2008). Single-strand break repair and genetic disease. Nat Rev Genet *9*, 619-631.

Caldecott, K.W., Aoufouchi, S., Johnson, P., and Shall, S. (1996). XRCC1 polypeptide interacts with DNA polymerase beta and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' in vitro. Nucleic Acids Res *24*, 4387-4394.

Caldecott, K.W., McKeown, C.K., Tucker, J.D., Ljungquist, S., and Thompson, L.H. (1994). An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. Mol Cell Biol *14*, 68-76.

Cantara, W.A., Crain, P.F., Rozenski, J., McCloskey, J.A., Harris, K.A., Zhang, X., Vendeix, F.A., Fabris, D., and Agris, P.F. (2011). The RNA Modification Database, RNAMDB: 2011 update. Nucleic Acids Res *39*, D195-201.

Casellas, R., Yamane, A., Kovalchuk, A.L., and Potter, M. (2009). Restricting activation-induced cytidine deaminase tumorigenic activity in B lymphocytes. Immunology *126*, 316-328.

Chahwan, R., Edelmann, W., Scharff, M.D., and Roa, S. (2012). AIDing antibody diversity by error-prone mismatch repair. Semin Immunol *24*, 293-300.

Chang, D.J., and Cimprich, K.A. (2009). DNA damage tolerance: when it's OK to make mistakes. Nat Chem Biol *5*, 82-90.

Chang, L.W., Hsia, S.M., Chan, P.C., and Hsieh, L.L. (1994). Macromolecular adducts: biomarkers for toxicity and carcinogenesis. Annu Rev Pharmacol Toxicol *34*, 41-67.

Chango, A., Abdel Nour, A.M., Niquet, C., and Tessier, F.J. (2009). Simultaneous determination of genomic DNA methylation and uracil misincorporation. Med Princ Pract *18*, 81-84.

Chapman, J.R., Taylor, M.R., and Boulton, S.J. (2012). Playing the end game: DNA double-strand break repair pathway choice. Mol Cell *47*, 497-510.

Chattopadhyay, R., Wiederhold, L., Szczesny, B., Boldogh, I., Hazra, T.K., Izumi, T., and Mitra, S. (2006). Identification and characterization of mitochondrial abasic (AP)-endonuclease in mammalian cells. Nucleic Acids Res *34*, 2067-2076.

Chaudhuri, J., Basu, U., Zarrin, A., Yan, C., Franco, S., Perlot, T., Vuong, B., Wang, J., Phan, R.T., Datta, A.*, et al.* (2007). Evolution of the immunoglobulin heavy chain class switch recombination mechanism. Adv Immunol *94*, 157-214.

Chaudhuri, J., Khuong, C., and Alt, F.W. (2004). Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. Nature *430*, 992-998.

Chen, J.M., Ferec, C., and Cooper, D.N. (2012). Transient hypermutability, chromothripsis and replication-based mechanisms in the generation of concurrent clustered mutations. Mutat Res-Rev Mutat *750*, 52-59.

Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J.*, et al.* (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell *147*, 107-119.

Chiba, T., and Marusawa, H. (2009). A novel mechanism for inflammation-associated carcinogenesis; an important role of activation-induced cytidine deaminase (AID) in mutation induction. J Mol Med (Berl) *87*, 1023-1027.

Chiu, Y.L., Soros, V.B., Kreisberg, J.F., Stopak, K., Yonemoto, W., and Greene, W.C. (2005). Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells. Nature *435*, 108-114.

Ciccia, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. Mol Cell *40*, 179-204.

Cleaver, J.E., Lam, E.T., and Revet, I. (2009). Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. Nat Rev Genet *10*, 756-768.

Collins, A.R., Cadet, J., Moller, L., Poulsen, H.E., and Vina, J. (2004). Are we sure we know how to measure 8-oxo-7,8-dihydroguanine in DNA from human cells? Arch Biochem Biophys *423*, 57-65.

Conticello, S.G. (2008). The AID/APOBEC family of nucleic acid mutators. Genome Biol *9*, 229.

Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F.*, et al.* (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419-423.

Daffis, S., Szretter, K.J., Schriewer, J., Li, J., Youn, S., Errett, J., Lin, T.Y., Schneller, S., Zust, R., Dong, H.*, et al.* (2010). 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. Nature *468*, 452-456.

Dalhus, B., Laerdahl, J.K., Backe, P.H., and Bjoras, M. (2009). DNA base repair--recognition and initiation of catalysis. FEMS Microbiol Rev *33*, 1044-1078.

Daroza, R., Friedberg, E.C., Duncan, B.K., and Warner, H.R. (1977). Enzymatic Degradation of Uracil-Containing DNA .4. Repair of Nitrous-Acid Damage to DNA in Escherichia-Coli. Biochemistry *16*, 4934-4939.

De Bont, R., and van Larebeke, N. (2004). Endogenous DNA damage in humans: a review of quantitative data. Mutagenesis *19*, 169-185.

De Vos, M., Schreiber, V., and Dantzer, F. (2012). The diverse roles and clinical relevance of PARPs in DNA damage repair: current state of the art. Biochem Pharmacol *84*, 137-146.

Dedeoglu, F., Horwitz, B., Chaudhuri, J., Alt, F.W., and Geha, R.S. (2004). Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB. Int Immunol *16*, 395-404.

Di Noia, J.M., and Neuberger, M.S. (2007). Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem *76*, 1-22.

Di Noia, J.M., Rada, C., and Neuberger, M.S. (2006). SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. Embo J *25*, 585-595.

Dimitriadis, E.K., Prasad, R., Vaske, M.K., Chen, L., Tomkinson, A.E., Lewis, M.S., and Wilson, S.H. (1998). Thermodynamics of human DNA ligase I trimerization and association with DNA polymerase beta. J Biol Chem *273*, 20540-20550.

Dizdaroglu, M., Karakaya, A., Jaruga, P., Slupphaug, G., and Krokan, H.E. (1996). Novel activities of human uracil DNA N-glycosylase for cytosine-derived products of oxidative DNA damage. Nucleic Acids Res *24*, 418-422.

Dong, M., and Dedon, P.C. (2006). 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*, 50-57.

Dong, M., Wang, C., Deen, W.M., and Dedon, P.C. (2003). Absence of 2'-deoxyoxanosine and presence of abasic sites in DNA exposed to nitric oxide at controlled physiological concentrations. Chem Res Toxicol *16*, 1044-1055.

Dorsett, Y., McBride, K.M., Jankovic, M., Gazumyan, A., Thai, T.H., Robbiani, D.F., Di Virgilio, M., Reina San-Martin, B., Heidkamp, G., Schwickert, T.A*., et al.* (2008). MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. Immunity *28*, 630-638.

Dorsett, Y., Robbiani, D.F., Jankovic, M., Reina-San-Martin, B., Eisenreich, T.R., and Nussenzweig, M.C. (2007). A role for AID in chromosome translocations between c-myc and the IgH variable region. J Exp Med *204*, 2225-2232.

Drablos, F., Feyzi, E., Aas, P.A., Vaagbo, C.B., Kavli, B., Bratlie, M.S., Pena-Diaz, J., Otterlei, M., Slupphaug, G., and Krokan, H.E. (2004). Alkylation damage in DNA and RNA--repair mechanisms and medical significance. DNA Repair (Amst) *3*, 1389-1407.

Ehrenstein, M.R., and Neuberger, M.S. (1999). Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. EMBO J *18*, 3484-3490.

Ellery, P.J., Tippett, E., Chiu, Y.L., Paukovics, G., Cameron, P.U., Solomon, A., Lewin, S.R., Gorry, P.R., Jaworowski, A., Greene, W.C*., et al.* (2007). The CD16+ monocyte subset is more permissive to infection and preferentially harbors HIV-1 in vivo. J Immunol *178*, 6581-6589.

Endo, Y., Marusawa, H., Kinoshita, K., Morisawa, T., Sakurai, T., Okazaki, I.M., Watashi, K., Shimotohno, K., Honjo, T., and Chiba, T. (2007). Expression of activation-induced cytidine deaminase in human hepatocytes via NF-kappaB signaling. Oncogene *26*, 5587-5595.

Engel, N., Tront, J.S., Erinle, T., Nguyen, N., Latham, K.E., Sapienza, C., Hoffman, B., and Liebermann, D.A. (2009). Conserved DNA methylation in Gadd45a(-/-) mice. Epigenetics *4*, 98-99.

Evans, M.D., Dizdaroglu, M., and Cooke, M.S. (2004). Oxidative DNA damage and disease: induction, repair and significance. Mutat Res *567*, 1-61.

Farmer, P.B., Phillips, D., Moller, L., Singh, R., van Schooten, F.-J., Godschalk, R., Mateuca, R., and Kirsch-Volders, M. (2006). Sensitivity and specificity of techniques for the identification of biomarkers. In (The Nofer Institute of Occupational Medicine).

Ferrari, G., Rossi, R., Arosio, D., Vindigni, A., Biamonti, G., and Montecucco, A. (2003). Cell cycle-dependent phosphorylation of human DNA ligase I at the cyclin-dependent kinase sites. J Biol Chem *278*, 37761-37767.

Fortini, P., and Dogliotti, E. (2007). Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways. DNA Repair (Amst) *6*, 398-409.

Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., and Ellenberger, T. (2006). DNA repair and mutagenesis, 2nd edn (Washington, D.C., ASM Press).

Frieder, D., Larijani, M., Tang, E., Parsa, J.Y., Basit, W., and Martin, A. (2006). Antibody diversification: mutational mechanisms and oncogenesis. Immunol Res *35*, 75-88.

Friedman, J.I., and Stivers, J.T. (2010). Detection of damaged DNA bases by DNA glycosylase enzymes. Biochemistry *49*, 4957-4967.

Fu, D., Calvo, J.A., and Samson, L.D. (2012). Balancing repair and tolerance of DNA damage caused by alkylating agents. Nature Reviews Cancer *12*, 104-120.

Fukui, K. (2010). DNA mismatch repair in eukaryotes and bacteria. J Nucleic Acids *2010*.

Fung, H., Bennett, R.A., and Demple, B. (2001). Key role of a downstream specificity protein 1 site in cell cycle-regulated transcription of the AP endonuclease gene APE1/APEX in NIH3T3 cells. J Biol Chem *276*, 42011-42017.

Fung, H., and Demple, B. (2005). A vital role for Ape1/Ref1 protein in repairing spontaneous DNA damage in human cells. Mol Cell *17*, 463-470.

Gao, Y., Katyal, S., Lee, Y., Zhao, J., Rehg, J.E., Russell, H.R., and McKinnon, P.J. (2011). DNA ligase III is critical for mtDNA integrity but not Xrcc1-mediated nuclear DNA repair. Nature *471*, 240-244.

Gavegnano, C., Kennedy, E.M., Kim, B., and Schinazi, R.F. (2012). The Impact of Macrophage Nucleotide Pools on HIV-1 Reverse Transcription, Viral Replication, and the Development of Novel Antiviral Agents. Mol Biol Int *2012*, 625983.

Gazumyan, A., Bothmer, A., Klein, I.A., Nussenzweig, M.C., and McBride, K.M. (2012). Activation-induced cytidine deaminase in antibody diversification and chromosome translocation. Adv Cancer Res *113*, 167-190.

Gazumyan, A., Timachova, K., Yuen, G., Siden, E., Di Virgilio, M., Woo, E.M., Chait, B.T., Reina San-Martin, B., Nussenzweig, M.C., and McBride, K.M. (2011). Amino-terminal phosphorylation of activation-induced cytidine deaminase suppresses c-myc/IgH translocation. Mol Cell Biol *31*, 442-449.

Gilljam, K.M., Feyzi, E., Aas, P.A., Sousa, M.M., Muller, R., Vagbo, C.B., Catterall, T.C., Liabakk, N.B., Slupphaug, G., Drablos, F.*, et al.* (2009). Identification of a novel, widespread, and functionally important PCNA-binding motif. J Cell Biol *186*, 645-654.

Gonda, H., Sugai, M., Nambu, Y., Katakai, T., Agata, Y., Mori, K.J., Yokota, Y., and Shimizu, A. (2003). The balance between Pax5 and Id2 activities is the key to AID gene expression. J Exp Med *198*, 1427-1437.

Gourzi, P., Leonova, T., and Papavasiliou, F.N. (2007). Viral induction of AID is independent of the interferon and the Toll-like receptor signaling pathways but requires NF-kappaB. J Exp Med *204*, 259-265.

Gu, H., Marth, J.D., Orban, P.C., Mossmann, H., and Rajewsky, K. (1994). Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. Science *265*, 103-106.

Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.F., Liu, W., Xie, Z.G., Shi, L., He, X., Jin, S.G.*, et al.* (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606-610.

Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell *145*, 423-434.

Hagen, L., Kavli, B., Sousa, M.M., Torseth, K., Liabakk, N.B., Sundheim, O., Pena-Diaz, J., Otterlei, M., Horning, O., Jensen, O.N.*, et al.* (2008). Cell cycle-specific UNG2 phosphorylations regulate protein turnover, activity and association with RPA. EMBO J *27*, 51-61.

Hakim, O., Resch, W., Yamane, A., Klein, I., Kieffer-Kwon, K.R., Jankovic, M., Oliveira, T., Bothmer, A., Voss, T.C., Ansarah-Sobrinho, C.*, et al.* (2012). DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. Nature *484*, 69-74.

Hanssen-Bauer, A., Solvang-Garten, K., Akbari, M., and Otterlei, M. (2012). X-ray repair cross complementing protein 1 in base excision repair. Int J Mol Sci *13*, 17210-17229.

Hardeland, U., Bentele, M., Lettieri, T., Steinacher, R., Jiricny, J., and Schar, P. (2001). Thymine DNA glycosylase. Prog Nucleic Acid Res Mol Biol *68*, 235-253.

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

Harper, J.W., and Elledge, S.J. (2007). The DNA damage response: ten years after. Mol Cell *28*, 739-745.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., and Malim, M.H. (2003). DNA deamination mediates innate immunity to retroviral infection. Cell *113*, 803-809.

Harris, R.S., and Liddament, M.T. (2004). Retroviral restriction by APOBEC proteins. Nat Rev Immunol *4*, 868-877.

Harris, R.S., Petersen-Mahrt, S.K., and Neuberger, M.S. (2002). RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. Mol Cell *10*, 1247-1253.

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

Hasler, J., Rada, C., and Neuberger, M.S. (2011). Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1alpha (eEF1A). Proc Natl Acad Sci U S A *108*, 18366-18371.

Haug, T., Skorpen, F., Aas, P.A., Malm, V., Skjelbred, C., and Krokan, H.E. (1998). Regulation of expression of nuclear and mitochondrial forms of human uracil-DNA glycosylase. Nucleic Acids Res *26*, 1449-1457.

Haug, T., Skorpen, F., Kvaloy, K., Eftedal, I., Lund, H., and Krokan, H.E. (1996). Human uracil-DNA glycosylase gene: sequence organization, methylation pattern, and mapping to chromosome 12q23-q24.1. Genomics *36*, 408-416.

Haushalter, K.A., Todd Stukenberg, M.W., Kirschner, M.W., and Verdine, G.L. (1999). Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. Curr Biol *9*, 174-185.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L.*, et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303-1307.

Helm, M. (2006). Post-transcriptional nucleotide modification and alternative folding of RNA. Nucleic Acids Res *34*, 721-733.

Hendrich, B., Hardeland, U., Ng, H.H., Jiricny, J., and Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature *401*, 301-304.

Henneke, G., Koundrioukoff, S., and Hubscher, U. (2003). Phosphorylation of human Fen1 by cyclin-dependent kinase modulates its role in replication fork regulation. Oncogene *22*, 4301-4313.

Hermann, A., Gowher, H., and Jeltsch, A. (2004). Biochemistry and biology of mammalian DNA methyltransferases. Cell Mol Life Sci *61*, 2571-2587.

Hidaka, K., Yamada, M., Kamiya, H., Masutani, C., Harashima, H., Hanaoka, F., and Nohmi, T. (2008). Specificity of mutations induced by incorporation of oxidized dNTPs into DNA by human DNA polymerase eta. DNA Repair (Amst) *7*, 497-506.

Horvath, A., and Vertessy, B.G. (2010). A one-step method for quantitative determination of uracil in DNA by real-time PCR. Nucleic Acids Res *38*, e196.

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

Hudson, T.J., Anderson, W., Artez, A., Barker, A.D., Bell, C., Bernabe, R.R., Bhan, M.K., Calvo, F., Eerola, I., Gerhard, D.S.*, et al.* (2010). International network of cancer genome projects. Nature *464*, 993-998.

Imai, K., Slupphaug, G., Lee, W.I., Revy, P., Nonoyama, S., Catalan, N., Yel, L., Forveille, M., Kavli, B., Krokan, H.E.*, et al.* (2003). Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. Nature Immunology *4*, 1023-1028.

Ito, S., Nagaoka, H., Shinkura, R., Begum, N., Muramatsu, M., Nakata, M., and Honjo, T. (2004). Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. Proc Natl Acad Sci U S A *101*, 1975-1980.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300-1303.

Izzotti, A., Cartiglia, C., Taningher, M., De Flora, S., and Balansky, R. (1999). Age-related increases of 8-hydroxy-2'-deoxyguanosine and DNA-protein crosslinks in mouse organs. Mutat Res *446*, 215-223.

Jacobs, A.L., and Schar, P. (2012). DNA glycosylases: in DNA repair and beyond. Chromosoma *121*, 1-20.

Jia, G., Fu, Y., and He, C. (2013). Reversible RNA adenosine methylation in biological regulation. Trends Genet *29*, 108-115.

Jobert, L., and Nilsen, H. (2014). Regulatory mechanisms of RNA function: emerging roles of DNA repair enzymes. Cell Mol Life Sci.

Kang, J.-S. (2012). Principles and Applications of LC-MS/MS for the Quantitative Bioanalysis of Analytes in Various Biological Samples. In Tandem Mass Spectrometry - Applications and Principles D.J. Prasain, ed.

Karpen, G.H. (2009). Preparation of high-molecular-weight DNA from Drosophila embryos. Cold Spring Harb Protoc *2009*, pdb prot5254.

Katyal, S., and McKinnon, P.J. (2008). DNA strand breaks, neurodegeneration and aging in the brain. Mech Ageing Dev *129*, 483-491.

Kavli, B., Otterlei, M., Slupphaug, G., and Krokan, H.E. (2007). Uracil in DNA--general mutagen, but normal intermediate in acquired immunity. DNA Repair (Amst) *6*, 505-516.

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.

Kelley, M.R., Georgiadis, M.M., and Fishel, M.L. (2012). APE1/Ref-1 role in redox signaling: translational applications of targeting the redox function of the DNA repair/redox protein APE1/Ref-1. Curr Mol Pharmacol *5*, 36-53.

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.

Khodyreva, S.N., Prasad, R., Ilina, E.S., Sukhanova, M.V., Kutuzov, M.M., Liu, Y., Hou, E.W., Wilson, S.H., and Lavrik, O.I. (2010). Apurinic/apyrimidinic (AP) site recognition by the 5'-dRP/AP lyase in poly(ADP-ribose) polymerase-1 (PARP-1). Proc Natl Acad Sci U S A *107*, 22090-22095.

Kim, W.C., King, D., and Lee, C.H. (2010). RNA-cleaving properties of human apurinic/apyrimidinic endonuclease 1 (APE1). Int J Biochem Mol Biol *1*, 12-25.

Kim, Y.J., and Wilson, D.M., 3rd (2012). Overview of base excision repair biochemistry. Curr Mol Pharmacol *5*, 3-13.

Klaene, J.J., Sharma, V.K., Glick, J., and Vouros, P. (2013). The analysis of DNA adducts: The transition from P-32-postlabeling to mass spectrometry. Cancer Lett *334*, 10-19.

Klauer, A.A., and van Hoof, A. (2012). Degradation of mRNAs that lack a stop codon: a decade of nonstop progress. Wiley Interdiscip Rev RNA *3*, 649-660.

Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F.*, et al.* (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell *147*, 95-106.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R.*, et al.* (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature *468*, 839-843.

Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G.*, et al.* (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell *8*, 200-213.

Kohn, K.W., Erickson, L.C., Ewig, R.A., and Friedman, C.A. (1976). Fractionation of DNA from mammalian cells by alkaline elution. Biochemistry *15*, 4629-4637.

Komeno, Y., Kitaura, J., Watanabe-Okochi, N., Kato, N., Oki, T., Nakahara, F., Harada, Y., Harada, H., Shinkura, R., Nagaoka, H.*, et al.* (2010). AID-induced T-lymphoma or B-leukemia/lymphoma in a mouse BMT model. Leukemia *24*, 1018-1024.

Komori, J., Marusawa, H., Machimoto, T., Endo, Y., Kinoshita, K., Kou, T., Haga, H., Ikai, I., Uemoto, S., and Chiba, T. (2008). Activation-induced cytidine deaminase links bile duct inflammation to human cholangiocarcinoma. Hepatology *47*, 888-896.

Kondo, E., Gu, Z., Horii, A., and Fukushige, S. (2005). The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. Mol Cell Biol *25*, 4388-4396.

Kondo, N., Takahashi, A., Ono, K., and Ohnishi, T. (2010). DNA damage induced by alkylating agents and repair pathways. J Nucleic Acids *2010*, 543531.

Kong, Q., and Lin, C.L. (2010). Oxidative damage to RNA: mechanisms, consequences, and diseases. Cell Mol Life Sci *67*, 1817-1829.

Kotani, A., Kakazu, N., Tsuruyama, T., Okazaki, I.M., Muramatsu, M., Kinoshita, K., Nagaoka, H., Yabe, D., and Honjo, T. (2007). Activation-induced cytidine deaminase (AID) promotes B cell lymphomagenesis in Emu-cmyc transgenic mice. Proc Natl Acad Sci U S A *104*, 1616-1620.

Kotani, A., Okazaki, I.M., Muramatsu, M., Kinoshita, K., Begum, N.A., Nakajima, T., Saito, H., and Honjo, T. (2005). A target selection of somatic hypermutations is regulated similarly between T and B cells upon activation-induced cytidine deaminase expression. Proc Natl Acad Sci U S A *102*, 4506-4511.

Krejci, L., Altmannova, V., Spirek, M., and Zhao, X. (2012). Homologous recombination and its regulation. Nucleic Acids Res *40*, 5795-5818.

Krishnakumar, R., and Kraus, W.L. (2010). The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. Mol Cell *39*, 8-24.

Krokan, H.E., and Bjoras, M. (2013). Base excision repair. Cold Spring Harb Perspect Biol *5*, a012583.

Krokan, H.E., Drablos, F., and Slupphaug, G. (2002). Uracil in DNA--occurrence, consequences and repair. Oncogene *21*, 8935-8948.

Krokan, H.E., Standal, R., and Slupphaug, G. (1997). DNA glycosylases in the base excision repair of DNA. Biochem J *325 (Pt 1)*, 1-16.

Kryston, T.B., Georgiev, A.B., Pissis, P., and Georgakilas, A.G. (2011). Role of oxidative stress and DNA damage in human carcinogenesis. Mutat Res *711*, 193-201.

Kunkel, T.A., and Erie, D.A. (2005). DNA mismatch repair. Annu Rev Biochem *74*, 681-710.

Kuppers, R. (2005). Mechanisms of B-cell lymphoma pathogenesis. Nature Reviews Cancer *5*, 251-262.

Kuraoka, M., Liao, D., Yang, K., Allgood, S.D., Levesque, M.C., Kelsoe, G., and Ueda, Y. (2009). Activation-induced cytidine deaminase expression and activity in the absence of germinal centers: insights into hyper-IgM syndrome. J Immunol *183*, 3237-3248.

Lada, A.G., Krick, C.F., Kozmin, S.G., Mayorov, V.I., Karpova, T.S., Rogozin, I.B., and Pavlov, Y.I. (2011). Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. Biochemistry (Mosc) *76*, 131-146.

Ladner, R.D., and Caradonna, S.J. (1997). The human dUTPase gene encodes both nuclear and mitochondrial isoforms. Differential expression of the isoforms and characterization of a cDNA encoding the mitochondrial species. J Biol Chem *272*, 19072-19080.

Lakshmipathy, U., and Campbell, C. (2000). Mitochondrial DNA ligase III function is independent of Xrcc1. Nucleic Acids Res *28*, 3880-3886.

Lange, S.S., Takata, K., and Wood, R.D. (2011). DNA polymerases and cancer. Nat Rev Cancer *11*, 96-110.

Langerak, P., Nygren, A.O.H., Krijger, P.H.L., van den Berk, P.C.M., and Jacobs, H. (2007). A/T mutagenesis in hypermutated ARTICLE immunoglobulin genes strongly depends on PCNA(K164) modification. J Exp Med *204*, 1989-1998.

Lari, S.U., Chen, C.Y., Vertessy, B.G., Morre, J., and Bennett, S.E. (2006). Quantitative determination of uracil residues in Escherichia coli DNA: Contribution of ung, dug, and dut genes to uracil avoidance. DNA Repair (Amst) *5*, 1407-1420.

Larsen, E., Gran, C., Saether, B.E., Seeberg, E., and Klungland, A. (2003). Proliferation failure and gamma radiation sensitivity of Fen1 null mutant mice at the blastocyst stage. Mol Cell Biol *23*, 5346-5353.

Lasken, R.S., Schuster, D.M., and Rashtchian, A. (1996). Archaebacterial DNA polymerases tightly bind uracil-containing DNA. J Biol Chem *271*, 17692-17696.

Lau, P.P., Xiong, W.J., Zhu, H.J., Chen, S.H., and Chan, L. (1991). Apolipoprotein B mRNA editing is an intranuclear event that occurs posttranscriptionally coincident with splicing and polyadenylation. J Biol Chem *266*, 20550-20554.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet *11*, 204-220.

Leibeling, D., Laspe, P., and Emmert, S. (2006). Nucleotide excision repair and cancer. J Mol Histol *37*, 225-238.

Leonard, B., Hart, S.N., Burns, M.B., Carpenter, M.A., Temiz, N.A., Rathore, A., Isaksson Vogel, R., Nikas, J.B., Law, E.K., Brown, W.L.*, et al.* (2013). APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. Cancer Res.

Leonhardt, H., Rahn, H.P., Weinzierl, P., Sporbert, A., Cremer, T., Zink, D., and Cardoso, M.C. (2000). Dynamics of DNA replication factories in living cells. J Cell Biol *149*, 271-280.

Leppard, J.B., Dong, Z., Mackey, Z.B., and Tomkinson, A.E. (2003). Physical and functional interaction between DNA ligase IIIalpha and poly(ADP-Ribose) polymerase 1 in DNA single-strand break repair. Mol Cell Biol *23*, 5919-5927.

Levin, D.S., Bai, W., Yao, N., O'Donnell, M., and Tomkinson, A.E. (1997). An interaction between DNA ligase I and proliferating cell nuclear antigen: implications for Okazaki fragment synthesis and joining. Proc Natl Acad Sci U S A *94*, 12863-12868.

Levin, D.S., McKenna, A.E., Motycka, T.A., Matsumoto, Y., and Tomkinson, A.E. (2000). Interaction between PCNA and DNA ligase I is critical for joining of Okazaki fragments and long-patch base-excision repair. Curr Biol *10*, 919-922.

Li, G.M. (2008). Mechanisms and functions of DNA mismatch repair. Cell Res *18*, 85-98.

Li, Y.Q., Zhou, P.Z., Zheng, X.D., Walsh, C.P., and Xu, G.L. (2007). Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair. Nucleic Acids Res *35*, 390-400.

Liao, W., Hong, S.H., Chan, B.H., Rudolph, F.B., Clark, S.C., and Chan, L. (1999). APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. Biochem Biophys Res Commun *260*, 398-404.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem *79*, 181-211.

Liehr, J.G. (2001). Genotoxicity of the steroidal oestrogens oestrone and oestradiol: possible mechanism of uterine and mammary cancer development. Hum Reprod Update *7*, 273-281.

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.D., Glass, C.K.*, et al.* (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell *139*, 1069-1083.

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

Lindahl, T., and Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. Biochemistry *11*, 3610-3618.

Lindahl, T., and Nyberg, B. (1974). Heat-induced deamination of cytosine residues in deoxyribonucleic acid. Biochemistry *13*, 3405-3410.

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

Liu, M., and Schatz, D.G. (2009). Balancing AID and DNA repair during somatic hypermutation. Trends Immunol *30*, 173-181.

Loeb, L.A., and Preston, B.D. (1986). Mutagenesis by apurinic/apyrimidinic sites. Annu Rev Genet *20*, 201-230.

Lohr, J.G., Stojanov, P., Lawrence, M.S., Auclair, D., Chapuy, B., Sougnez, C., Cruz-Gordillo, P., Knoechel, B., Asmann, Y.W., Slager, S.L., *et al.* (2012). Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proc Natl Acad Sci U S A *109*, 3879-3884.

Longerich, S., Basu, U., Alt, F., and Storb, U. (2006). AID in somatic hypermutation and class switch recombination. Curr Opin Immunol *18*, 164-174.

Longley, D.B., Harkin, D.P., and Johnston, P.G. (2003). 5-fluorouracil: mechanisms of action and clinical strategies. Nat Rev Cancer *3*, 330-338.

Lu, T., Pan, Y., Kao, S.Y., Li, C., Kohane, I., Chan, J., and Yankner, B.A. (2004). Gene regulation and DNA damage in the ageing human brain. Nature *429*, 883-891.

Luo, X., and Kraus, W.L. (2012). On PAR with PARP: cellular stress signaling through poly(ADP-ribose) and PARP-1. Genes Dev *26*, 417-432.

Maga, G., and Hubscher, U. (2003). Proliferating cell nuclear antigen (PCNA): a dancer with many partners. J Cell Sci *116*, 3051-3060.

Mahaney, B.L., Meek, K., and Lees-Miller, S.P. (2009). Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. Biochem J *417*, 639-650.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J Biol Chem *286*, 35334-35338.

Malu, S., Malshetty, V., Francis, D., and Cortes, P. (2012). Role of non-homologous end joining in V(D)J recombination. Immunol Res *54*, 233-246.

Mano, J. (2012). Reactive carbonyl species: their production from lipid peroxides, action in environmental stress, and the detoxification mechanism. Plant Physiol Biochem *59*, 90-97.

Manvilla, B.A., Maiti, A., Begley, M.C., Toth, E.A., and Drohat, A.C. (2012). Crystal structure of human methyl-binding domain IV glycosylase bound to abasic DNA. J Mol Biol *420*, 164-175.

Masaoka, A., Matsubara, M., Hasegawa, R., Tanaka, T., Kurisu, S., Terato, H., Ohyama, Y., Karino, N., Matsuda, A., and Ide, H. (2003). 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 *42*, 5003-5012.

Mashiyama, S.T., Courtemanche, C., Elson-Schwab, I., Crott, J., Lee, B.L., Ong, C.N., Fenech, M., and Ames, B.N. (2004). Uracil in DNA, determined by an improved assay, is increased when deoxynucleosides are added to folate-deficient cultured human lymphocytes. Anal Biochem *330*, 58-69.

Mashiyama, S.T., Hansen, C.M., Roitman, E., Sarmiento, S., Leklem, J.E., Shultz, T.D., and Ames, B.N. (2008). An assay for uracil in human DNA at baseline: effect of marginal vitamin B6 deficiency. Anal Biochem *372*, 21-31.

Matsumoto, Y., Marusawa, H., Kinoshita, K., Endo, Y., Kou, T., Morisawa, T., Azuma, T., Okazaki, I.M., Honjo, T., and Chiba, T. (2007). Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. Nat Med *13*, 470-476.

Matsuoka, S., Ballif, B.A., Smogorzewska, A., McDonald, E.R., 3rd, Hurov, K.E., Luo, J., Bakalarski, C.E., Zhao, Z., Solimini, N., Lerenthal, Y., *et al.* (2007). ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science *316*, 1160-1166.

Maul, R.W., and Gearhart, P.J. (2010). AID and somatic hypermutation. Adv Immunol *105*, 159-191.

Maul, R.W., and Gearhart, P.J. (2014). Refining the Neuberger Model: uracil processing by activated B cells. Eur J Immunol.

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

McBride, K.M., Barreto, V., Ramiro, A.R., Stavropoulos, P., and Nussenzweig, M.C. (2004). Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. J Exp Med *199*, 1235-1244.

McBride, K.M., Gazumyan, A., Woo, E.M., Barreto, V.M., Robbiani, D.F., Chait, B.T., and Nussenzweig, M.C. (2006). Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. Proc Natl Acad Sci U S A *103*, 8798-8803.

McBride, K.M., Gazumyan, A., Woo, E.M., Schwickert, T.A., Chait, B.T., and Nussenzweig, M.C. (2008). Regulation of class switch recombination and somatic mutation by AID phosphorylation. J Exp Med *205*, 2585-2594.

Menissier de Murcia, J., Ricoul, M., Tartier, L., Niedergang, C., Huber, A., Dantzer, F., Schreiber, V., Ame, J.-C., Dierich, A., LeMeur, M*., et al.* (2003). Functional interaction between PARP-1 and PARP-2 in chromosome stability and embryonic development in mouse. EMBO J *22*, 2255-2263.

Millar, C.B., Guy, J., Sansom, O.J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P.D., Bishop, S.M., Clarke, A.R., and Bird, A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. Science *297*, 403-405.

Missero, C., Pirro, M.T., Simeone, S., Pischetola, M., and Di Lauro, R. (2001). The DNA glycosylase T:G mismatch-specific thymine DNA glycosylase represses thyroid transcription factor-1-activated transcription. J Biol Chem *276*, 33569-33575.

Mladenov, E., and Iliakis, G. (2011). Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. Mutat Res *711*, 61-72.

Mol, C.D., Arvai, A.S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H.E., and Tainer, J.A. (1995). Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. Cell *80*, 869-878.

Moldovan, G.L., Pfander, B., and Jentsch, S. (2007). PCNA, the maestro of the replication fork. Cell *129*, 665-679.

Morisawa, T., Marusawa, H., Ueda, Y., Iwai, A., Okazaki, I.M., Honjo, T., and Chiba, T. (2008). Organ-specific profiles of genetic changes in cancers caused by activation-induced cytidine deaminase expression. Int J Cancer *123*, 2735-2740.

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

Moser, J., Kool, H., Giakzidis, I., Caldecott, K., Mullenders, L.H., and Fousteri, M.I. (2007). Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. Mol Cell *27*, 311-323.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell *102*, 553-563.

Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K., and Honjo, T. (2000). Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene. Genomics *68*, 85-88.

Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y., and Kohli, R.M. (2012a). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nat Chem Biol *8*, 751-758.

Nabel, C.S., Manning, S.A., and Kohli, R.M. (2012b). The curious chemical biology of cytosine: deamination, methylation, and oxidation as modulators of genomic potential. ACS Chem Biol *7*, 20-30.

Nair, U., Bartsch, H., and Nair, J. (2007). Lipid peroxidation-induced DNA damage in cancer-prone inflammatory diseases: a review of published adduct types and levels in humans. Free Radic Biol Med *43*, 1109-1120.

Navarro, F., Bollman, B., Chen, H., Konig, R., Yu, Q., Chiles, K., and Landau, N.R. (2005). Complementary function of the two catalytic domains of APOBEC3G. Virology *333*, 374-386.

Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J.J., Wiebauer, K., and Jiricny, J. (1996). Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. J Biol Chem *271*, 12767-12774.

Neddermann, P., and Jiricny, J. (1994). 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*, 1642-1646.

Nelson, J.R., Lawrence, C.W., and Hinkle, D.C. (1996). Deoxycytidyl transferase activity of yeast REV1 protein. Nature *382*, 729-731.

Nie, B., Gan, W., Shi, F., Hu, G.X., Chen, L.G., Hayakawa, H., Sekiguchi, M., and Cai, J.P. (2013). Age-dependent accumulation of 8-oxoguanine in the DNA and RNA in various rat tissues. Oxid Med Cell Longev *2013*, 303181.

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

Nilsen, H., Rosewell, I., Robins, P., Skjelbred, C.F., Andersen, S., Slupphaug, G., Daly, G., Krokan, H.E., Lindahl, T., and Barnes, D.E. (2000). Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication. Mol Cell *5*, 1059-1065.

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.

Offer, S.M., Pan-Hammarstrom, Q., Hammarstrom, L., and Harris, R.S. (2010). Unique DNA repair gene variations and potential associations with the primary antibody deficiency syndromes IgAD and CVID. PLoS One *5*, e12260.

Okazaki, I.M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. J Exp Med *197*, 1173-1181.

Okuyama, S., Marusawa, H., Matsumoto, T., Ueda, Y., Matsumoto, Y., Endo, Y., Takai, A., and Chiba, T. (2012). Excessive activity of apolipoprotein B mRNA editing enzyme catalytic polypeptide 2 (APOBEC2) contributes to liver and lung tumorigenesis. Int J Cancer *130*, 1294-1301.

Olinski, R., Jurgowiak, M., and Zaremba, T. (2010). Uracil in DNA--its biological significance. Mutat Res *705*, 239-245.

Orthwein, A., Patenaude, A.M., Affar el, B., Lamarre, A., Young, J.C., and Di Noia, J.M. (2010). Regulation of activation-induced deaminase stability and antibody gene diversification by Hsp90. J Exp Med *207*, 2751-2765.

Orthwein, A., Zahn, A., Methot, S.P., Godin, D., Conticello, S.G., Terada, K., and Di Noia, J.M. (2012). Optimal functional levels of activation-induced deaminase specifically require the Hsp40 DnaJa1. EMBO J *31*, 679-691.

Osheroff, W.P., Jung, H.K., Beard, W.A., Wilson, S.H., and Kunkel, T.A. (1999). The fidelity of DNA polymerase beta during distributive and processive DNA synthesis. J Biol Chem *274*, 3642-3650.

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.

Park, S.R., Zan, H., Pal, Z., Zhang, J., Al-Qahtani, A., Pone, E.J., Xu, Z., Mai, T., and Casali, P. (2009). HoxC4 binds to the promoter of the cytidine deaminase AID gene to induce AID expression, class-switch DNA recombination and somatic hypermutation. Nat Immunol *10*, 540-550.

Parlanti, E., Locatelli, G., Maga, G., and Dogliotti, E. (2007). Human base excision repair complex is physically associated to DNA replication and cell cycle regulatory proteins. Nucleic Acids Res *35*, 1569-1577.

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

Pasqualucci, L., Kitaura, Y., Gu, H., and Dalla-Favera, R. (2006). PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. Proc Natl Acad Sci U S A *103*, 395-400.

Pasquinelli, A.E. (2012). MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. Nat Rev Genet *13*, 271-282.

Patenaude, A.M., Orthwein, A., Hu, Y., Campo, V.A., Kavli, B., Buschiazzo, A., and Di Noia, J.M. (2009). Active nuclear import and cytoplasmic retention of activation-induced deaminase. Nat Struct Mol Biol *16*, 517-527.

Pauklin, S., Sernandez, I.V., Bachmann, G., Ramiro, A.R., and Petersen-Mahrt, S.K. (2009). Estrogen directly activates AID transcription and function. J Exp Med *206*, 99-111.

Peron, S., Metin, A., Gardes, P., Alyanakian, M.A., Sheridan, E., Kratz, C.P., Fischer, A., and Durandy, A. (2008). Human PMS2 deficiency is associated with impaired immunoglobulin class switch recombination. J Exp Med *205*, 2465-2472.

Petersen-Mahrt, S.K., and Neuberger, M.S. (2003). In vitro deamination of cytosine to uracil in single-stranded DNA by apolipoprotein B editing complex catalytic subunit 1 (APOBEC1). J Biol Chem *278*, 19583-19586.

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

Pettersen, H.S., Visnes, T., Vagbo, C.B., Svaasand, E.K., Doseth, B., Slupphaug, G., Kavli, B., and Krokan, H.E. (2011). UNG-initiated base excision repair is the major repair route for 5-fluorouracil in DNA, but 5-fluorouracil cytotoxicity depends mainly on RNA incorporation. Nucleic Acids Res *39*, 8430-8444.

Pfeifer, G.P., Kadam, S., and Jin, S.G. (2013). 5-hydroxymethylcytosine and its potential roles in development and cancer. Epigenetics Chromatin *6*, 10.

Phan, R.T., and Dalla-Favera, R. (2004). The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. Nature *432*, 635-639.

Phillips, D.H., and Arlt, V.M. (2007). The 32P-postlabeling assay for DNA adducts. Nat Protoc *2*, 2772-2781.

Phillips, D.H., Farmer, P.B., Beland, F.A., Nath, R.G., Poirier, M.C., Reddy, M.V., and Turteltaub, K.W. (2000). Methods of DNA adduct determination and their application to testing compounds for genotoxicity. Environ Mol Mutagen *35*, 222-233.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature *463*, 1101-1105.

Poulsen, H.E., Specht, E., Broedbaek, K., Henriksen, T., Ellervik, C., Mandrup-Poulsen, T., Tonnesen, M., Nielsen, P.E., Andersen, H.U., and Weimann, A. (2012). RNA modifications by oxidation: a novel disease mechanism? Free Radic Biol Med *52*, 1353-1361.

Powers, H.J. (2005). Interaction among folate, riboflavin, genotype, and cancer, with reference to colorectal and cervical cancer. J Nutr *135*, 2960s-2966s.

Prochnow, C., Bransteitter, R., Klein, M.G., Goodman, M.F., and Chen, X.S. (2007). The APOBEC-2 crystal structure and functional implications for the deaminase AID. Nature *445*, 447-451.

Puebla-Osorio, N., Lacey, D.B., Alt, F.W., and Zhu, C. (2006). Early embryonic lethality due to targeted inactivation of DNA ligase III. Mol Cell Biol *26*, 3935-3941.

Racz, I., Kiraly, I., and Lasztily, D. (1978). Effect of light on the nucleotide composition of rRNA of wheat seedlings. Planta *142*, 263-267.

Rada, C., Di Noia, J.M., and Neuberger, M.S. (2004). Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. Mol Cell *16*, 163-171.

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.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell *135*, 1201-1212.

Ramiro, A.R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H.T., McBride, K.M., Eisenreich, T.R., Chen, J., Dickins, R.A., Lowe, S.W*., et al.* (2006). Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. Nature *440*, 105-109.

Ramiro, A.R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S., Muramatsu, M., Hongo, T., Nussenzweig, A., and Nussenzweig, M.C. (2004). AID is required for c-myc/IgH chromosome translocations in vivo. Cell *118*, 431-438.

Ranalli, T.A., Tom, S., and Bambara, R.A. (2002). AP endonuclease 1 coordinates flap endonuclease 1 and DNA ligase I activity in long patch base excision repair. J Biol Chem *277*, 41715-41724.

Ren, J., Ulvik, A., Refsum, H., and Ueland, P.M. (2002). 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*, 295-299.

Robbiani, D.F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A., *et al.* (2008). AID Is Required for the Chromosomal Breaks in c-myc that Lead to c-myc/IgH Translocations. Cell *135*, 1028-1038.

Robbiani, D.F., Bunting, S., Feldhahn, N., Bothmer, A., Camps, J., Deroubaix, S., McBride, K.M., Klein, I.A., Stone, G., Eisenreich, T.R., *et al.* (2009). AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. Mol Cell *36*, 631-641.

Robbiani, D.F., and Nussenzweig, M.C. (2013). Chromosome translocation, B cell lymphoma, and activation-induced cytidine deaminase. Annu Rev Pathol *8*, 79-103.

Roberts, M.J., Wondrak, G.T., Laurean, D.C., Jacobson, M.K., and Jacobson, E.L. (2003). DNA damage by carbonyl stress in human skin cells. Mutat Res *522*, 45-56.

Rogozin, I.B., Basu, M.K., Jordan, I.K., Pavlov, Y.I., and Koonin, E.V. (2005). APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. Cell Cycle *4*, 1281-1285.

Rosenberg, B.R., Hamilton, C.E., Mwangi, M.M., Dewell, S., and Papavasiliou, F.N. (2011). Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs. Nat Struct Mol Biol *18*, 230-236.

Rulten, S.L., Fisher, A.E., Robert, I., Zuma, M.C., Rouleau, M., Ju, L., Poirier, G., Reina-San-Martin, B., and Caldecott, K.W. (2011). PARP-3 and APLF function together to accelerate nonhomologous end-joining. Mol Cell *41*, 33-45.

Santella, R.M. (1999). Immunological methods for detection of carcinogen-DNA damage in humans. Cancer Epidemiol Biomarkers Prev *8*, 733-739.

Saribasak, H., and Gearhart, P.J. (2012). Does DNA repair occur during somatic hypermutation? Seminars in Immunology *24*, 287-292.

Sayegh, C.E., Quong, M.W., Agata, Y., and Murre, C. (2003). E-proteins directly regulate expression of activation-induced deaminase in mature B cells. Nat Immunol *4*, 586-593.

Schatz, D.G., and Swanson, P.C. (2011). V(D)J recombination: mechanisms of initiation. Annu Rev Genet *45*, 167-202.

Schiesser, S., Hackner, B., Pfaffeneder, T., Muller, M., Hagemeier, C., Truss, M., and Carell, T. (2012). Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. Angew Chem Int Ed Engl *51*, 6516-6520.

Schmitz, O.J., Worth, C.C., Stach, D., and Wiessler, M. (2002). Capillary electrophoresis analysis of DNA adducts as biomarkers for carcinogenesis. Angew Chem Int Ed Engl *41*, 445-448.

Schrader, C.E., Edelmann, W., Kucherlapati, R., and Stavnezer, J. (1999). Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes. J Exp Med *190*, 323-330.

Schroeder, H.W., Jr., and Cavacini, L. (2010). Structure and function of immunoglobulins. J Allergy Clin Immunol *125*, S41-52.

Schrofelbauer, B., Hakata, Y., and Landau, N.R. (2007). HIV-1 Vpr function is mediated by interaction with the damage-specific DNA-binding protein DDB1. Proc Natl Acad Sci U S A *104*, 4130-4135.

Schrofelbauer, B., Yu, Q., Zeitlin, S.G., and Landau, N.R. (2005). Human immunodeficiency virus type 1 Vpr induces the degradation of the UNG and SMUG uracil-DNA glycosylases. J Virol *79*, 10978-10987.

Screaton, R.A., Kiessling, S., Sansom, O.J., Millar, C.B., Maddison, K., Bird, A., Clarke, A.R., and Frisch, S.M. (2003). Fas-associated death domain protein interacts with methyl-CpG binding domain protein 4: a potential link between genome surveillance and apoptosis. Proc Natl Acad Sci U S A *100*, 5211-5216.

Sepe, S., Payan-Gomez, C., Milanese, C., Hoeijmakers, J.H., and Mastroberardino, P.G. (2013). Nucleotide excision repair in chronic neurodegenerative diseases. DNA Repair (Amst).

Shaffer, A.L., Lin, K.I., Kuo, T.C., Yu, X., Hurt, E.M., Rosenwald, A., Giltnane, J.M., Yang, L., Zhao, H., Calame, K.*, et al.* (2002). Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. Immunity *17*, 51-62.

Shall, S., and de Murcia, G. (2000). Poly(ADP-ribose) polymerase-1: what have we learned from the deficient mouse model? Mutat Res *460*, 1-15.

Shapiro, R. (1981). Damage to DNA caused by hydrolysis. In Chromosome Danage and Repair (Plenum Press, New York, NY), pp. pp. 3-18.

Sharbeen, G., Yee, C.W.Y., Smith, A.L., and Jolly, C.J. (2012). Ectopic restriction of DNA repair reveals that UNG2 excises AID-induced uracils predominantly or exclusively during G1 phase. The Journal of experimental medicine *209*, 965-974.

Shen, B., Singh, P., Liu, R., Qiu, J., Zheng, L., Finger, L.D., and Alas, S. (2005). Multiple but dissectible functions of FEN-1 nucleases in nucleic acid processing, genome stability and diseases. Bioessays *27*, 717-729.

Shen, H.M., Peters, A., Baron, B., Zhu, X., and Storb, U. (1998). Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. Science *280*, 1750-1752.

Shen, H.M., Tanaka, A., Bozek, G., Nicolae, D., and Storb, U. (2006). Somatic hypermutation and class switch recombination in Msh6(-/-)Ung(-/-) double-knockout mice. J Immunol *177*, 5386-5392.

Shrivastav, N., Li, D., and Essigmann, J.M. (2010). Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. Carcinogenesis *31*, 59-70.

Simsek, D., Furda, A., Gao, Y., Artus, J., Brunet, E., Hadjantonakis, A.K., Van Houten, B., Shuman, S., McKinnon, P.J., and Jasin, M. (2011). Crucial role for DNA ligase III in mitochondria but not in Xrcc1-dependent repair. Nature *471*, 245-248.

Singh, R., and Farmer, P.B. (2006). Liquid chromatography-electrospray ionization-mass spectrometry: the future of DNA adduct detection. Carcinogenesis *27*, 178-196.

Slupphaug, G., Eftedal, I., Kavli, B., Bharati, S., Helle, N.M., Haug, T., Levine, D.W., and Krokan, H.E. (1995). Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase. Biochemistry *34*, 128-138.

Slupphaug, G., Mol, C.D., Kavli, B., Arvai, A.S., Krokan, H.E., and Tainer, J.A. (1996). A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. Nature *384*, 87-92.

Smiglak, M., Reichert, W.M., Holbrey, J.D., Wilkes, J.S., Sun, L., Thrasher, J.S., Kirichenko, K., Singh, S., Katritzky, A.R., and Rogers, R.D. (2006). Combustible ionic liquids by design: is laboratory safety another ionic liquid myth? Chem Commun (Camb), 2554-2556.

Smith, H.C., Bennett, R.P., Kizilyer, A., McDougall, W.M., and Prohaska, K.M. (2012). Functions and regulation of the APOBEC family of proteins. Semin Cell Dev Biol *23*, 258-268.

Sousa, M.M., Krokan, H.E., and Slupphaug, G. (2007). DNA-uracil and human pathology. Mol Aspects Med *28*, 276-306.

Sowden, M.P., Ballatori, N., Jensen, K.L., Reed, L.H., and Smith, H.C. (2002). The editosome for cytidine to uridine mRNA editing has a native complexity of 27S: identification of intracellular domains containing active and inactive editing factors. J Cell Sci *115*, 1027-1039.

Stallons, L.J., and McGregor, W.G. (2010). Translesion synthesis polymerases in the prevention and promotion of carcinogenesis. J Nucleic Acids *2010*.

Stavnezer, J., Guikema, J.E., and Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. Annu Rev Immunol *26*, 261-292.

Stevnsner, T., Thorslund, T., de Souza-Pinto, N.C., and Bohr, V.A. (2002). Mitochondrial repair of 8-oxoguanine and changes with aging. Exp Gerontol *37*, 1189-1196.

Stopak, K.S., Chiu, Y.L., Kropp, J., Grant, R.M., and Greene, W.C. (2007). Distinct patterns of cytokine regulation of APOBEC3G expression and activity in primary lymphocytes, macrophages, and dendritic cells. J Biol Chem *282*, 3539-3546.

Stratton, M.R. (2011). Exploring the Genomes of Cancer Cells: Progress and Promise. Science *331*, 1553-1558.

Suspene, R., Aynaud, M.M., Guetard, D., Henry, M., Eckhoff, G., Marchio, A., Pineau, P., Dejean, A., Vartanian, J.P., and Wain-Hobson, S. (2011). Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. Proc Natl Acad Sci U S A *108*, 4858-4863.

Svilar, D., Goellner, E.M., Almeida, K.H., and Sobol, R.W. (2011). Base excision repair and lesion-dependent subpathways for repair of oxidative DNA damage. Antioxid Redox Signal *14*, 2491-2507.

Takai, A., Toyoshima, T., Uemura, M., Kitawaki, Y., Marusawa, H., Hiai, H., Yamada, S., Okazaki, I.M., Honjo, T., Chiba, T.*, et al.* (2009). A novel mouse model of hepatocarcinogenesis triggered by AID causing deleterious p53 mutations. Oncogene *28*, 469-478.

Takizawa, M., Tolarova, H., Li, Z., Dubois, W., Lim, S., Callen, E., Franco, S., Mosaico, M., Feigenbaum, L., Alt, F.W.*, et al.* (2008). AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. J Exp Med *205*, 1949-1957.

Tan, L., and Shi, Y.G. (2012). Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development *139*, 1895-1902.

Tang, E.S., and Martin, A. (2007). Immunoglobulin gene conversion: synthesizing antibody diversification and DNA repair. DNA Repair (Amst) *6*, 1557-1571.

Taylor, B.J., Nik-Zainal, S., Wu, Y.L., Stebbings, L.A., Raine, K., Campbell, P.J., Rada, C., Stratton, M.R., and Neuberger, M.S. (2013). DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. Elife *2*, e00534.

Tebbs, R.S., Flannery, M.L., Meneses, J.J., Hartmann, A., Tucker, J.D., Thompson, L.H., Cleaver, J.E., and Pedersen, R.A. (1999). Requirement for the Xrcc1 DNA base excision repair gene during early mouse development. Dev Biol *208*, 513-529.

Tell, G., Fantini, D., and Quadrifoglio, F. (2010). Understanding different functions of mammalian AP endonuclease (APE1) as a promising tool for cancer treatment. Cell Mol Life Sci *67*, 3589-3608.

Teng, B., Burant, C.F., and Davidson, N.O. (1993). Molecular cloning of an apolipoprotein B messenger RNA editing protein. Science *260*, 1816-1819.

Teng, G., Hakimpour, P., Landgraf, P., Rice, A., Tuschl, T., Casellas, R., and Papavasiliou, F.N. (2008). MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. Immunity *28*, 621-629.

Teperek-Tkacz, M., Pasque, V., Gentsch, G., and Ferguson-Smith, A.C. (2011). Epigenetic reprogramming: is deamination key to active DNA demethylation? Reproduction *142*, 621-632.

Tessman, I., Liu, S.K., and Kennedy, M.A. (1992). Mechanism of SOS mutagenesis of UV-irradiated DNA: mostly error-free processing of deaminated cytosine. Proc Natl Acad Sci U S A *89*, 1159-1163.

Thompson, L.H., and Schild, D. (2002). Recombinational DNA repair and human disease. Mutat Res *509*, 49-78.

Tini, M., Benecke, A., Um, S.J., Torchia, J., Evans, R.M., and Chambon, P. (2002). Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. Mol Cell *9*, 265-277.

Tran, T.H., Nakata, M., Suzuki, K., Begum, N.A., Shinkura, R., Fagarasan, S., Honjo, T., and Nagaoka, H. (2010). B cell-specific and stimulation-responsive enhancers derepress Aicda by overcoming the effects of silencers. Nat Immunol *11*, 148-154.

Traut, T.W. (1994). Physiological concentrations of purines and pyrimidines. Mol Cell Biochem *140*, 1-22.

Uchimura, A., Hidaka, Y., Hirabayashi, T., Hirabayashi, M., and Yagi, T. (2009). DNA polymerase delta is required for early mammalian embryogenesis. PLoS One *4*, e4184.

Um, S., Harbers, M., Benecke, A., Pierrat, B., Losson, R., and Chambon, P. (1998). Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. J Biol Chem *273*, 20728-20736.

Vagbo, C.B., Svaasand, E.K., Aas, P.A., and Krokan, H.E. (2013). Methylation damage to RNA induced in vivo in Escherichia coli is repaired by endogenous AlkB as part of the adaptive response. DNA Repair (Amst) *12*, 188-195.

Vascotto, C., Fantini, D., Romanello, M., Cesaratto, L., Deganuto, M., Leonardi, A., Radicella, J.P., Kelley, M.R., D'Ambrosio, C., Scaloni, A*., et al.* (2009). APE1/Ref-1 interacts with NPM1 within nucleoli and plays a role in the rRNA quality control process. Mol Cell Biol *29*, 1834-1854.

Vieira, V.C., and Soares, M.A. (2013). The role of cytidine deaminases on innate immune responses against human viral infections. Biomed Res Int *2013*, 683095.

Vigorito, E., Perks, K.L., Abreu-Goodger, C., Bunting, S., Xiang, Z., Kohlhaas, S., Das, P.P., Miska, E.A., Rodriguez, A., Bradley, A*., et al.* (2007). microRNA-155 regulates the generation of immunoglobulin class-switched plasma cells. Immunity *27*, 847-859.

Visnes, T., Doseth, B., Pettersen, H.S., Hagen, L., Sousa, M.M., Akbari, M., Otterlei, M., Kavli, B., Slupphaug, G., and Krokan, H.E. (2009). Uracil in DNA and its processing by different DNA glycosylases. Philos Trans R Soc Lond B Biol Sci *364*, 563-568.

Vora, K.A., Tumas-Brundage, K.M., Lentz, V.M., Cranston, A., Fishel, R., and Manser, T. (1999). Severe attenuation of the B cell immune response in Msh2-deficient mice. J Exp Med *189*, 471-482.

Wallenius, A., Hauser, J., Aas, P.A., Sarno, A., Kavli, B., Krokan, H.E., and Grundstrom, T. (2014). Expression and recruitment of uracil-DNA glycosylase are regulated by E2A during antibody diversification. Mol Immunol *60*, 23-31.

Wang, M., Wu, W., Rosidi, B., Zhang, L., Wang, H., and Iliakis, G. (2006). PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. Nucleic Acids Res *34*, 6170-6182.

Warbrick, E. (2000). The puzzle of PCNA's many partners. Bioessays *22*, 997-1006.

Waters, L.S., Minesinger, B.K., Wiltrout, M.E., D'Souza, S., Woodruff, R.V., and Walker, G.C. (2009). Eukaryotic Translesion Polymerases and Their Roles and Regulation in DNA Damage Tolerance. Microbiol Mol Biol R *73*, 134-+.

Wedekind, J.E., Gillilan, R., Janda, A., Krucinska, J., Salter, J.D., Bennett, R.P., Raina, J., and Smith, H.C. (2006). Nanostructures of APOBEC3G support a hierarchical assembly model of high molecular mass ribonucleoprotein particles from dimeric subunits. J Biol Chem *281*, 38122-38126.

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

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

Wu, P., Qiu, C., Sohail, A., Zhang, X., Bhagwat, A.S., and Cheng, X. (2003). Mismatch repair in methylated DNA. Structure and activity of the mismatch-specific thymine glycosylase domain of methyl-CpG-binding protein MBD4. J Biol Chem *278*, 5285-5291.

Wu, X., Li, J., Li, X., Hsieh, C.L., Burgers, P.M., and Lieber, M.R. (1996). Processing of branched DNA intermediates by a complex of human FEN-1 and PCNA. Nucleic Acids Res *24*, 2036-2043.

Wurtmann, E.J., and Wolin, S.L. (2009). RNA under attack: cellular handling of RNA damage. Crit Rev Biochem Mol Biol *44*, 34-49.

Xanthoudakis, S., Smeyne, R.J., Wallace, J.D., and Curran, T. (1996). The redox/DNA repair protein, Ref-1, is essential for early embryonic development in mice. Proc Natl Acad Sci U S A *93*, 8919-8923.

Xu, Z., Fulop, Z., Wu, G., Pone, E.J., Zhang, J., Mai, T., Thomas, L.M., Al-Qahtani, A., White, C.A., Park, S.R.*, et al.* (2010). 14-3-3 adaptor proteins recruit AID to 5'-AGCT-3'-rich switch regions for class switch recombination. Nat Struct Mol Biol *17*, 1124-1135.

Yadav, A., Olaru, A., Saltis, M., Setren, A., Cerny, J., and Livak, F. (2006). Identification of a ubiquitously active promoter of the murine activation-induced cytidine deaminase (AICDA) gene. Mol Immunol *43*, 529-541.

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

Yan, N., O'Day, E., Wheeler, L.A., Engelman, A., and Lieberman, J. (2011). HIV DNA is heavily uracilated, which protects it from autointegration. Proc Natl Acad Sci U S A *108*, 9244-9249.

Yang, B., Chen, K., Zhang, C., Huang, S., and Zhang, H. (2007). 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*, 11667-11675.

Zan, H., and Casali, P. (2013). Regulation of Aicda expression and AID activity. Autoimmunity *46*, 83-101.

Zhang, W., Liu, Z., Crombet, L., Amaya, M.F., Liu, Y., Zhang, X., Kuang, W., Ma, P., Niu, L., and Qi, C. (2011). Crystal structure of the mismatch-specific thymine glycosylase domain of human methyl-CpG-binding protein MBD4. Biochem Biophys Res Commun *412*, 425-428.

Zharkov, D.O. (2008). Base excision DNA repair. Cellular and Molecular Life Sciences *65*, 1544-1565.

Zheng, L., Dai, H., Hegde, M.L., Zhou, M., Guo, Z., Wu, X., Wu, J., Su, L., Zhong, X., Mitra, S.*, et al.* (2011). Fen1 mutations that specifically disrupt its interaction with PCNA cause aneuploidy-associated cancer. Cell Res *21*, 1052-1067.

Zheng, L., Dai, H., Qiu, J., Huang, Q., and Shen, B. (2007). Disruption of the FEN-1/PCNA interaction results in DNA replication defects, pulmonary hypoplasia, pancytopenia, and newborn lethality in mice. Mol Cell Biol *27*, 3176-3186.

Zou, L., Liu, D., and Elledge, S.J. (2003). Replication protein A-mediated recruitment and activation of Rad17 complexes. Proc Natl Acad Sci U S A *100*, 13827-13832.

Zust, R., Cervantes-Barragan, L., Habjan, M., Maier, R., Neuman, B.W., Ziebuhr, J., Szretter, K.J., Baker, S.C., Barchet, W., Diamond, M.S.*, et al.* (2011). Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. Nat Immunol *12*, 137-143.

Paper I

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

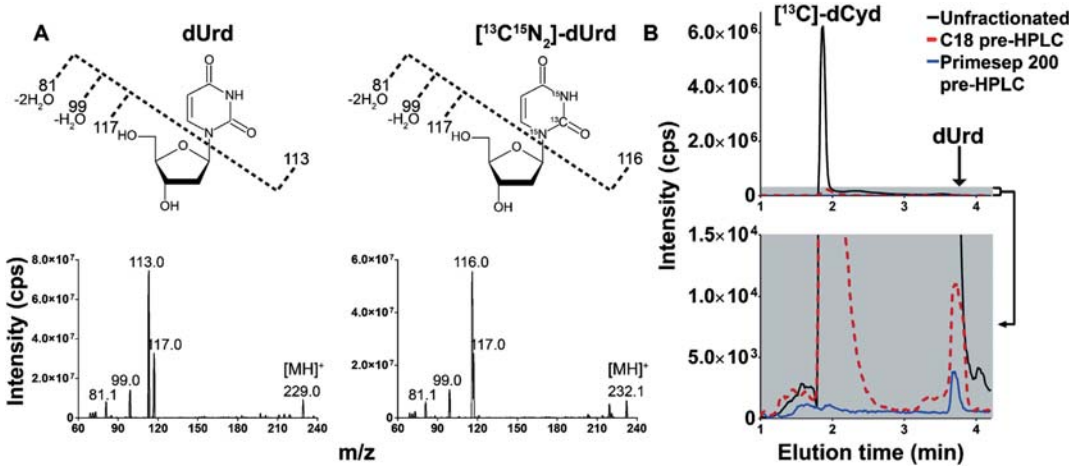


Fig. 1. Optimized LC/MS/MS conditions ensure method specificity. (A) MS/MS spectra of 2′-deoxyuridine (*m*/*z* 229) and [2-^{13}C,1,3-^{15}N]-2′-deoxyuridine (*m*/*z* 232.1) showing parent [MH]$^+$ to product ion transitions. The proposed origins of key fragments are indicated. Note that the collision 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

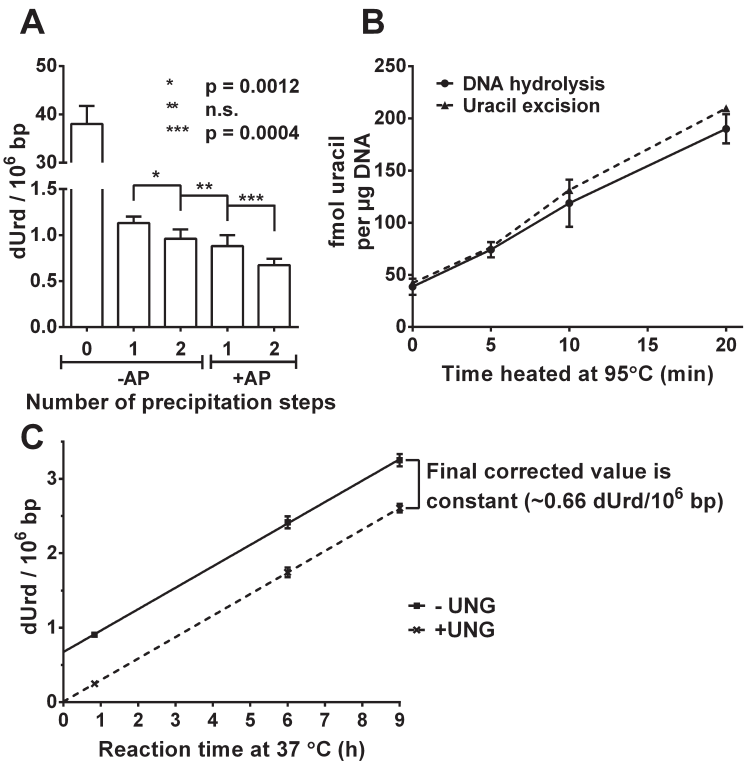


Fig. 2. Sample contamination with intracellular 2′-deoxyribonucleotides and *in vitro* dCyd deamination leads to overestimation of genomic dUrd. (A) Alkaline phosphatase (AP) pretreatment of commercially prepared DNA followed by repeated isopropanol precipitation steps prior to DNA hydrolysis decreases 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-R88 C 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

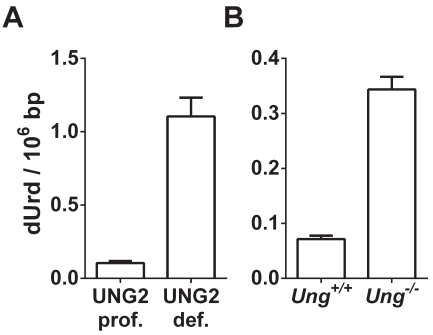


Fig. 3. Quantification of genomic dUrd in $Ung^{+/+}$ and $Ung^{-/-}$ mouse embryonic fibroblasts and human lymphoblastoid cell lines carrying *UNG* mutations. (A) UNG2 dysfunctional lymphoblastoid cells (UNG2-F251S) had 11-fold higher genomic dUrd level than lymphoblastoid cells with functional UNG2 (UNG2-R88 C). (B) Genomic dUrd level was 5-fold higher in $Ung^{-/-}$ than in $Ung^{+/+}$ mouse embryonic fibroblasts. Results represent triplicate experiments ±SD.

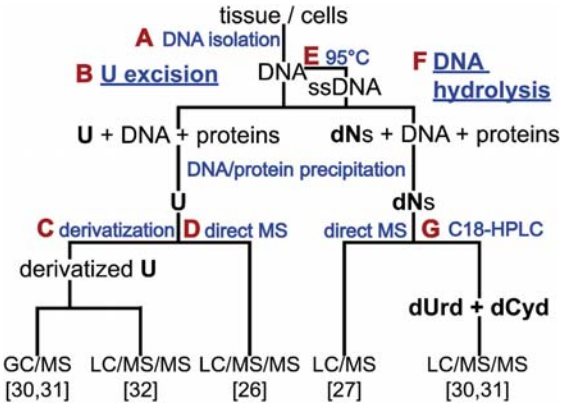


Fig. 4. Overview of possible errors in the methods for absolute quantification of genomic U/dUrd. (A) Intracellular 2′-deoxyribonucleotides can co-elute with DNA and be subsequently included in quantification. (B) Unspecific contaminants are usually more abundant with decreasing molecular weight of the 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.

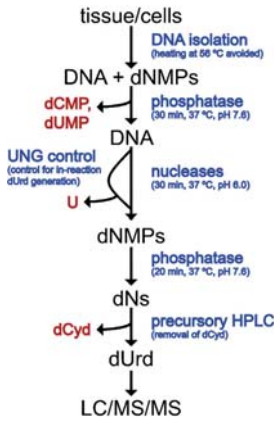


Fig. 5. Summary of improved genomic dUrd quantification. DNA isolation is improved by avoiding sample heating at 56 °C. A phosphatase pre-treatment step removes intracellular dCMP and dUMP, which otherwise co-purify with DNA. UNG2 is used to deuracilate DNA as a control processed in parallel to estimate 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.

Paper II

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

Henrik Sahlin Pettersen[1,3], Anastasia Galashevskaya[1], Berit Doseth[1], Mirta M. L. Sousa[1], Antonio Sarno[1], [§]Torkild Visnes[1], Per Arne Aas[1], Nina Beate Liabakk[1], Geir Slupphaug, Pål Sætrom[1,2], *Bodil Kavli[1] and *Hans E. Krokan[1]

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

[2] Department of Computer and Information Science, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway.

[3] Liaison Committee between the Central Norway Regional Health Authority (RHA) and the Norwegian University of Science and Technology (NTNU) and St. Olav's Hospital, Trondheim University Hospital, NO-7006 Trondheim, Norway

[§]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

*Corresponding authors: Hans E. Krokan; E-mail address: hans.krokan@ntnu.no

Bodil Kavli; E-mail address: bodil.kavli@ntnu.no

Department of Cancer Research and Molecular Medicine, NTNU, Erling Skjalgssons gate 1, N-7489 Trondheim, Norway.

Fax: + 47 72 57 64 00

Phone: + 47 72 57 30 74 (HEK); +47 72 573221 (BK)

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 direct 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.

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 (Kavli et al., 2007)). 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 (Muramatsu et al., 1999). However, evidence that AID was a DNA mutator in *E. coli* (Petersen-Mahrt et al., 2002) and its functional interaction with uracil-DNA glycosylase UNG in adaptive immunity (Di Noia and Neuberger, 2002; Imai et al., 2003; Rada et al., 2002), 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* (Conticello, 2008; Harris et al., 2002). 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 (Conticello, 2008). 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 mutations in human cancer, highlighting enzymatic deamination of cytosine to uracil as a potential source of mutagenesis (Alexandrov et al., 2013; Greenman et al., 2007; Zhang et al., 2013). However, the actual uracil level in normal and various cancer genomes has remained elusive.

Here, a sensitive LC/MS/MS-based method for directly 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.

RESULTS

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 (Figure 1A). The origin and major characteristics of cell lines is displayed in Figure 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 lymphocytes from blood donors. The mean value for the genomic uracil level in B-cell lymphoma cell lines (2.5 dU/10^6 nt) was 13.2-fold higher than in blood donor lymphocytes and significantly higher (4.4-fold, $p<0.001$) than the mean for non-lymphoma cell lines (0.57 dU/10^6 nt). This may suggest that B-cell

lymphoma cells may be exposed to enzymatic untargeted cytosine deamination throughout the genome.

AID expression correlates with genomic uracil accumulation

AID has previously been shown to be expressed in several lymphoma subtypes (Greeve et al., 2003; Lossos et al., 2004; Pasqualucci et al., 2004; Smit et al., 2003) and AID/APOBEC family enzymes were suggested to contribute to mutational signatures in a number of cancers by deaminating cytosine to uracil in DNA (Alexandrov et al., 2013). 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 (Figure 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 they were expressed in all cell lines as well as in the normal lymphocytes (Figure 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 (Figure 2B). This is a highly selective method allowing quantification of many protein targets in a single sample (Gallien et al., 2012; Peterson et al., 2012). In agreement with mRNA data, MS quantification revealed higher amounts of AID protein in lymphoma

cells with increased genomic uracil (Figure 2B, upper panel). Furthermore, similar to mRNA data (Figure 2A, middle panel), APOBEC3B, -3D, -3F and-3G proteins were expressed in all cell lines (Figure 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 (Figure 2C). As an additional control, we also quantified AID by western analysis, which yielded results similar to the MS analysis (Figure 2D). Importantly, AID expression significantly correlated with genomic uracil also at the protein level (R^2=0.65, P<0.0001) (Figure 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.

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 (Hu et al., 2014). AID is mostly localized in the cytoplasm (Figure 3A), but is actively imported into the nucleus where it may access the genome (Hu et al., 2013). We found that the cells expressing AID-YFP displayed an almost six-fold higher level of genomic uracil compared to the YFP control (Figure 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 hours (Figure 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 reduce proliferation (Figure 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 (Figure 2B and 2D). We found that a 60% knockdown of AID reduced genomic uracil level by 38% ($p = 0.005$; Figure 3C). Taken together these results strongly support the view that enzymatic cytosine deamination is the major source of genomic uracil in AID-expressing cells.

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 in human cells (Kavli et al., 2002). We have previously shown that UNG deficiency in human and mouse cells results in a several-fold increase in genomic uracil (Galashevskaya et al., 2013). The other uracil-DNA glycosylases, *i.e.* SMUG1, TDG, and MBD4, are thought to be quantitatively less important contributors, at least in proliferating cells (Kavli et al., 2002; Krokan and Bjoras, 2013; Pettersen et al., 2007). Furthermore, the DNA repair machinery has been shown to protect against AID-induced mutagenesis (Hasham et al., 2010; Liu et al., 2008; Yamane et al., 2011). Therefore, we measured uracil excision activity of cell free extracts prepared form 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 (Figure 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 (Figure 4B), although the correlation was 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 (Figure 4C). Using these activity values, a stronger correlation with genomic uracil level was observed (Figure 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 (Nilsen et al., 1997; Nilsen et al., 2000). 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 (Figure 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) (Figure 4F) correlated strongly with total uracil excision activity (Figure 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 (Figure 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 (Figure 4H) when analyzed separately (Table 2) or in combination with AID or other glycosylases.

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 (Andersen et al., 2005b). Indeed, we observed a significant inverse relationship between genomic uracil and doubling time in non-lymphoma cancer cells ($R^2 = 0.57$; $p = 0.048$; Figure 4I). Furthermore, since AID has been shown to act in the G_1 phase of the cell cycle (Ordinario et al., 2009; Schrader et al., 2007; Sharbeen et al., 2012), 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 (Figure 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 (Hagen et al., 2008; Hardeland et al., 2007). In contrast, TDG is mainly expressed in the G_1 phase of the cell cycle (Hagen et al., 2008; Hardeland et al., 2007). 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 (Pena-Diaz et al., 2013) and may contribute in all cell cycle phases, but is a rather slow acting enzyme (Kavli et al., 2002). 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 (Hagen et al., 2008), 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 (Akbari et al., 2004; Akbari et al., 2009; Visnes et al., 2008). To examine UNG, SMUG1 and TDG separately, we used a combination of neutralizing antibodies against UNG, SMUG1 and TDG. UNG was found to be a major contributor to initiate BER in all phases of the cell cycle, but SMUG1, and particularly TDG, contributed significantly in G_1 and G_2 (Figure 4J). Thus, a role of TDG and SMUG1 in BER of U:G mismatches in the G_1 phase, and a smaller role in the S-phase would seem likely from our *in vitro* data.

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 TpCpN sequence contexts (Alexandrov et al., 2013). As mentioned, these *kataegis* patterns might be different from those found in lymphomas and CLL (Alexandrov et al., 2013), 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 (Figure 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* (Figure 5B). This strongly implicates AID-induced genomic uracil formation in the development of localized hypermutation in B-cell malignancies, in accordance with our genomic uracil measurements and the published associations between AID and lymphomas (Bodor et al., 2005; Deutsch et al., 2007; Greeve et al., 2003; Hardianti et al., 2004; Lossos et al., 2004; Pasqualucci et al., 2004; Smit et al., 2003) and CLL (McCarthy et al., 2003; Palacios et al., 2010).

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. In general agreement with other studies (Greeve et al., 2003; Lossos et al., 2004; Pasqualucci et al., 2004; Smit et al., 2003) we find that AID is expressed in a large fraction of the lymphoma cell lines. AID is normally only expressed in activated germinal center B-cells (Crouch et al., 2007; Muramatsu et al., 1999) and at low but detectable levels in early developing B-cells in the bone marrow (Han et al., 2007). This is apparently a risky process because AID strongly promotes the generation of germinal center-derived lymphomas (Kotani et al., 2007; Pasqualucci et al., 2008; Smit et al., 2003), in which off-target activity of AID may contribute to point mutations and translocations during lymphomagenesis (Hakim et al., 2012; Klein et al., 2011; Liu et al., 2008). 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 (Alexandrov et al., 2013; Greenman et al., 2007; Zhang et al., 2013). However, direct

evidence from measurements of uracil in the cancer genomes is largely missing. There is also a lack of information on the modulating effects of uracil repair proteins, as well as other proteins that may influence genomic uracil levels.

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 hours. 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. We also found that recombinant AID overexpression further increases genomic uracil, whereas AID knockdown decreases it. It therefore appears to be rather clear that AID expression is the cause of genomic uracil accumulation, presumably as U:G mismatches, in B-cell lymphomas. In contrast, expression of other APOBEC-family members did not clearly correlate with genomic uracil levels. This does not rule out these as DNA 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 (Alexandrov et al., 2013; Burns et al., 2013).

Although AID expression levels correlated with variation in genomic uracil in the cells we tested, our results indicate that additonal 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 (Visnes et al., 2008) although UNG and SMUG1

may have complementary roles in uracil repair (Kavli et al., 2002; Nilsen et al., 2001; Pettersen et al., 2007) and in the prevention of mutagenesis (An et al., 2005). Studies on *UNG*$^{-/-}$ cells have documented an important function for UNG in keeping genomic uracil levels low (Galashevskaya et al., 2013). 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 (Liu et al., 2008). 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 (Hauser et al., 2008; Wallenius et al., 2014), but represses both UNG-expression and its binding to relevant regions in the Ig genes (Wallenius et al., 2014). Furthermore, p53 is actively reduced in germinal center B cells, presumably to allow mutagenic processing required for antibody maturation (Phan and Dalla-Favera, 2004). 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 (Kuppers and Dalla-Favera, 2001). In contrast, AID knockout mice have fewer translocations (Dorsett et al., 2007) and accumulate fewer mutations in genes linked to B cell tumorigenesis (Liu et al., 2008). AID expression has also been shown to confer a mutator phenotype in established lymphomas (Bodor et al., 2005; Deutsch et al., 2007; Hardianti et al., 2004), but the role of AID in cancer progression remains unsettled (Leuenberger et al., 2010; Lossos et al., 2004; Willenbrock et al., 2009). Interestingly, AID expression has been reported in numerous cancers of non-B-cell origin, including breast, prostate, stomach, liver, and lung (Orthwein and Di Noia, 2012). 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 (Andersen et al., 2005a; Nilsen et al., 2003). 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 (Figure 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 (Bodor et al., 2005; Deutsch et al., 2007; Hardianti et al., 2004) and CLL (McCarthy et al., 2003; Palacios et al., 2010). Interestingly, progression of established cancers through expression of AID was also demonstrated in other blood cell cancers, such as ALL (Gruber et al., 2010) and chronic myelogenous leukemia (CML), in which AID expression

may lead to fatal lymphoblastoid crisis (Klemm et al., 2009). 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 found that the level of genomic uracil is increased several-fold in B-cell lymphoma cell lines, correlating highly significantly with AID expression. Furthermore, the function of AID as a cytosine deaminase was demonstrated by increased genomic uracil levels after induction of AID, as well as after AID overexpression. 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.

Upon preparation for submission of our manuscript, a paper in press that in part overlaps with our results became available. Using a different method that indirectly measures relative genomic uracil levels the authors reported increased uracil in DNA from AID expressing B-cell lymphoma and CLL cells (Shalhout et al., 2014).

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.

Materials and methods

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. "Buffy" 1-3 lymphocytes were purified from buffy coats from healthy blood donors using the Lymphoprep^{TM} (Progen) kit 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 ºC 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 (1x final), 0.1 mg/ml gentamicin, and 2.3 µg/ml fungizone, and 20% heat inactivated (56 ºC, 20 min) FCS at 37 ºC and 5% CO_2. For quantitative rtPCR and uracil measurements cells were harvested at densities between 750 000 - 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, 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 (Hu et al., 2013). CH12F3 cells ($2x10^6$ 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 #39-2500, clone ZA001, 500 μg/ml (Invitrogen). Nuclear extracts from synchronized HeLa cells were prepared essentially as described (Hagen et al., 2008; Visnes et al., 2008).

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.

Quantification of uracil in DNA by LC/MS/MS

Genomic uracil was quantified as previously described (Galashevskaya et al., 2013). 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 x 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 x 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.

***In vitro* uracil DNA excision activity and complete BER assays**

Standard UDG activity assay was performed as described (Kavli et al., 2002). Briefly, 20 µl reaction mixtures containing (final) 1.8 µM nick translated [^{3}H]-dUMP-labeled calf thymus DNA (U:A substrate), 1x 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 (Akbari et al., 2004). Oligodeoxynucleotide UDG assays were performed as described (Kavli et al., 2002). 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, 1x 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 (Akbari et al., 2004; Visnes et al., 2008). 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.

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 ºC for 30 min prior to DNA staining with 200 µl propidium iodide (50 µg/ml in PBS) at 37 ºC for 30 min. Cell cycle analyses were performed using a FACS Canto flow cytometer (BD-Life Science).

Sample preparation and targeted mass spectrometry

Cell pellets were resuspended in 1x packed cell volume in buffer I: 10 mM Tris-HCl pH 8.0, 200 mM KCl, 1x complete protease inhibitor, and 5x 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, 1x complete protease inhibitor, and 5x 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 minutes followed by alkylation with 1 µmol/mg protein of iodoacetamide for 45 minutes in the dark. Proteins were precipitated using a methanol-chloroform method as described (Batth et al., 2012), 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. x 2 cm, C18, 5 µm, 100 Å) (Thermo Scientific) and further separated on a Acclaim PepMap100 C-18 analytical column (75 µm i.d. x 50 cm, C18, 3 µm, 100 Å) (Thermo Scientific). A 120 minute method was used and consisted of a 300 nl/minute 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 minutes, followed by an increase to 35% Buffer B over 98 minutes and a rapid increase to 100% buffer B in 6 minutes, where it was held for 5.5 minutes. The solvent composition was quickly ramped to 0% buffer B, where it was subsequently held for 8 minutes 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 $2x10^5$, 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 (MacLean et al., 2010). *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 minutes. 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.

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 (Alexandrov et al., 2013). 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.

REFERENCES

Akbari, M., M. Otterlei, J. Pena-Diaz, P.A. Aas, B. Kavli, N.B. Liabakk, L. Hagen, K. Imai, A. Durandy, G. Slupphaug, and H.E. Krokan. 2004. 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 research* 32:5486-5498.

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

Alexandrov, L.B., 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, and M.R. Stratton. 2013. Signatures of mutational processes in human cancer. *Nature* 500:415-421.

An, Q., P. Robins, T. Lindahl, and D.E. Barnes. 2005. C --> T mutagenesis and gamma-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases. *The EMBO journal* 24:2205-2213.

Andersen, S., M. Ericsson, H.Y. Dai, J. Pena-Diaz, G. Slupphaug, H. Nilsen, H. Aarset, and H.E. Krokan. 2005a. Monoclonal B-cell hyperplasia and leukocyte imbalance precede development of B-cell malignancies in uracil-DNA glycosylase deficient mice. *DNA repair* 4:1432-1441.

Andersen, S., T. Heine, R. Sneve, I. Konig, H.E. Krokan, B. Epe, and H. Nilsen. 2005b. Incorporation of dUMP into DNA is a major source of spontaneous DNA damage, while excision of uracil is not required for cytotoxicity of fluoropyrimidines in mouse embryonic fibroblasts. *Carcinogenesis* 26:547-555.

Batth, T.S., J.D. Keasling, and C.J. Petzold. 2012. Targeted proteomics for metabolic pathway optimization. *Methods in molecular biology* 944:237-249.

Bodor, C., A. Bognar, L. Reiniger, A. Szepesi, E. Toth, L. Kopper, and A. Matolcsy. 2005. Aberrant somatic hypermutation and expression of activation-induced cytidine deaminase mRNA in mediastinal large B-cell lymphoma. *British journal of haematology* 129:373-376.

Burns, M.B., 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, and R.S. Harris. 2013. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 494:366-370.

Conticello, S.G. 2008. The AID/APOBEC family of nucleic acid mutators. *Genome biology* 9:229.

Crouch, E.E., Z. Li, M. Takizawa, S. Fichtner-Feigl, P. Gourzi, C. Montano, L. Feigenbaum, P. Wilson, S. Janz, F.N. Papavasiliou, and R. Casellas. 2007. Regulation of AID expression in the immune response. *The Journal of experimental medicine* 204:1145-1156.

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

Di Noia, J., and M.S. Neuberger. 2002. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* 419:43-48.

Dorsett, Y., D.F. Robbiani, M. Jankovic, B. Reina-San-Martin, T.R. Eisenreich, and M.C. Nussenzweig. 2007. A role for AID in chromosome translocations between c-myc and the IgH variable region. *The Journal of experimental medicine* 204:2225-2232.

Galashevskaya, A., A. Sarno, C.B. Vagbo, P.A. Aas, L. Hagen, G. Slupphaug, and H.E. Krokan. 2013. A robust, sensitive assay for genomic uracil determination by LC/MS/MS reveals lower levels than previously reported. *DNA repair* 12:699-706.

Gallien, S., E. Duriez, C. Crone, M. Kellmann, T. Moehring, and B. Domon. 2012. Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Molecular & cellular proteomics : MCP* 11:1709-1723.

Greenman, C., 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, and M.R. Stratton. 2007. Patterns of somatic mutation in human cancer genomes. *Nature* 446:153-158.

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

Gruber, T.A., M.S. Chang, R. Sposto, and M. Muschen. 2010. Activation-induced cytidine deaminase accelerates clonal evolution in BCR-ABL1-driven B-cell lineage acute lymphoblastic leukemia. *Cancer research* 70:7411-7420.

Hagen, L., 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, and G. Slupphaug. 2008. Cell cycle-specific UNG2 phosphorylations regulate protein turnover, activity and association with RPA. *The EMBO journal* 27:51-61.

Hakim, O., 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, and R. Casellas. 2012. DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. *Nature* 484:69-74.

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

Hardeland, U., C. Kunz, F. Focke, M. Szadkowski, and P. Schar. 2007. Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2. *Nucleic acids research* 35:3859-3867.

Hardianti, M.S., E. Tatsumi, M. Syampurnawati, K. Furuta, K. Saigo, Y. Nakamachi, S. Kumagai, H. Ohno, S. Tanabe, M. Uchida, and N. Yasuda. 2004. Activation-induced

cytidine deaminase expression in follicular lymphoma: association between AID expression and ongoing mutation in FL. *Leukemia* 18:826-831.

Harris, R.S., S.K. Petersen-Mahrt, and M.S. Neuberger. 2002. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Molecular cell* 10:1247-1253.

Hasham, M.G., N.M. Donghia, E. Coffey, J. Maynard, K.J. Snow, J. Ames, R.Y. Wilpan, Y. He, B.L. King, and K.D. Mills. 2010. Widespread genomic breaks generated by activation-induced cytidine deaminase are prevented by homologous recombination. *Nature immunology* 11:820-826.

Hauser, J., N. Sveshnikova, A. Wallenius, S. Baradaran, J. Saarikettu, and T. Grundstrom. 2008. B-cell receptor activation inhibits AID expression through calmodulin inhibition of E-proteins. *Proceedings of the National Academy of Sciences of the United States of America* 105:1267-1272.

Hu, Y., I. Ericsson, B. Doseth, N.B. Liabakk, H.E. Krokan, and B. Kavli. 2014. Activation-induced cytidine deaminase (AID) is localized to subnuclear domains enriched in splicing factors. *Experimental cell research* 322:178-192.

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

Imai, K., 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, and A. Durandy. 2003. Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nature immunology* 4:1023-1028.

Kavli, B., M. Otterlei, G. Slupphaug, and H.E. Krokan. 2007. Uracil in DNA--general mutagen, but normal intermediate in acquired immunity. *DNA repair* 6:505-516.

Kavli, B., O. Sundheim, M. Akbari, M. Otterlei, H. Nilsen, F. Skorpen, P.A. Aas, L. Hagen, H.E. Krokan, and G. Slupphaug. 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. *The Journal of biological chemistry* 277:39926-39936.

Klein, I.A., W. Resch, M. Jankovic, T. Oliveira, A. Yamane, H. Nakahashi, M. Di Virgilio, A. Bothmer, A. Nussenzweig, D.F. Robbiani, R. Casellas, and M.C. Nussenzweig. 2011. Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* 147:95-106.

Klemm, L., 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, and M. Muschen. 2009. The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia. *Cancer cell* 16:232-245.

Kotani, A., N. Kakazu, T. Tsuruyama, I.M. Okazaki, M. Muramatsu, K. Kinoshita, H. Nagaoka, D. Yabe, and T. Honjo. 2007. Activation-induced cytidine deaminase (AID) promotes B cell lymphomagenesis in Emu-cmyc transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* 104:1616-1620.

Krokan, H.E., and M. Bjoras. 2013. Base excision repair. *Cold Spring Harbor perspectives in biology* 5:a012583.

Kuppers, R., and R. Dalla-Favera. 2001. Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene* 20:5580-5594.

Leuenberger, M., S. Frigerio, P.J. Wild, F. Noetzli, D. Korol, D.R. Zimmermann, C. Gengler, N.M. Probst-Hensch, H. Moch, and M. Tinguely. 2010. AID protein expression in

chronic lymphocytic leukemia/small lymphocytic lymphoma is associated with poor prognosis and complex genetic alterations. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 23:177-186.

Liu, M., J.L. Duke, D.J. Richter, C.G. Vinuesa, C.C. Goodnow, S.H. Kleinstein, and D.G. Schatz. 2008. Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 451:841-845.

Lossos, I.S., R. Levy, and A.A. Alizadeh. 2004. 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* 18:1775-1779.

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

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

Muramatsu, M., V.S. Sankaranand, S. Anant, M. Sugai, K. Kinoshita, N.O. Davidson, and T. Honjo. 1999. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *The Journal of biological chemistry* 274:18470-18476.

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

Nilsen, H., M. Otterlei, T. Haug, K. Solum, T.A. Nagelhus, F. Skorpen, and H.E. Krokan. 1997. Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene. *Nucleic acids research* 25:750-755.

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

Nilsen, H., K.S. Steinsbekk, M. Otterlei, G. Slupphaug, P.A. Aas, and H.E. Krokan. 2000. 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 research* 28:2277-2285.

Ordinario, E.C., M. Yabuki, R.P. Larson, and N. Maizels. 2009. Temporal regulation of Ig gene diversification revealed by single-cell imaging. *Journal of immunology* 183:4545-4553.

Orthwein, A., and J.M. Di Noia. 2012. Activation induced deaminase: how much and where? *Seminars in immunology* 24:246-254.

Palacios, F., P. Moreno, P. Morande, C. Abreu, A. Correa, V. Porro, A.I. Landoni, R. Gabus, M. Giordano, G. Dighiero, O. Pritsch, and P. Oppezzo. 2010. 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* 115:4488-4496.

Pasqualucci, L., G. Bhagat, M. Jankovic, M. Compagno, P. Smith, M. Muramatsu, T. Honjo, H.C. Morse, 3rd, M.C. Nussenzweig, and R. Dalla-Favera. 2008. AID is required for germinal center-derived lymphomagenesis. *Nature genetics* 40:108-112.

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

Pena-Diaz, J., S.A. Hegre, E. Anderssen, P.A. Aas, R. Mjelle, G.D. Gilfillan, R. Lyle, F. Drablos, H.E. Krokan, and P. Saetrom. 2013. Transcription profiling during the cell cycle shows that a subset of Polycomb-targeted genes is upregulated during DNA replication. *Nucleic acids research* 41:2846-2856.

Petersen-Mahrt, S.K., R.S. Harris, and M.S. Neuberger. 2002. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature* 418:99-103.

Peterson, A.C., J.D. Russell, D.J. Bailey, M.S. Westphall, and J.J. Coon. 2012. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Molecular & cellular proteomics : MCP* 11:1475-1488.

Pettersen, H.S., O. Sundheim, K.M. Gilljam, G. Slupphaug, H.E. Krokan, and B. Kavli. 2007. Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. *Nucleic acids research* 35:3879-3892.

Phan, R.T., and R. Dalla-Favera. 2004. The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. *Nature* 432:635-639.

Rada, C., G.T. Williams, H. Nilsen, D.E. Barnes, T. Lindahl, and M.S. Neuberger. 2002. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Current biology : CB* 12:1748-1755.

Schrader, C.E., J.E. Guikema, E.K. Linehan, E. Selsing, and J. Stavnezer. 2007. Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair. *Journal of immunology* 179:6064-6071.

Shalhout, S., D. Haddad, A. Sosin, T.C. Holland, A. Al-Katib, A. Martin, and A.S. Bhagwat. 2014. Genomic Uracil Homeostasis during Normal B Cell Maturation and Loss of this Balance During B Cell Cancer Development. *Molecular and cellular biology*

Sharbeen, G., C.W. Yee, A.L. Smith, and C.J. Jolly. 2012. Ectopic restriction of DNA repair reveals that UNG2 excises AID-induced uracils predominantly or exclusively during G1 phase. *The Journal of experimental medicine* 209:965-974.

Smit, L.A., R.J. Bende, J. Aten, J.E. Guikema, W.M. Aarts, and C.J. van Noesel. 2003. Expression of activation-induced cytidine deaminase is confined to B-cell non-Hodgkin's lymphomas of germinal-center phenotype. *Cancer research* 63:3894-3898.

Visnes, T., M. Akbari, L. Hagen, G. Slupphaug, and H.E. Krokan. 2008. The rate of base excision repair of uracil is controlled by the initiating glycosylase. *DNA repair* 7:1869-1881.

Wallenius, A., J. Hauser, P.A. Aas, A. Sarno, B. Kavli, H.E. Krokan, and T. Grundstrom. 2014. Expression and recruitment of uracil-DNA glycosylase are regulated by E2A during antibody diversification. *Molecular immunology* 60:23-31.

Willenbrock, K., C. Renne, M. Rottenkolber, W. Klapper, M. Dreyling, M. Engelhard, R. Kuppers, M.L. Hansmann, and B. Jungnickel. 2009. The expression of activation induced cytidine deaminase in follicular lymphoma is independent of prognosis and stage. *Histopathology* 54:509-512.

Yamane, A., W. Resch, N. Kuo, S. Kuchen, Z. Li, H.W. Sun, D.F. Robbiani, K. McBride, M.C. Nussenzweig, and R. Casellas. 2011. Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nature immunology* 12:62-69.

Zhang, J., 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, and S.S. Dave. 2013. Genetic heterogeneity of diffuse large B-cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* 110:1398-1403.

TABLES

Table 1: Linear regression analysis of genomic uracil levels (linear) versus AID and APOBEC protein expression (log) normalized to total protein. Bold green indicates significant positive correlation.

	All cell lines including buffy		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

Table 2: Linear regression analysis of genomic uracil levels (linear) versus expression of uracil-DNA repair glycosylases (linear) normalized either to total protein or to total protein per cell. Bold red indicates significant negative correlation.

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

FIGURE LEGENDS

Figure 1. Genomic uracil levels in lymphoma-/non-lymphoma cell lines, and buffy coat lymphocytes. (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 lymphocytes isolated from buffy coats from blood donors (red). Asterisk (*) signifies measurements significantly ($p < 0.05$) different from average genomic uracil levels in buffy coat lymphocytes 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 used in the study and their origin. Buffy coat lymphocytes were isolated by Lymphoprep^{TM} (Progen). B-NHL: B-cell non-Hodgkin lymphoma.

Figure 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 lymphocytes isolated from buffy coats in red. Cell lines within each group are ordered along the x-axis according to increasing genomic uracil levels, as in Figure 1. mRNA levels have been normalized to GAPDH mRNA, and protein levels to MS signal counts per total injected protein. Linear regression plots of genomic uracil (dU/10^6 nt) vs. AID mRNA and protein levels are presented in the lower panels in Figure 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.

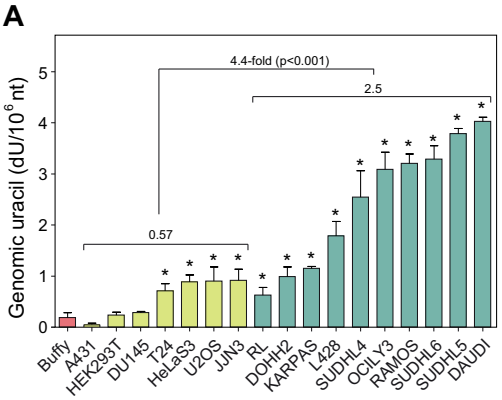
Figure 3. Genomic uracil levels after stimulation of endogenous AID expression, AID-YFP overexpression, and AID knockdown. (**A**) Genomic uracil levels in DNA isolated from mouse lymphoma cells (CH12F3) stably transfected with AID-YFP or YFP, and confocal microscopy showing subcellular distribution 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.

Figure 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 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. The data points represent mean and SD of at least two parallel experiments.

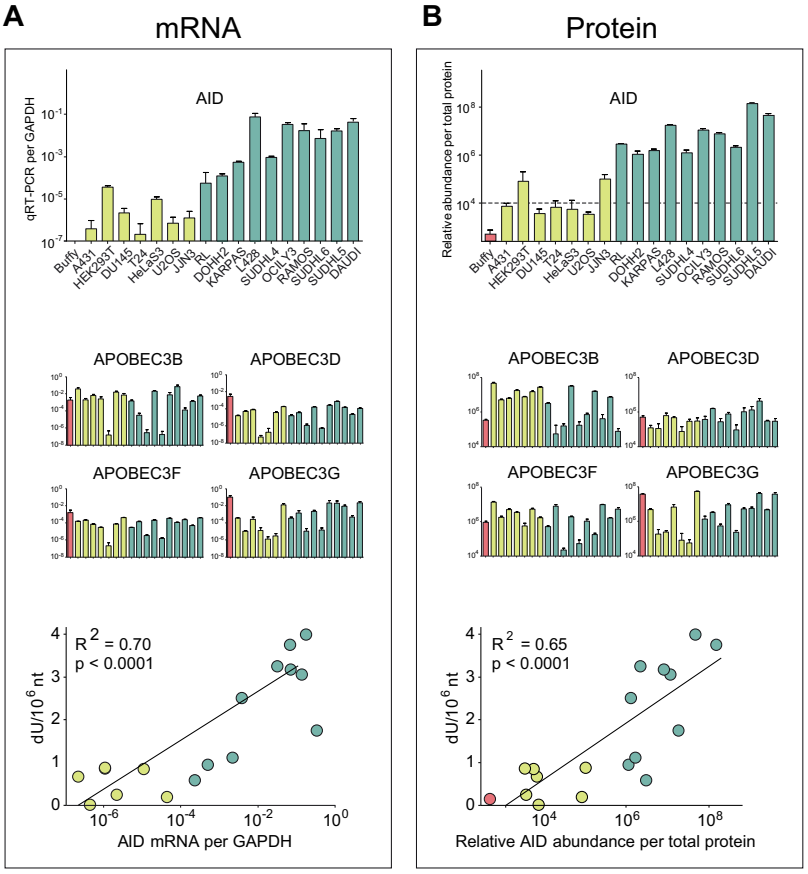
Figure 5. Sequence context of C to T transitions in *kataegis* regions of lymphomas. Sequence context of C to T transitions in *kataegis* regions of lymphomas (n = 21; 1102 single mutation sites) and CLL (n = 15; 290 single mutation sites) showing an AID-hotspot consensus sequence (-AG**C**TN-), where N represent 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). Sequence analyses are based on exome sequencing data obtained from (Alexandrov et al., 2013).

A



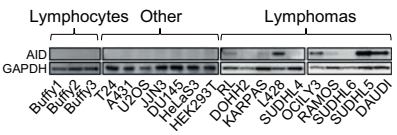
B

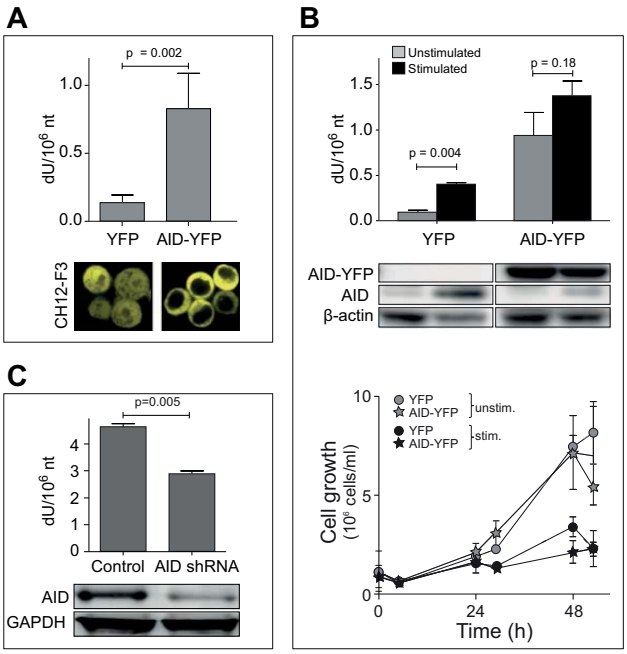
Name	Cell/tissue type	Disease
Buffy coat	Lymphocyte/blood	Healthy blood donors
A431	Epithelial/skin	Epidermoid carcinoma
HEK293T	Embryonic kidney	NA
DU145	Epithelial/prostate	Prostate cancer
T24	Epithelial/bladder	Bladder carcinoma
HeLaS3	Epithelial/cervix	Adenocarcinoma
U2OS	Osteocyte/bone	Osteosarcoma
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

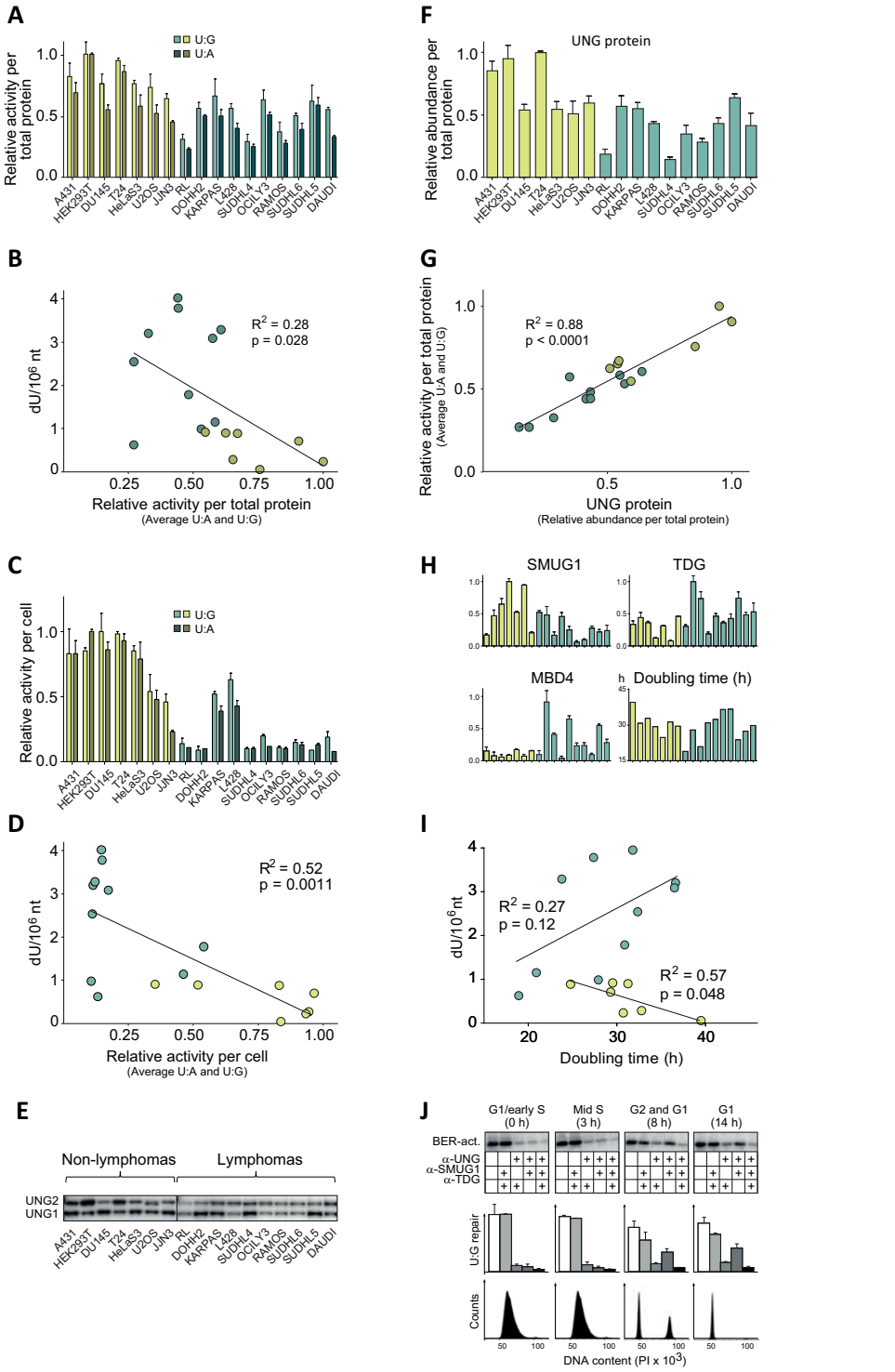


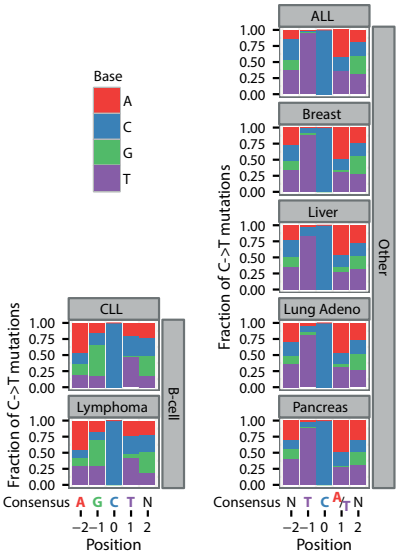
C

	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




A
p = 0.002
dU/10^6 nt
1.0
0.5
0.0
YFP
AID-YFP
CH12-F3
B
Unstimulated
Stimulated
p = 0.18
p = 0.004
dU/10^6 nt
1.5
1.0
0.5
0.0
YFP
AID-YFP
AID-YFP
AID
β-actin
C
p=0.005
dU/10^6 nt
4
3
2
1
0
Control
AID shRNA
AID
GAPDH
YFP
AID-YFP
unstim.
YFP
AID-YFP
stim.
Cell growth
(10^5 cells/ml)
10
5
0
0
24
48
Time (h)


A
U:G
U:A
Relative activity per total protein
1.0
0.5
0.0
A431
HEK293T
DU145
T24
HeLaS3
U2OS
JJN3
RL
DOHH2
KARPAS
L428
SUDHL4
OCILY3
RAMOS
SUDHL6
SUDHL5
DAUDI
F
UNG protein
Relative abundance per total protein
B
dU/10^6 nt
R^2 = 0.28
p = 0.028
Relative activity per total protein
(Average U:A and U:G)
G
Relative activity per total protein
(Average U:A and U:G)
R^2 = 0.88
p < 0.0001
UNG protein
(Relative abundance per total protein)
C
Relative activity per cell
H
SMUG1
TDG
MBD4
h
Doubling time (h)
D
R^2 = 0.52
p = 0.0011
Relative activity per cell
(Average U:A and U:G)
I
dU/10^6nt
R^2 = 0.27
p = 0.12
R^2 = 0.57
p = 0.048
Doubling time (h)
E
Non-lymphomas
Lymphomas
UNG2
UNG1
J
G1/early S
(0 h)
Mid S
(3 h)
G2 and G1
(8 h)
G1
(14 h)
BER-act.
α-UNG
α-SMUG1
α-TDG
U:G repair
Counts
DNA content (PI x 10^3)


Base
A
C
G
T
CLL
Lymphoma
B-cell
ALL
Breast
Liver
Lung Adeno
Pancreas
Other
Fraction of C->T mutations
1.00
0.75
0.50
0.25
0.00
Consensus A G C T N
Consensus N T C A/T N
−2 −1 0 1 2
Position

Paper III

Molecular Cell
Short Article

The Human Base Excision Repair Enzyme SMUG1 Directly Interacts with DKC1 and Contributes to RNA Quality Control

Laure Jobert,[1] Hanne K. Skjeldam,[1] Bjørn Dalhus,[2] Anastasia Galashevskaya,[3] Cathrine Broberg Vågbø,[3] Magnar Bjørås,[2] and Hilde Nilsen[1,*]
[1]The Biotechnology Centre, University of Oslo, P.O. Box 1125 Blindern, 0317 Oslo, Norway
[2]Department of Microbiology, University of Oslo, Oslo University Hospital, Rikshospitalet, P.O. Box 4950 Nydalen, 0424 Oslo, Norway
[3]Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, 7489 Trondheim, Norway
*Correspondence: hilde.nilsen@biotek.uio.no
http://dx.doi.org/10.1016/j.molcel.2012.11.010

SUMMARY

Single-strand-selective monofunctional uracil-DNA glycosylase 1 (SMUG1) is a base excision repair enzyme that removes uracil and oxidised pyrimidines from DNA. We show that SMUG1 interacts with the pseudouridine synthase Dyskerin (DKC1) and colocalizes with DKC1 in nucleoli and Cajal bodies. As DKC1 functions in RNA processing, we tested whether SMUG1 excised modified bases in RNA and demonstrated that SMUG1 has activity on single-stranded RNA containing 5-hydroxymethyldeoxyuridine, but not pseudouridine, the nucleoside resulting from isomerization of uridine by DKC1. Moreover, SMUG1 associates with the 47S rRNA precursor processed by DKC1, and depletion of SMUG1 leads to a reduction in the levels of mature rRNA accompanied by an increase in polyadenylated rRNA. Depletion of SMUG1, and, in particular, the combined loss of SMUG1 and DKC1, leads to accumulation of 5-hydroxymethyluridine in rRNA. In conclusion, SMUG1 is a DKC1 interaction partner that contributes to rRNA quality control, partly by regulating 5-hydroxymethyluridine levels.

INTRODUCTION

Single-strand-selective monofunctional uracil-DNA glycosylase 1 (SMUG1) (Haushalter et al., 1999) initiates repair of DNA base damage via the base excision repair (BER) pathway. SMUG1 is the main uracil-excision activity in *Ung*$^{-/-}$ mice (Nilsen et al., 2001; Nilsen et al., 2000), and the combined loss of UNG and SMUG1 leads to a dramatic loss of cellular UDG activity (Kemmerich et al., 2012). In addition to uracil, SMUG1 removes several pyrimidine oxidation products (e.g. 5-formyluracil [Masaoka et al., 2003] and 5-carboxyuracil [Darwanto et al., 2009]) and has a specific function to remove the thymine oxidation product 5-hydroxymethyl uracil from DNA (Boorstein et al., 2001; Kemmerich et al., 2012; Masaoka et al., 2003). In contrast to UNG2, which is excluded from nucleoli, SMUG1 has a broad nuclear localization with some enrichment in nucleoli (Kavli et al., 2002). Several DNA repair proteins have been observed in this organelle (Marciniak et al., 1998; Vascotto et al., 2009). As the nucleoli are organelles in which ribosomal RNA (rRNA) synthesis and processing, rather than DNA metabolism, take place (Boisvert et al., 2007), the functional relevance of the nucleolar localization of DNA repair proteins is poorly understood.

Here, we aimed to assess the functional relevance of the SMUG1 localization pattern. We confirmed the nucleolar enrichment, but also observed SMUG1 localization in discrete nuclear spots corresponding to Cajal bodies. Consistent with this localization, we found that SMUG1 directly interacts with DKC1 which, when mutated, causes the severe bone marrow maturation syndrome Dyskeratosis congenita (Dokal, 2011). Interestingly, DKC1 is the main pseudouridine synthase in mammals, which processes nucleolar rRNA and small nuclear RNA (snRNA) species in Cajal bodies. DKC1 has been suggested to mediate the degradation of damaged rRNA by the nuclear exosome (Hoskins and Butler, 2008) and thus to participate in rRNA quality control. As DKC1 functions in RNA processing, we asked whether SMUG1 could excise modified bases in RNA and identified an activity of SMUG1 on 5-hydroxymethyl deoxyuridine [5-hm(dUrd)]-containing single-stranded RNA (ssRNA). Furthermore, we demonstrate a specific in vivo function for SMUG1 in rRNA quality control as SMUG1 associates with 47S rRNA and depletion of SMUG1 leads to the downregulation of the mature rRNA species. Depletion of SMUG1, and, in particular, the combined loss of SMUG1 and DKC1, leads to an accumulation of 5-hm(Urd) in the mature 28S and 18S rRNA species. Hence, we conclude that SMUG1 functions in rRNA quality control in part by regulating 5-hm(Urd) levels in rRNA.

RESULTS

SMUG1 Directly Interacts with DKC1

We reassessed the intracellular localization pattern of SMUG1 in HeLa cells. SMUG1 fused to EYFP was found to localize mainly in the nucleus and be enriched in organelles resembling nucleoli as previously observed (Kavli et al., 2002) (Figure 1A, top panel). SMUG1 also localized in discrete nuclear spots that corresponded to Cajal bodies (Figure 1A, top panel), as is evident by costaining with an antibody specific for Dyskerin (DKC1), an

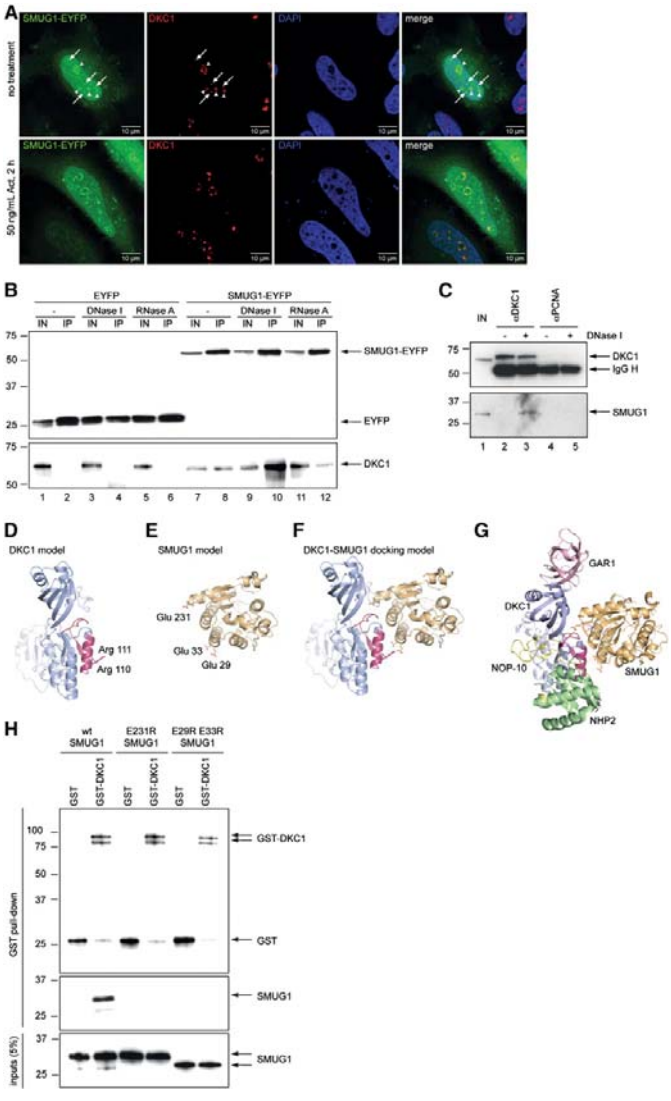


Figure 1. SMUG1 Directly Interacts with DKC1

(A) HeLa cells expressing SMUG1-EYFP were treated or not with Actinomycin D, fixed, and stained with a DKC1-specific antibody and 4′,6-diamidino-2-phenylindole (DAPI). Confocal fluorescent images were obtained on a Zeiss LSM510 confocal microscope. Nucleoli and Cajal bodies are indicated by arrowheads and arrows, respectively.

(B) Coimmunoprecipitations were performed in cells expressing EYFP (lanes 1–6) or SMUG1-EYFP (lanes 7–12). Lysates were treated or not with DNase I or RNase A prior to immunoprecipitation. Coimmunoprecipitated proteins were detected by western blot analysis with EYFP- and DKC1-specific antibodies. IN, 10% input; IP, immunoprecipitate.

(C) Coimmunoprecipitations were performed with antibodies specific for DKC1 (lanes 2 and 3) or PCNA (lanes 4 and 5), as a negative control. Coimmunoprecipitated proteins were detected by western blot analysis with DKC1- and SMUG1-specific antibodies. IN, 10% input; IP, immunoprecipitate; IgG H, heavy chain of Immunoglobulin G.

(D) DKC1 model (blue) with the SMUG1-interacting peptide (amino acids 103–131, magenta). Amino acids used as restriction criteria are indicated.

(E) SMUG1 model (gold) with amino acids used as restriction criteria indicated.

(F) DKC1-SMUG1 docking model.

(G) Three-dimensional model of the interaction between SMUG1 and DKC1 together with the DKC1 partners GAR1 (purple), NOP-10 (yellow), and NHP2 (green).

(H) GST pull-downs were performed with recombinant purified GST or GST-DKC1 as baits and equivalent amounts of recombinant WT, E231R, or E29R E33R SMUG1 proteins. Results were analyzed by SDS-PAGE followed by western blotting analysis with the antibodies specific for GST and SMUG1. The bottom panel shows 5% input of each SMUG1 variant used for the GST pull-down.

See also Figure S1.

established marker for both nucleoli and Cajal bodies. SMUG1 colocalized with DKC1 in both organelles. Colocalization of SMUG1 and DKC1 was lost upon RNA polymerase I inhibition by actinomycin D, which resulted in redistribution of SMUG1 and DKC1 to separate regions of the nucleolar caps (Figure 1A, bottom panel). Thus, colocalization of SMUG1 and DKC1 depended on rRNA biogenesis, which indicated a functional relevance of the interaction.

To confirm that SMUG1 associates with DKC1 in vivo, we performed coimmunoprecipitation (coIP) experiments. IP of overexpressed SMUG1-EYFP significantly recovered DKC1, which was barely detectable in the control IP from extracts prepared from cells expressing EYFP alone (Figure 1B). The SMUG1/DKC1 association was not mediated by nucleic acids since it could be detected in DNase I- and RNase A-treated cells (Figure 1B). We confirmed the SMUG1/DKC1 interaction as an IP of DKC1 pulled down a fraction of endogenous SMUG1 when the lysate was pretreated with DNase I (Figure 1C). Improved association of a DNA repair protein and RNA binding protein after DNase I treatment was previously reported also for the AP-endonuclease 1 (APE1) and NPM1 association (Vascotto et al., 2009) and is believed to be due to the equilibrium of APE1 (and SMUG1) binding to different nucleic acids, being skewed in favor of RNA binding upon treatment with DNase I.

To map the interaction domains of SMUG1 and DKC1, we designed peptide arrays. The DKC1 (1–514) sequence was synthesized as 20-mer peptides and spotted on cellulose membranes offset by 3 amino acids and analyzed for SMUG1 binding by

overlay with purified, recombinant glutathione S-transferase (GST)-tagged SMUG1 protein (Figure S1A available online), followed by anti-GST immunoblotting (Figures S1C and S1D). Five potential SMUG1-binding DKC1 peptides emerged: amino acids 16–29, 112–122, 247–260, 400–410, and 475–491 (Figure S1D, left panel). Similarly, the SMUG1 (1–270) sequence was spotted and analyzed for DKC1 binding (Figure S1C) by incubation with purified GST-DKC1 protein (Figure S1B). Two potential DKC1-binding sequences in SMUG1 were suggested: amino acids 25–35 and 220–233 (Figure S1D, right panel).

Based on the binding studies, we created a structural model of the DKC1/SMUG1 complex using homology models of the two proteins in ZDOCK (Figures 1D and 1E, respectively). The program was run with bias toward docking solutions involving Glu29 and Glu33 in SMUG1 and the residues Arg110 and Arg111 in DKC1. Of the 25 top-ranked models, most showed incompatible poses with steric conflicts between SMUG1 and the NOP-10 subunit in the *Saccharomyces cerevisiae* complex corresponding to the DKC1/NOP-10/GAR1 RNP particle (Zhou et al., 2011). The model selected from the remaining six docking solutions (Figure 1F) involved SMUG1 Glu 231 in the interaction surface, which was among the suggested interacting peptides in the peptide array (Figure S1D) without being included as a restriction criterion during modeling. This model suggested an interaction surface that does not comprise the nucleic acid binding domains of either protein and does not interfere with binding of NOP-10 and GAR1. Moreover, SMUG1 does not interfere with binding of NHP2, which is also part of the H/ACA RNP, as shown by superposition of *Pyrococcus furiosus* structure of all four proteins (Figure 1G) (Li and Ye, 2006).

Mutations expected to disrupt the interaction surface were designed and two mutants, SMUG1 E29R/E33R and SMUG1 E231R, were purified (Figure S1E) and tested for their ability to mediate direct binding to DKC1 in GST pull-down experiments. Purified GST-DKC1 was immobilized on glutathione Sepharose beads and incubated with the purified recombinant proteins. The immunoprecipitated proteins were analyzed by western blotting with GST- and SMUG1-specific antibodies (Figure 1H). While GST-DKC1 efficiently recovered WT SMUG1, it failed to pull down the two SMUG1 mutants, demonstrating that SMUG1 amino acids 29, 33, and 231 are required for binding to DKC1. As an additional control, GST-DKC1 bound to the beads was also detected with a DKC1-specific antibody to confirm the identity of the double band (Figure S1F). As DKC1 is part of a bigger complex (Figure 1G), we tested whether other subunits of the DKC1-containing H/ACA RNP were pulled down with overexpressed SMUG1-EYFP, and we found GAR1 and NHP2 present in the immunoprecipitated complex in the DNase I-treated lysate (Figure S1G).

SMUG1 Has Activity on 5-hm(dUrd)-Containing ssRNA

Both localization of SMUG1 in nucleoli and Cajal bodies and its interaction with the pseudouridine synthase DKC1 raised the hypothesis that SMUG1 may have activity on RNA. As 5-hm(dUrd) is a specific substrate for SMUG1 in DNA, we tested whether SMUG1 might have activity on this modification also in RNA substrates using standard oligonucleotide-nicking assays on 25-mer ssRNA substrates containing a centrally placed 5-hm(dUrd). Incision activity was monitored as the appearance of a faster migrating 12-mer fragment detected by PAGE following incubation with APE1, which has incision activity on AP-site containing RNA (Vascotto et al., 2009). Interestingly, a significant proportion of the 5-hm(dUrd)-containing RNA substrate was cleaved in the presence of recombinant SMUG1 (Figure 2A). No activity was seen on ssRNA containing Urd or pseudo(Urd) [Ψ(Urd)] (Figure 2A). Moreover, we found no activity on dUrd-containing ssRNA, showing that there is no direct overlap between DNA and RNA substrates of SMUG1 (Figure S2A). Thus, specific modifications of the base are required for the ability of SMUG1 to excise a modified base from ssRNA.

To confirm that the incision depended on SMUG1 catalytic activity, we made use of a previously characterized mutation in the catalytic residue histidine 239 that abrogated activity on DNA (Matsubara et al., 2004). The purified recombinant SMUG1 H239L mutant (Figure S1E) exhibited no detectable activity, in contrast to WT SMUG1 (Figure 2B), which reached 21% incision activity on the same substrate (Figure 2C). There was no activity on double-stranded RNA substrates (data not shown). Nicking assays performed on a corresponding 5-hm(dUrd)-containing ssDNA showed that while 100 ng SMUG1 was able to convert 21% of the RNA substrate, about 46% of the DNA substrate was cleaved (Figure S2B). This gave an RNA/DNA activity ratio of 0.44 ± 0.03 calculated from four independent experiments. Thus, the activity of SMUG1 was approximately 2-fold higher on ssDNA than ssRNA under our experimental conditions.

Hence, SMUG1 has activity on ssRNA containing 5-hm(dUrd), but not on substrates containing the functional RNA bases U or ΨU.

SMUG1 Associates with rRNA In Vivo and Contributes to RNA Quality Control

To test whether SMUG1 also associates with RNA species processed by DKC1 in vivo, we established a native RNA coimmunoprecipitation assay to measure the recruitment of SMUG1 to specific RNAs. In this assay, RNA copurified with SMUG1-EYFP was quantified by reverse-transcription quantitative PCR. The presence of RNA species processed by DKC1 in nucleoli (47S, 28S, 18S, and 5.8S rRNAs) and in Cajal bodies (U2 snRNA), as well as the abundant mRNA GAPDH, was quantified. Interestingly, the 47S precursor RNA was found to copurify with SMUG1-EYFP (Figure 3A). Hence, SMUG1 associates with the 47S precursor RNA but not with the processed mature 28S, 18S, and 5.8S rRNA species, U2snRNA, or GAPDH.

To directly explore whether SMUG1 functions in RNA metabolism, we depleted SMUG1 by small interfering RNAs (siRNAs) achieving a knockdown efficiency of more than 90% (Figure S3) and measured the levels of DKC1-associated RNA. The 47S precursor rRNA was equally abundant in SMUG1 depleted as in the control cells, suggesting that rRNA transcription by RNA polymerase I in the nucleolus was unaffected (Figure 3B). Similarly, the spliceosomal U2 snRNA was unchanged in SMUG1-depleted cells, which is consistent with the fact that U2 snRNA was not copurified with SMUG1-EYFP (Figure 3A). The 47S rRNA precursor is subsequently processed to generate the mature 28S, 18S, and 5.8S rRNA species. Consequently, rRNA

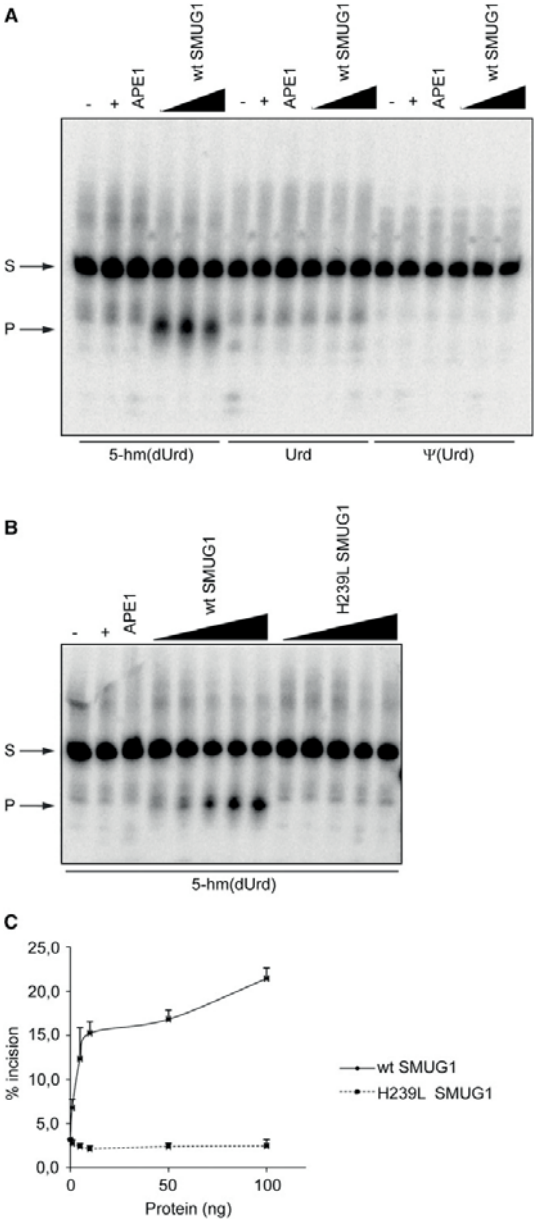


Figure 2. SMUG1 Has Activity on 5-hm(dUrd)-Containing ssRNA

(A) Activity of recombinant wt SMUG1 was assayed on a 25-mer single stranded and 5′-^{32}P-end-labeled oligoribonucleotide substrate containing a centrally placed 5-hm(dUrd), Urd, or Ψ(Urd), as indicated. The substrates were incubated with no enzyme (–), UDG and APE1 (+), APE1 alone, or with increasing amounts of SMUG1 (10, 100, and 200 ng) and APE1. The 12-mer radiolabeled product was resolved by denaturing PAGE and detected by phosphorimager. S, substrate; P, product.

(B) Activity of recombinant WT and mutant H239L SMUG1 was assayed on 5-hm(dUrd)-containing ssRNA. Increasing amounts (1, 5, 10, 50, and 100 ng) of WT and mutant H239L SMUG1 were incubated with the 5′-^{32}P-end-labeled substrate and APE1.

(C) SMUG1 excision activity (%) on 5-hm(dUrd)-containing ssRNA was calculated from three independent experiments and given as the mean ± SD. See also Figure S2.

quality control defects are associated with reduced expression of processed rRNA, which reflects degradation of damaged or inappropriately processed rRNA. The reduced levels of all three mature rRNAs in SMUG1-depleted cells (Figure 3B) were therefore consistent with an in vivo function of SMUG1 in rRNA quality control. The main route for rRNA degradation is the nuclear exosome which involves addition of a poly(A)-tail to the RNA molecule to be degraded (Andersen et al., 2008). Thus, the increased accumulation of polyadenylated 28S rRNA, but not 5.8S or 18S rRNA, in SMUG1-depleted cells is consistent with rRNA degradation in the absence of SMUG1 (Figure 3C). Hence, damaged or inappropriately processed rRNA accumulated in the absence of SMUG1.

SMUG1 and DKC1 Prevent Accumulation of 5-hm(Urd) in 28S and 18S rRNAs In Vivo

As some of the 47S rRNA precursor processing steps are DKC1-dependent, we tested whether SMUG1 might be required for some aspect of DKC1 function. Depletion of SMUG1 by siRNA gave no indication that SMUG1 affected DKC1 stability (Figure 4A). Nor did SMUG1 depletion affect DKC1 activity, as Ψ(Urd) levels in 28S and 18S rRNAs were reduced only in DKC1 knockdown cells (Figure 4B) as previously shown (Jack et al., 2011).

However, quantification of 5-hm(Urd) in 28S and 18S rRNAs by liquid chromatography-tandem mass spectrometry (LC-MS/MS) showed that the 5-hm(Urd) levels were 2.5-fold ($p < 0.005$) higher in 28S and 18S rRNAs when SMUG1 was depleted (Figure 4C). While no significant effect was seen in cells depleted for DKC1 alone (Figure 4C), 28S and 18S rRNA purified from cells depleted for both SMUG1 and DKC1 showed a 3.9- and 4.3-fold ($p < 0.05$), respectively, increased 5-hm(Urd) content. Hence, loss of DKC1 potentiates the effect of SMUG1 on 5-hm(Urd) levels in those rRNAs.

The latter results show that DKC1 somehow facilitates SMUG1 function. As no stimulation of SMUG1 activity was achieved by DKC1 (Figure S4), we tested whether the interaction with DKC1 was important for SMUG1 localization by transfecting cells with the SMUG1 E29R/E33R mutant, which is active (Figure S5) but unable to interact with DKC1 (Figure 1H). These cells still express DKC1 (Figure 4D) but unlike SMUG1-EYFP, the SMUG1 E29R/E33R mutant was diffusely located in the nucleoplasm and not enriched in nucleoli and Cajal bodies. Moreover, cells transfected with this mutant failed to show the expected DKC1 staining pattern in nucleoli (Figure 4D). Comparison of two neighboring cells, one that did not express the mutated construct and one that did, clearly showed that while DKC1 has the expected localization pattern in both nucleoli (arrow heads) and Cajal bodies (arrows) in the untransfected control (Figure 4E), localization of DKC1 within nucleoli was perturbed

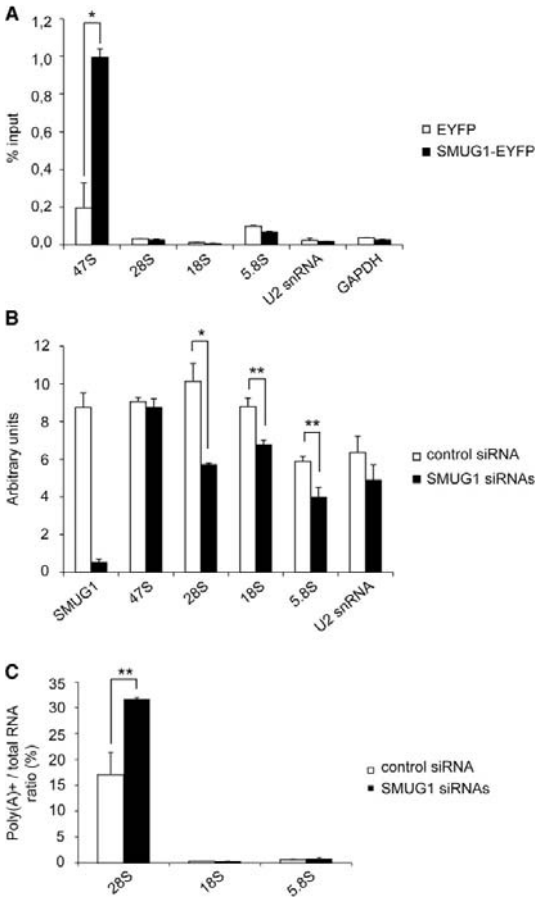


Figure 3. SMUG1 Contributes to rRNA Quality Control

(A) Native RNA coimmunoprecipitations were performed in cells overexpressing SMUG1-EYFP or EYFP alone. Reverse-transcription quantitative PCRs were performed with primers specific for the indicated RNAs. The data shown are the mean ± SEM from two independent experiments. Statistical significance was evaluated with the Student t test. *$p < 0.02$.

(B) After transfection with control or SMUG1 siRNAs for 48 hr, total RNA was purified and analyzed by reverse-transcription quantitative PCR with specific primers, as indicated. The data shown are the mean ± SEM from two independent experiments. Statistical significance was evaluated with the Student t test. *$p < 0.005$, **$p < 0.05$.

(C) Total and poly(A)$^+$ RNA were prepared from cells transfected with control or SMUG1 siRNAs. Equal volumes of RNA samples were used as template in the RT reactions containing either random hexamers or oligo d(T) primers. Equal amounts of cDNAs were used in reverse-transcription quantitative PCR reactions with primers specific for the indicated RNAs. The results shown are the average of four PCRs from two independent RNA extractions. Statistical significance was evaluated using the Student t test. **$p < 0.05$.

See also Figure S3.

in the presence of the SMUG1 mutant (Figure 4F). Hence, disruption of the SMUG1/DKC1-interaction surface perturbed proper localization of both SMUG1 and DKC1.

DISCUSSION

In the present study we show that SMUG1 interacts directly with DKC1 and contributes to rRNA quality control in vivo. We demonstrate that SMUG1 has activity on 5-hm(dUrd)-containing ssRNA which adds significantly to the increasing body of evidence linking DNA repair proteins to different aspects of RNA metabolism (Tell et al., 2010). Moreover, we characterized DKC1 as a SMUG1-interacting protein, and SMUG1 amino acids involved in the binding to DKC1 were identified and experimentally verified (Figure 1). The SMUG1 mutants E29R/E33R and E231R were still active (Figure S5), strongly indicating correct folding. Both mutants nevertheless failed to bind purified recombinant DKC1 (Figure 1H), strongly arguing that the SMUG1/DKC1 interaction is direct.

In line with the main role of SMUG1 being in DNA repair, the in vivo data showed that a fraction of SMUG1 colocalized with DKC1. The predicted interaction surface was compatible with the mutual binding of SMUG1 and the GAR1, NOP-10, and NHP2 subunits of the H/ACA RNP particle to DKC1 (Figure 1G). Consistently, SMUG1-EYFP pulled down GAR1 and NHP2 (Figure S1G). Moreover, as the interaction surface did not comprise the substrate binding domains of either protein, in vitro nicking assays confirmed that SMUG1 activity was not inhibited by addition of DKC1 (Figure S4) and LC-MS/MS data showed that SMUG1 depletion did not affect Ψ(Urd) levels in 28S and 18S rRNAs (Figure 4B). The available evidence therefore suggests that the interaction with DKC1 may serve to target, or tether, SMUG1 to a select group of RNA substrates in the nucleoli and Cajal bodies. In line with this interpretation, we found strong association of SMUG1 only with the precursor 47S rRNA in vivo (Figure 3A) although SMUG1 associated with several RNA species after formaldehyde crosslinking in an RNA-IP assay (Figure S6). A specific role of SMUG1 in rRNA quality control was substantiated by the demonstration of a reduced amount of mature rRNA accompanied with increased polyadenylation of 28S rRNA in SMUG1-depleted cells. Hence, damaged or inappropriately processed rRNA accumulated in the absence of SMUG1. The observation that the 5.8S and 18S levels were reduced without being polyadenylated may be explained by the polarity of the rRNA locus transcribed as a single pre-rRNA by RNA polymerase I, as previously suggested (Fang et al., 2004). The 28S rRNA is located at the 3′ side of the locus and is, thus, the first rRNA being produced by the oligo(dT)-primed reverse transcriptase. Therefore, the inability of the enzyme to reach the end of the long polyadenylated rRNA template may explain the absence of polyadenylated 18S and 5.8S rRNAs. Alternatively, different mature rRNA species may be substrates for different RNA-degradation pathways (Andersen et al., 2008), as suggested from studies in *S. cerevisiae* (Hoskins and Butler, 2008).

The fact that we saw increased polyadenylation and decreased levels of mature 28S rRNA in SMUG1-depleted cells strongly indicated that SMUG1 has a normal function in rRNA surveillance. The accumulation of 5-hm(Urd) in mature rRNA upon SMUG1 knockdown showed that 5-hm(Urd) is an in vivo substrate for SMUG1. Taken together, this suggests that

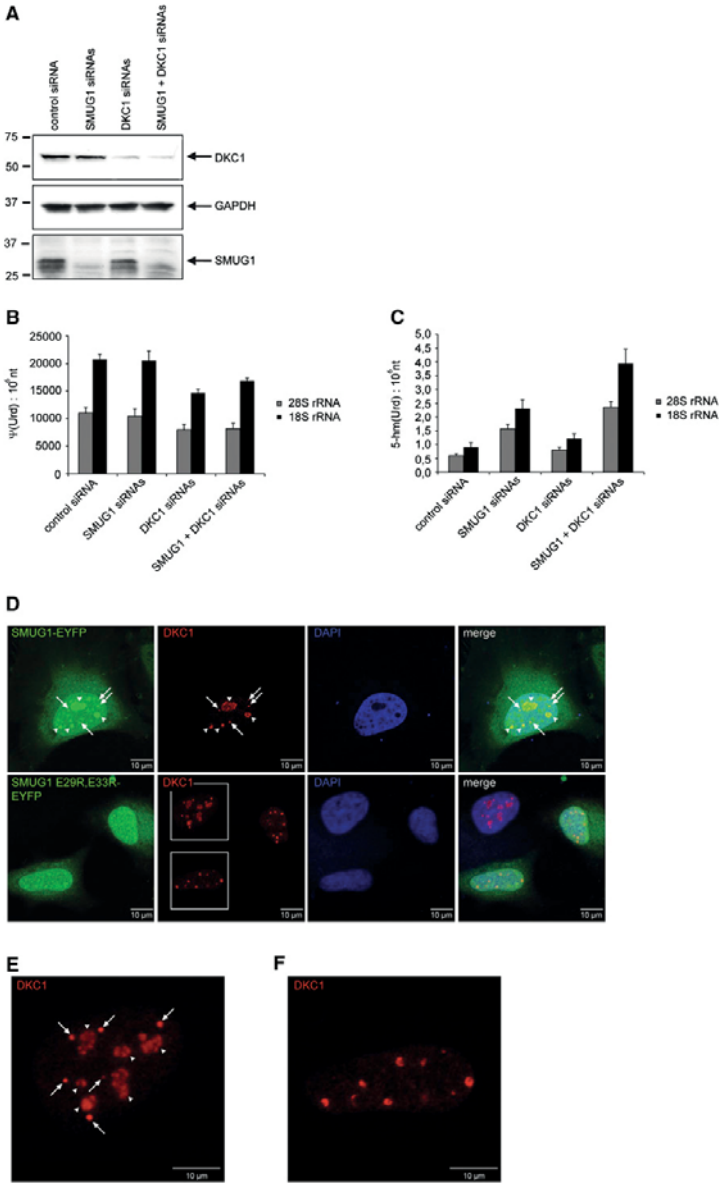


Figure 4. SMUG1 and DKC1 Prevent Accumulation of 5-hm(Urd) in 28S and 18S rRNAs In Vivo

(A) HeLa cells were transfected with control, SMUG1, and/or DKC1 siRNAs for 48 hr and whole-cell extracts were subjected to western blot analysis using DKC1-, GAPDH-, and SMUG1-specific antibodies.

(B and C) Quantification of incorporated Ψ(Urd) and 5-hm(Urd) per nucleotide RNA. Cells were harvested and 28S and 18S rRNA species were isolated, hydrolyzed, and analyzed for Ψ(Urd) or 5-hm(Urd) content by LC-MS/MS. Ψ(Urd) or 5-hm(Urd) levels are normalized relative to the total number of normal nucleosides measured. The data shown are the mean ± SD from three independent experiments.

(D–F) Cells expressing SMUG1-EYFP or SMUG1 E29R E33R-EYFP were fixed and stained with a DKC1-specific antibody and DAPI (D). Close-up of two neighboring cells from that do not (E) or do (F) express the SMUG1 E29R E33R-EYFP construct. Confocal fluorescent images were obtained by a Zeiss LSM510 confocal microscope. Nucleoli and Cajal bodies are indicated by arrowheads and arrows, respectively.

5-hm(Urd) modified rRNA may be degraded by the exosome in the absence of SMUG1. The source of 5-hm(Urd) in RNA is not known. However, 5-hm(Cyt) is a natural modification in 18S and 26S rRNA in eukaryotes (Rácz et al., 1978), and 5-hm(Urd) may therefore result from hydrolytic deamination of 5-hm(Cyt). Alternatively, it can result from incorporation of 5-hm(dUrd)/5-hm(Urd) recycled from damaged DNA, as previously suggested (Pettersen et al., 2011). Although 5-hm(dUrd) was not detected, our LC-MS/MS data demonstrate that 5-hm(Urd) is present in mature rRNA in vivo.

In summary, the BER enzyme SMUG1 is a DKC1 interaction partner. The SMUG1/DKC1 interaction targets the complex to nucleoli, where it contributes to rRNA quality control, in part by regulating the level of 5-hm(Urd).

EXPERIMENTAL PROCEDURES

RNA Coimmunoprecipitation Assay

Approximately 325 μg dialyzed whole-cell extract was subjected to immunoprecipitation with an anti-GFP antibody (Roche). After incubation with protein G Sepharose and stringent washes (see the Supplemental Experimental Procedures), immune complexes were treated with 20 U DNase I-RNase free (Fermentas) for 30 min at 37°C in the presence of 20 U RNaseOUT recombinant ribonuclease inhibitor (Invitrogen). The reaction was stopped with 2.5 mM EDTA for 10 min at 65°C. Then, immunoprecipitates were treated with 50 μg proteinase K (QIAGEN) for 50 min at 50°C. Coimmunoprecipitated RNA was purified by phenol/chloroform/isoamyl alcohol extraction, precipitated with ethanol for 2 hr at −20°C and analyzed by reverse-transcription quantitative PCR.

Oligonucleotide Nicking Assay

Standard oligonucleotide nicking assays were performed essentially as previously described (Nilsen et al., 2001). In brief, 3.5 pmol end-labeled 25-mer RNA oligonucleotide were reacted with purified wild-type or mutant hSMUG1 in the reaction buffer (20 mM Tris-HCl [pH 7.5], 60 mM NaCl, 1 mM EDTA, 0.7 mg/ml

BSA, and 1 mM DTT) for 30 min at 37°C for 30 min before addition of 5 U APE-1 at 37°C for 30 min. The products were run on an 8% denaturing polyacrylamide gel.

LC-MS/MS Analysis of 5-hmUrd and PseudoUrd

RNA was enzymatically hydrolyzed to nucleosides essentially as described (Crain, 1990), with 3 volumes of methanol added, and centrifuged (16,000 g, 20 min, 4°C). The supernatants were dried and dissolved in 50 μl 5% methanol in water (v/v) for LC-MS/MS analysis. Chromatographic separation was performed on a Shimadzu Prominence HPLC system with a Zorbax SB-C18 2.1 × 150 mm internal diameter (i.d.) (3.5 μm) column equipped with an Eclipse XDB-C8 2.1x12.5 mm i.d. (5 μm) guard column (Agilent Technologies). Mass spectrometry detection was performed with an MDS Sciex API5000 triple quadrupole (Applied Biosystems) operating in negative electrospray ionization mode for modified nucleosides, and positive electrospray ionization mode for unmodified nucleosides, monitoring the mass transitions 273.1/140.1 (5-hmUrd, quantifier ion), 273.1/230.1 (5-hmUrd, qualifier ion), 243.1/183.1 (pseudoUrd, quantifier ion), 243.1/153.1 (pseudoUrd, qualifier ion), 268.1/136.1 (Ado), 244.1/112.1 (Cyt), 284.1/152.1 (Guo), and 245.1/113.1 (Urd).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and six figures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2012.11.010.

ACKNOWLEDGMENTS

Funding was provided by The Research Council of Norway FRIBIO (185911/V40) and Functional Genomics (183330/S10) programs. We thank Ian G. Mills for sharing equipment, Hans E. Krokan for SMUG1 antibody, and Thomas U. Meier for NHP2 antibody. B.D. and M.B. were supported by the South-Eastern Norway Regional Health Authority (grant numbers 2009100 and 2011040) for establishing the Regional Core Facility for Structural Biology and Bioinformatics. C.B.V. and A.G. were supported by the Research Council of Norway grant numbers 185308 and 191408, respectively.

Received: June 1, 2012
Revised: October 18, 2012
Accepted: November 7, 2012
Published: December 13, 2012

REFERENCES

Andersen, K.R., Jensen, T.H., and Brodersen, D.E. (2008). Take the "A" tail—quality control of ribosomal and transfer RNA. Biochim. Biophys. Acta *1779*, 532–537.

Boisvert, F.M., van Koningsbruggen, S., Navascués, J., and Lamond, A.I. (2007). The multifunctional nucleolus. Nat. Rev. Mol. Cell Biol. *8*, 574–585.

Boorstein, R.J., Cummings, A., Jr., 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.

Crain, P.F. (1990). Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. Methods Enzymol. *193*, 782–790.

Darwanto, A., Theruvathu, J.A., Sowers, J.L., Rogstad, D.K., Pascal, T., Goddard, W., 3rd, and Sowers, L.C. (2009). Mechanisms of base selection by human single-stranded selective monofunctional uracil-DNA glycosylase. J. Biol. Chem. *284*, 15835–15846.

Dokal, I. (2011). Dyskeratosis congenita. Hematology (Am Soc Hematol Educ Program) *2011*, 480–486.

Fang, F., Hoskins, J., and Butler, J.S. (2004). 5-fluorouracil enhances exosome-dependent accumulation of polyadenylated rRNAs. Mol. Cell. Biol. *24*, 10766–10776.

Haushalter, K.A., Todd Stukenberg, M.W., Kirschner, M.W., and Verdine, G.L. (1999). Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. Curr. Biol. *9*, 174–185.

Hoskins, J., and Butler, J.S. (2008). RNA-based 5-fluorouracil toxicity requires the pseudouridylation activity of Cbf5p. Genetics *179*, 323–330.

Jack, K., Bellodi, C., Landry, D.M., Niederer, R.O., Meskauskas, A., Musalgaonkar, S., Kopmar, N., Krasnykh, O., Dean, A.M., Thompson, S.R., et al. (2011). rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells. Mol. Cell *44*, 660–666.

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.

Li, L., and Ye, K. (2006). Crystal structure of an H/ACA box ribonucleoprotein particle. Nature *443*, 302–307.

Marciniak, R.A., Lombard, D.B., Johnson, F.B., and Guarente, L. (1998). Nucleolar localization of the Werner syndrome protein in human cells. Proc. Natl. Acad. Sci. USA *95*, 6887–6892.

Masaoka, A., Matsubara, M., Hasegawa, R., Tanaka, T., Kurisu, S., Terato, H., Ohyama, Y., Karino, N., Matsuda, A., and Ide, H. (2003). 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 *42*, 5003–5012.

Matsubara, M., Tanaka, T., Terato, H., Ohmae, E., Izumi, S., Katayanagi, K., and Ide, H. (2004). Mutational analysis of the damage-recognition and catalytic mechanism of human SMUG1 DNA glycosylase. Nucleic Acids Res. *32*, 5291–5302.

Nilsen, H., Rosewell, I., Robins, P., Skjelbred, C.F., Andersen, S., Slupphaug, G., Daly, G., Krokan, H.E., Lindahl, T., and Barnes, D.E. (2000). Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication. Mol. Cell *5*, 1059–1065.

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.

Pettersen, H.S., Visnes, T., Vågbø, C.B., Svaasand, E.K., Doseth, B., Slupphaug, G., Kavli, B., and Krokan, H.E. (2011). UNG-initiated base excision repair is the major repair route for 5-fluorouracil in DNA, but 5-fluorouracil cytotoxicity depends mainly on RNA incorporation. Nucleic Acids Res. *39*, 8430–8444.

Rácz, I., Király, I., and Lásztity, D. (1978). Effect of light on the nucleotide composition of rRNA of wheat seedlings. Planta *142*, 263–267.

Tell, G., Wilson, D.M., 3rd, and Lee, C.H. (2010). Intrusion of a DNA repair protein in the RNome world: is this the beginning of a new era? Mol. Cell. Biol. *30*, 366–371.

Vascotto, C., Fantini, D., Romanello, M., Cesaratto, L., Deganuto, M., Leonardi, A., Radicella, J.P., Kelley, M.R., D'Ambrosio, C., Scaloni, A., et al. (2009). APE1/Ref-1 interacts with NPM1 within nucleoli and plays a role in the rRNA quality control process. Mol. Cell. Biol. *29*, 1834–1854.

Zhou, J., Liang, B., and Li, H. (2011). Structural and functional evidence of high specificity of Cbf5 for ACA trinucleotide. RNA *17*, 244–250.

Molecular Cell, Volume *49*

Supplemental Information

The Human Base Excision Repair Enzyme SMUG1 Directly Interacts with DKC1 and Contributes to RNA Quality Control

Laure Jobert, Hanne K. Skjeldam, Bjørn Dalhus, Anastasia Galashevskaya, Cathrine Broberg Vågbø, Magnar Bjørås, and Hilde Nilsen

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cell lines and treatments

HeLa cells were grown in DMEM (4.5 g / L glucose, GlutaMAX™) containing 10% FBS, 100 U / mL penicillin 100 µg / mL streptomycin, 0.1 mM non essential amino acids solution and 1 mM Na pyruvate at 37 °C and 5 % CO_2. Inhibition of RNA polymerase I was achieved by incubating HeLa cells with 50 ng / mL actinomycin D (Sigma-Aldrich, A9415) for 2 h.

Transfections with DNA or siRNA

HeLa cells were transfected for 24 h with either pEYFP-N1-SMUG1 or pEYFP-N1 (Clontech) plasmids using the FuGENE® 6 (Roche) transfection reagent according to the manufacturer's instructions. HeLa cells were transfected for 48 h with siRNAs using the Lipofectamine™ RNAiMAX (Invitrogen) transfection reagent according to the manufacturer's instructions. For SMUG1 knock-down, a mix of two siRNAs (final concentration 25 mM each) was used. siRNAs were purchased from Ambion: *Silencer*® Negative Control #1 siRNA (AM4611), SMUG1 siRNA ID # 21109 and SMUG1 siRNA ID # 21193, DKC1 siRNA ID # s4110, DKC1 siRNA ID # s 4111 and DKC1 siRNA ID # s4112.

Whole cell and nuclear extractions

Approximately 3.10^6 cells were harvested and washed twice with ice-cold PBS. For whole cell extraction, cells were suspended in 200 µL lysis buffer (20 mM Tris-HCl pH 7.5, 400 mM KCl, 20 % glycerol, 1 mM DTT, 1X Complete EDTA-free protease inhibitor cocktail) and incubated on ice for 10 min before three freeze-thaw cycles in liquid N_2 and ice. Cellular debris were discarded by centrifugation at 15000 g for 15 min at 4 °C and the whole cell extract was dialysed overnight at 4 °C against 1 L dialysis buffer (25 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 100 mM KCl, 10 % glycerol, 1 mM DTT, 0.5 mM PMSF). For nuclear extraction, cells were washed with 200 µL isotonic buffer (20 mM HEPES-NaOH pH 7.8, 1 mM $MgCl_2$, 5 mM KCl, 250 mM sucrose, 1 mM DTT). Cells were suspended in 200 µL hypotonic buffer (20 mM HEPES-NaOH pH 7.8, 1 mM $MgCl_2$, 5 mM KCl, 1 mM DTT, 1X Complete EDTA-free protease inhibitor cocktail) and incubated on ice for 45 min. Cells were lysed by 15 strokes in a dounce homogenizer (B pestle) and centrifuged at 600 g for 5 min at 4 °C. The nuclear pellet was suspended in 200 µL hypertonic buffer (20 mM HEPES-NaOH pH 7.8, 1 mM $MgCl_2$, 5 mM KCl, 0.5 M NaCl, 25 % glycerol, 1 mM DTT, 1X Complete EDTA-free protease inhibitor cocktail) and incubated on ice for 30 min. Nuclear debris were discarded by centrifugation at 15000 g for 15 min at 4 °C and the nuclear extract was dialysed overnight at 4 °C against 1 L dialysis buffer.

Immunoprecipitation and Western blotting

Approximately 250 µg cell lysate was used per immunoprecipitation. The cell lysate was preliminary treated with 40 U DNase I-RNase free or 25 µg / mL RNase A for 30 min at 37 °C.

The cell lysate was then incubated with 10 µL anti-GFP antibody (Roche, Cat. No. 11 814 460 001) overnight at 4 °C. The following day, 20 µL Protein G Sepharose™ 4 Fast Flow (GE Healthcare) pre-equilibrated with IP buffer (25 mM Tris-HCl pH 7.9, 5 mM $MgCl_2$, 10 % glycerol, 0.1 % NP-40, 1 mM DTT, 1X Complete EDTA-free protease inhibitor cocktail) containing 100 mM KCl was added and gently mixed for 2 h at 4 °C. Immune complexes were washed three times with IP buffer containing 150 mM KCl and two times with IP buffer containing 100 mM KCl, each for 5 min at 4 °C. Immunoprecipitates were boiled in SDS-sample buffer (25 mM Tris-HCl pH 6.8, 2 % SDS, 10 % glycerol, 0.05 % bromophenol blue, 5 % 2-mercaptoethanol) for 5 min, separated by SDS-polyacrylamide gel electrophoresis (PAGE) and subject to Western blot analysis, performed as follows. Nitrocellulose membranes were blocked in PBS containing 5 % non-fat dry milk for 1 h at room temperature and incubated with the following primary antibodies: anti-GFP (Roche, 1:1000), anti-dyskerin (Santa Cruz Biotechnology, sc-48794, 1:1000), anti-SMUG1 (PSM1, kind gift from H. E. Krokan, 1:1000 or sc-26880, Santa Cruz Biotechnology, 1:200), anti-GAR1 (Proteintech, Cat No 11711-1-AP) or anti-NHP2 (CG2 rabbit polyclonal antiserum, kind gift from U. T. Meier). Secondary antibodies were horseradish peroxidase-conjugated anti-mouse (Santa Cruz Biotechnology, sc-2314, 1:5000) and anti-rabbit (Cell Signaling, #7074, 1:2000) immunoglobulin Gs.

GST pull-down

Approximately 20 µL Glutathione 4B Sepharose (GE Healthcare) were incubated with 20 pmol recombinant GST-DKC1 protein or with GST for 1 h at 4 °C on a rotary shaker. After washing with IP buffer containing 100 mM KCl, 50 pmol recombinant SMUG1 were added to the beads and incubated for 2 h at 4 °C on a rotary shaker. The beads were then washed three times with IP buffer containing 100 mM KCl, resuspended in 20 µL SDS-sample buffer and subject to Western blot analysis.

Immunofluorescence

Cells were plated onto coverslips, fixed with 4 % paraformaldehyde for 5 min at room temperature and permeabilized with PBS containing 0.1 % Triton X-100, 0.01 % Tween-20 and 2 % BSA for 15 min. Cells were incubated with the anti-DKC1 antibody (1:50) for 5 h at room temperature, washed three times with PBS containing 0.1 % Triton X-100, each for 5 min. Cells were incubated with the secondary antibody, Alexa Fluor 555 goat anti-rabbit IgG (Invitrogen, 1:500) for 45 min at room temperature. DNA was labelled with 0.25 mg / mL 4,6-diamidino-2-phenylindole (DAPI). Confocal fluorescent images were obtained using a LSM 510 META (Carl Zeiss) microscope with a 63x objective.

RNA purification and qPCR

Total RNA was purified with TRIzol® reagent (Invitrogen) according to manufacturer's instructions. Reverse transcription was performed with SuperScript™ II Reverse Transcriptase (Invitrogen). Quantitative PCR was carried out on a MyiQ real-time PCR detection system with iQ™ SYBR® Green (Bio-Rad). Primers used are available upon request.

SMUG1 overexpression and purification

E. coli BL21(DE3) harboring pGEX-4T1-hSMUG1 (or empty pGEX-4T1) was grown in LB media (2 L) containing ampicillin (100 µg / mL) at 37 °C until the OD_{600} reached 0.6. After addition of isopropyl β-D-thiogalactopyranoside (final concentration 0.5 mM), the cell culture was continued at 20 °C for 6 h. The following procedures were performed at 4 °C or on ice. Cells were harvested by centrifugation, suspended in 20 mL buffer A (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 % glycerol, 1 mM DTT, 1X Complete EDTA-free protease inhibitor cocktail) and lysed by the addition of 1 mg / mL lysozyme. Cells were incubated for 30 min at 4 °C with gentle shaking. The cell lysate was supplemented with 0.5 % NP-40, 5 mM $MgCl_2$ and 40 µg / mL DNase I, incubated for an additional 30 min at 4 °C with gentle shaking and centrifuged (10000 g,

30 min). To the clarified supernatant was added 0.5 ml (bed volume) of Glutathione 4B Sepharose (GE Healthcare), pre-equilibrated in PBS. The cell lysate and glutathione resin were allowed to mix for 1 h using a mechanical rotary platform. For purification of the hSMUG1 protein, the resin was washed three times with ice-cold PBS and incubated with 0.5 mL elution buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 2.5 mM $CaCl_2$, 1 mM DTT) containing 150 U thrombin (Sigma-Aldrich, T4648) overnight at 4 °C. The eluate was collected and the resin was incubated with 0.5 mL elution buffer without thrombin for 1 h at 4 °C. The two eluates were combined, dialysed against 1 L dialysis buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM DTT, 1 mM EDTA, 0.5 mM PMSF) overnight at 4 °C. The digested sample was loaded onto a 1 ml Resource Q FPLC column (Amersham Biosciences), washed with dialysis buffer and then eluted with a linear gradient of NaCl (50-1000 mM) in dialysis buffer. Fractions containing hSMUG1 were collected, combined and concentrated by ultrafiltration (Amicon Ultra-0.5 mL, 10K, Millipore). The purified protein was stored in 50 % glycerol at -20 °C.

For purification of the GST-hSMUG1 fusion protein, the resin was washed three times with ice-cold PBS and incubated with 0.5 mL glutathione elution buffer (15 mM reduced glutathione, 100 mM Tris-HCl pH 8.0, 200 mM NaCl, 0.1 % Triton X-100, 1 mM DTT, 1X Complete EDTA-free protease inhibitor cocktail) overnight at 4 °C. The eluate was collected and the resin was incubated with 0.5 mL glutathione elution buffer for an additional 30 min at 4 °C. The two eluates were combined, concentrated by ultrafiltration (Ultrafree®-0.5, 30K, Millipore) and dialysed against 1 L dialysis buffer overnight at 4 °C. The purified protein was stored in 50 % glycerol at -20 °C.

GST-DKC1 overexpression and purification

E. coli BL21(DE3)-RIL (Stratagene) harboring pGEX-4T1-hDKC1 (or empty pGEX-4T1) was grown in LB media (2 L) containing 100 µg / mL ampicillin at 37 °C until the OD_{600} reached 0.6. After addition of isopropyl β-D-thiogalactopyranoside (final concentration 0.5 mM), the cell culture was continued at 20 °C for 4 h. Bacterial lysis, GST batch purification and elution with reduced glutathione were performed as for GST-hSMUG1. The eluate was dialysed against 1 L dialysis buffer (50 mM HEPES pH 8, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.5 mM PMSF) overnight at 4 °C. The dialysed sample was loaded onto a 1 ml HiTrap SP FF FPLC column (Amersham Biosciences), washed with dialysis buffer and then eluted with a stepwise gradient at 600 mM NaCl. Fractions containing GST-hDKC1 were collected, combined and concentrated by ultrafiltration (Amicon Ultra-0.5 mL, 10K, Millipore). The buffer was exchanged with dialysis buffer. The purified protein was stored in 10 % glycerol and snap-frozen at -80 °C.

Peptide arrays

Peptide arrays were synthesized on cellulose membranes using a MultiPep automated peptide synthesizer (INTAVIS Bioanalytical Instruments AG). Peptide interactions with GST and GST fusion proteins were determined by overlaying the cellulose membranes with 1 µg / mL protein. Bound proteins were detected with the horseradish peroxidase-conjugated anti-GST antibody (GE Healthcare, RPN1236) and visualized by ECL.

Structure modeling and protein-protein docking

Homology models of human SMUG1 and DKC1 were built using Phyre2.0 (Kelley and Sternberg, 2009). The models, based on templates 1oe4.pdb (Wibley et al., 2003) and 2apo.pdb (Hamma et al., 2005), respectively, were used for protein-protein docking by the ZDOCK Fast Fourier Transform docking program (Chen et al., 2003). The ZDOCK program was run with preferences for docking poses with residues Glu29 and Glu33 in SMUG1 and residues Arg110 and Arg111 in DKC1 located in the protein-protein interface. All crystal structure figures were designed using PyMOL (DeLano Scientific).

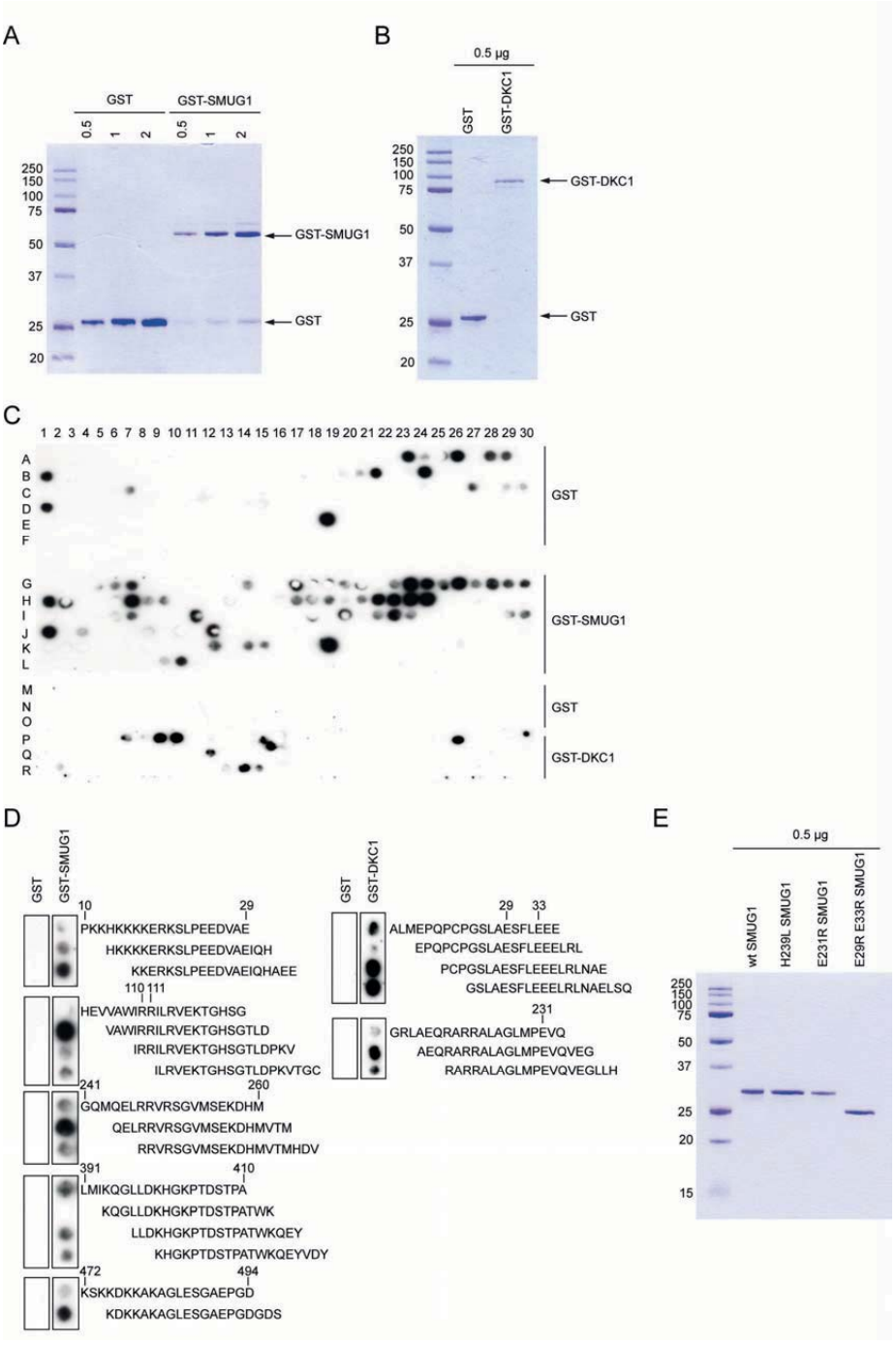
Oligonucleotide nicking assay

3.5 pmol of each 25-mer oligonucleotide (see sequences below) was 5′-end labelled using T4 polynucleotide kinase (Sigma) and [γ-^{32}P]ATP (PerkinElmer). Unincorporated [γ-^{32}P]ATP was removed by centrifugation on a MicroSpin G50 spin column (GE Healthcare) for DNA substrates and on a NucAway™ spin column (Ambion) for RNA substrates. The standard assay mixture for uracil-DNA glycosylase activity contained 5′-end labelled oligonucleotide, nuclear extracts or purified wild type or mutant hSMUG1 in the reaction buffer (20 mM Tris-HCl pH 7.5, 60 mM NaCl, 1 mM EDTA, 0.7 mg / mL BSA, 1 mM DTT). The reaction mixture was incubated at 37°C for 30 min followed by cleavage of abasic sites by 5 U APE-1 at 37 °C for 30 min. The products of the reactions were supplemented with 1 volume STOP solution (80 % formamide, 7 M urea, 10 mM EDTA, 0.025 % xylene cyanol), boiled for 5 min and chilled on ice. The products were run on an 8 % denaturing polyacrylamide gel, visualized on a Typhoon PhosphorImager and quantified using the ImageQuant 5.1 software (Molecular Dynamics, Sunnyvale, CA).

The sequences of RNA substrates are the following: RNA-U25, 5'-rCrCrArCrArCrArArArGrGrGrUrArArArGrCrCrGrGrGrGrCrA-3'; RNA-(dUrd)25, 5'-rCrCrArCrArCrArArArGrGrGdUrArArArGrCrCrGrGrGrGrCrA-3'; RNA-5-hm(dUrd)25, 5'-rCrCrArCrArCrArArArGrGrG5-hmdUrArArArGrCrCrGrGrGrGrCrA-3'; RNA-ψU25, 5'-rCrCrArCrArCrArArArGrGrGrΨUrArArArGrCrCrGrGrGrGrCrA-3'.

LC-MS/MS

RNA was enzymatically hydrolyzed to nucleosides essentially as described (Crain, 1990), added 3 volumes of methanol and centrifuged (16,000 g, 20 min, 4°C). The supernatants were dried and dissolved in 50 µl 5% methanol in water (v/v) for LC-MS/MS analysis of 5-hmUrd, pseudoUrd, and unmodified nucleosides. Chromatographic separation was performed on a Shimadzu Prominence HPLC system with a Zorbax SB-C18 2.1x150 mm i.d. (3.5 µm) column equipped with an Eclipse XDB-C8 2.1x12.5 mm i.d. (5 µm) guard column (Agilent Technologies). For modified nucleosides, the mobile phase consisted of water and methanol, for 5-hmUrd starting with a 4.4-min gradient of 5-60 % methanol, followed by 7-min re-equilibration with 5 % methanol, and for pseudoUrd starting with a 2.5-min gradient of 5-35% methanol, followed by 5 min re-equilibration with 5 % methanol. For unmodified nucleosides, a mobile phase of water and methanol added 0.1 % formic acid was maintained isocratically with 30 % methanol. Mass spectrometry detection was performed using an MDS Sciex API5000 triple quadrupole (Applied Biosystems) operating in negative electrospray ionization mode for modified nucleosides, or positive electrospray ionization mode for unmodified nucleosides, monitoring the mass transitions 273.1/ 140.1 (5-hmUrd, quantifier ion), 273.1/230.1 (5-hmUrd, qualifier ion), 243.1/183.1 (pseudoUrd, quantifier ion), 243.1/153.1 (pseudoUrd, qualifier ion), 268.1/136.1 (Ado), 244.1/112.1 (Cyt), 284.1/152.1 (Guo), and 245.1/113.1 (Urd).


A
GST
GST-SMUG1
0.5
1
2
0.5
1
2
250
150
100
75
50
37
25
20
GST-SMUG1
GST
B
0.5 µg
GST
GST-DKC1
250
150
100
75
50
37
25
20
GST-DKC1
GST
C
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
A
B
C
D
E
F
GST
G
H
I
J
K
L
GST-SMUG1
M
N
O
P
Q
R
GST
GST-DKC1
D
GST
GST-SMUG1
10
29
PKKHKKKKERKSLPEEDVAE
HKKKKERKSLPEEDVAEIQH
KKERKSLPEEDVAEIQHAEE
110 111
HEVVAWIRRILRVEKTGHSG
VAWIRRILRVEKTGHSGTLD
IRRILRVEKTGHSGTLDPKV
ILRVEKTGHSGTLDPKVTGC
241
260
GQMQELRRVRSGVMSEKDHM
QELRRVRSGVMSEKDHMVTM
RRVRSGVMSEKDHMVTMHDV
391
410
LMIKQGLLDKHGKPTDSTPA
KQGLLDKHGKPTDSTPATWK
LLDKHGKPTDSTPATWKQEY
KHGKPTDSTPATWKQEYVDY
472
494
KSKKDKKAKAGLESGAEPGD
KDKKAKAGLESGAEPGDGDS
GST
GST-DKC1
29
33
ALMEPQPCPGSLAESFLEEE
EPQPCPGSLAESFLEEELRL
PCPGSLAESFLEEELRLNAE
GSLAESFLEEELRLNAELSQ
231
GRLAEQRARRALAGLMPEVQ
AEQRARRALAGLMPEVQVEG
RARRALAGLMPEVQVEGLLH
E
0.5 µg
wt SMUG1
H239L SMUG1
E231R SMUG1
E29R E33R SMUG1
250
150
100
75
50
37
25
20
15

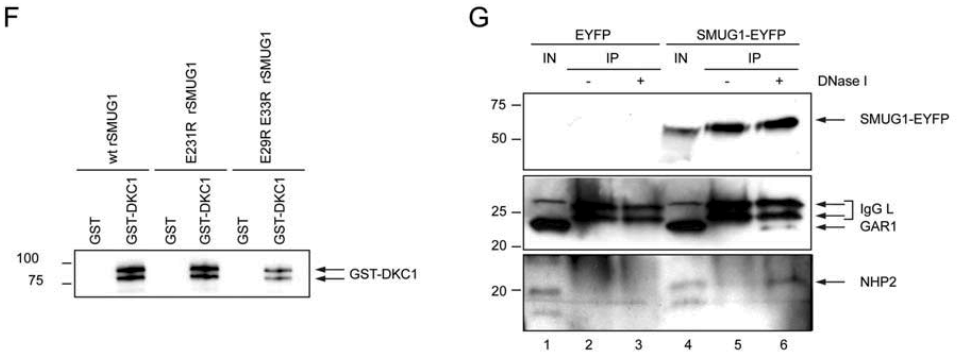


Figure S1. SMUG1 directly interacts with DKC1, Related to Figure 1

A. Purified recombinant human GST and GST-SMUG1 proteins were analysed by PAGE and Coomassie staining. 0.5, 1 or 2 µg of each protein sample was loaded. **B**. Purified GST and GST-DKC1 proteins were analysed by PAGE and Coomassie staining. **C**. Peptide arrays harboring either DKC1 (duplicates, lanes A-F and G-L) or SMUG1 peptides (duplicates, lanes M-O and P-R) were analysed for GST/GST-SMUG1 or GST/GST-DKC1 binding, respectively. **D**. Selected spots from the peptide arrays presented in C. The amino acid sequences of the cognate hybridized peptides are shown. **E**. Purified wild type (wt), H239L, E231R and E29R E33R SMUG1 proteins were analysed by PAGE and Coomassie staining. **F.** GST and GST-DKC1 proteins immobilized on the glutathione sepharose beads used in the GST pull-down were analysed by Western blotting. An antibody specific for DKC1 was used to confirm the identity of the double band. **G.** Co-immunoprecipitations were performed from HeLa cells overexpressing EYFP (lanes 1-3) or SMUG1-EYFP (lanes 4-6). Lysates were preliminary treated (+) or not (-) with 50 U DNase I prior to immunoprecipitation. Co-immunoprecipitated proteins were detected by Western blot analysis with EYFP-, GAR1- and NHP2-specific antibodies. IN, 10 % input; IP, immunoprecipitate; IgG L, light chain of Immunoglobulin G.

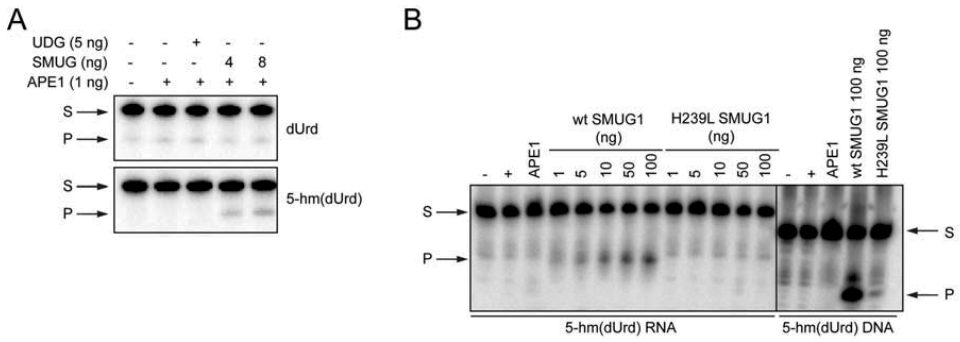


Figure S2. SMUG1 has activity on 5-hm(dUrd)-containing ssRNA, Related to Figure 2
A. Activity of recombinant wt SMUG1 was assayed on a 25–mer single stranded and 5'-^{32}P-end-labeled RNA substrate containing a centrally placed (dUrd) or 5-hm(dUrd) nucleoside, as indicated. The substrates were incubated with no enzyme, APE1 alone, UDG and APE1, or with increasing amounts of SMUG1 (4 and 8 ng) and APE1. The 12-mer radiolabeled product was resolved by denaturing PAGE and detected by phosphorimager. S, substrate; P, product. **B.** Activity of recombinant wt or H239L SMUG1 was assayed on a 25–mer single stranded and 5'-^{32}P-end-labeled oligoribonucleotide (RNA) or oligodeoxyribonucleotide (DNA) substrate containing a centrally placed 5-hm(dUrd), as indicated. The substrates were incubated with no enzyme (-), UDG and APE1 (+), APE1 alone, or with increasing amounts of SMUG1 (1, 5, 10, and 100 ng) and APE1. The 12-mer radiolabeled product was resolved by denaturing PAGE and detected by phosphorimager. S, substrate; P, product.

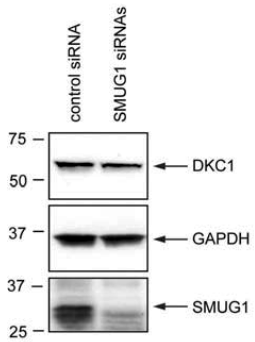


Figure S3. Knock-down efficiency of SMUG1 siRNAs, Related to Figure 3
HeLa cells were transfected with control or SMUG1 siRNAs for 48 h and whole cell extracts were subjected to Western blot analysis using DKC1-, GAPDH- and SMUG1-specific antibodies.

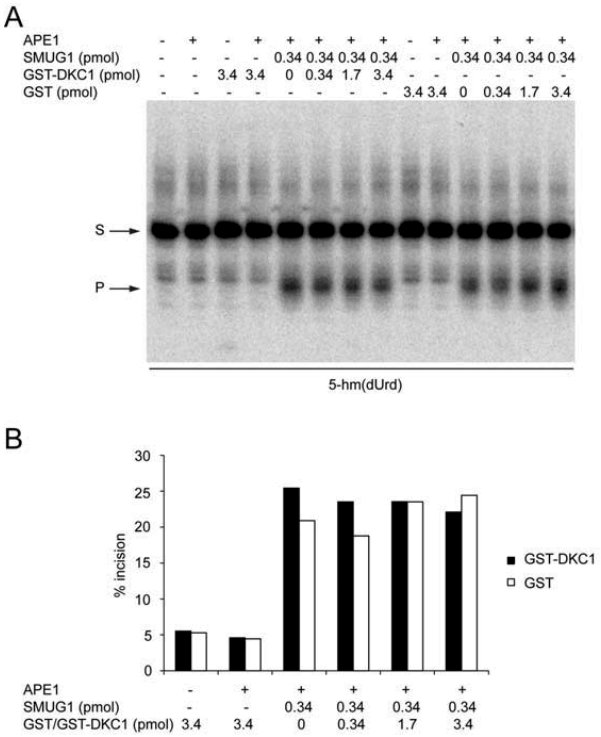


Figure S4. DKC1 does not affect *in vitro* SMUG1 incision activity on 5-hm(dUrd)-containing ssRNA, Related to the Results

A. Purified recombinant SMUG1 and GST-DKC1 proteins were mixed for 10 min at 37 °C and then incubated with the 5'-^{32}P-end-labeled 5-hm(dUrd) ssRNA substrate. After cleavage by APE1, RNA products were resolved by denaturing PAGE and detected by phosphorimager. **B**. SMUG1 incision activity (%) on 5-hm(dUrd)-containing ssRNA in the presence of GST-DKC1 or GST is presented.

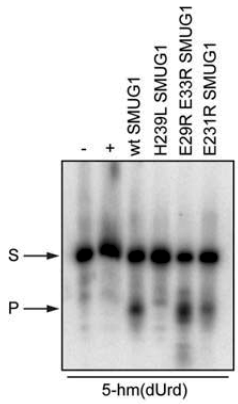


Figure S5. Activity of SMUG1 mutants on the 5-hm(dUrd)-containing RNA substrate, Related to the Results

Activity of recombinant wt, H239L, E29R E33R or E231R SMUG1 was assayed on a 25-mer single stranded and 5'-^{32}P-end-labeled oligoribonucleotide substrate containing a centrally placed 5-hm(dUrd) base. The substrates were incubated with no enzyme (-), UDG and APE1 (+) or with 100 ng SMUG1 and APE1. The 12-mer radiolabeled product was resolved by denaturing PAGE and detected by phosphorimager. S, substrate; P, product.

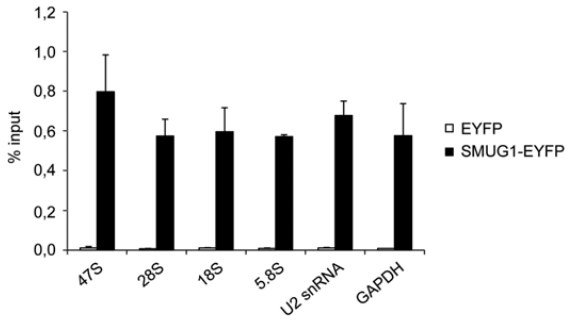


Figure S6. SMUG1 associates with several RNA species after formaldehyde crosslinking in an RNA-IP assay, Related to the Discussion

In vivo RNA-immunoprecipitation assays were performed in HeLa cells overexpressing SMUG1-EYFP or EYFP alone. RT-qPCRs were performed with primers specific for the indicated RNAs. The data shown are the mean ± SEM from three independent experiments.

SUPPLEMENTAL REFERENCES

Chen, R., Li, L., and Weng, Z. (2003). ZDOCK: an initial-stage protein-docking algorithm. Proteins *52*, 80-87.

Crain, P.F. (1990). Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. Methods in enzymology *193*, 782-790.

Hamma, T., Reichow, S.L., Varani, G., and Ferre-D'Amare, A.R. (2005). The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. Nature structural & molecular biology *12*, 1101-1107.

Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols *4*, 363-371.

Wibley, J.E., Waters, T.R., Haushalter, K., Verdine, G.L., and Pearl, L.H. (2003). Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. Molecular cell *11*, 1647-1659.