MEDIZINISCHE
UNIVERSITÄT WIEN

Effect of Obesity on Mutational Signature Generation in Gastrointestinal Cancers

Doctoral Thesis at the Medical University of Vienna for obtaining the academic degree

Doctor of Philosophy

Submitted by:

Mathilde Meyenberg, BSc

Supervisor:

Joanna Loizou, PhD

CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences,
Lazarettgasse 14, 1090 Vienna, Austria
Center for Cancer Research, Medical University of Vienna, Borschkegasse 8a, 1090 Vienna, Austria

Vienna, January 2023

Acknowledgements

Pursuing a PhD is quite a different learning experience than anything else I have experienced before. It has been a tumultuous journey with many ups and downs, due to many changes in institutes and lastly also supervisors. I want to thank my supervisor Dr. Joanna Loizou for accompanying me on the first part of the journey and for seeing this through to the end, despite all the changes. It means a lot to finish this journey with you on board. I wish you all the best for your future journey in industry.

One of the large breaks in my PhD journey was not changing institution twice, or even changing supervisors. In fact, switching environments and encountering new and awesome colleagues have helped to always gain new perspectives and to thoroughly shine light on aspects of my research I would not thought of by myself. Thank you to all colleagues and friends I found over the years in all the places I worked! The biggest upheaval in my research journey was the switch from wet lab science to computational science. This change required to not only learn completely new concepts and ways of approaching research, but also came with a huge learning curve to overcome the technical hurdle of learning how to program. I want to thank Dr. Israel Tojal da Silva for his patience in helping me gain the programmer's way of thinking during his exchange stay in Vienna and for his continued support of me from halfway across the world. I especially want to thank Anna Hakobyan for her part in this endeavor and for her friendship as a fellow PhD student navigating uncharted territory for both our groups. I also want to thank Dr. Jörg Menche for welcoming me to his group when our Lab dissolved and for giving me a fantastic space to explore exciting computational concepts in the future.

The switch to computational biology took me on a crazy adventure which involved taking evening classes to obtain a master's in bioinformatics. Disclosing this time-consuming idea to my wife was met with nothing but loving support. I want to thank my wonderful wife Selina with all my heart for her amazing support throughout all parts of the journey. From the high points to the pits of despair, she was there to catch me and to listen, make me laugh, and gave ingenious advice, because often an outsider's perspective is refreshingly insightful. This achievement took a lot of sacrifice from both of us - time not spent together and time not spent with our two children, who were as patient as children can be when work interfered with weekend playtime. I want to thoroughly acknowledge the work and dedication Selina put in to keep me sane and clear time and space for me to work, to study, and still be able to be a mom and a partner. Without you, I would never have found the correct balance in all this. Thank you for being the amazing woman you are! I love you with all my heart and I can't wait to embark on the next adventure with you. The next chapter of our lives is waiting to be written by us.

And finally, I want to thank my parents, my Mama and Baba, for their unwavering support in all aspects of life. I now understand how much effort goes into raising children and thus I know how lucky I was to get to be your daughter. Thank you for believing in me without putting pressure on my shoulders. Thank you for helping me reach for whatever I set my mind to. You raised me with the confidence to seek happiness, always accepting and loving. Thank you for being my shelter when I was little, my guidance when I was older, and my friends now.

Declaration

All experimental work showcased in this thesis was carried out in the laboratory of Dr. Joanna Loizou, performed at the Center for Molecular Research (CeMM) of the Austrian Academy of Sciences, in Vienna, Austria (2018 -2020), or at the Center for Cancer Research of the Medical University of Vienna, Austria (2020 -2022). All computational work was performed in the lab of Dr. Jörg Menche at the Department of Structural and Computational Biology in the Max Perutz Laboratories of the University of Vienna, Austria (2021-2023).

All experimental work, analysis of experimental data, and interpretation of the results were performed by the author, Mathilde Meyenberg, under the supervision of Dr. Joanna Loizou, Dr. Christoph Binder, and Dr. Bon-Kyoung Koo and with the assistance of Dr. Nikolina Papac-Milicevic, Laura Göderle, and Dr. Ji-Hyun Lee.

Collaborator Name	Affiliation
Dr. Nikolina Papac-Milicevic	Department of Laboratory Medicine, Medical University of Vienna, Austria
Laura Göderle	Department of Laboratory Medicine, Medical University of Vienna, Austria
Dr. Christoph Binder	Department of Laboratory Medicine, Medical University of Vienna, Austria
Dr. Ji-Hyun Lee	Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Austria Center for Genome Engineering, Institute for Basic Science, Daejeon, Republic of Korea
Dr. Bon-Kyoung Koo	Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Austria Center for Genome Engineering, Institute for Basic Science, Daejeon, Republic of Korea

The computational analysis of all experimentally acquired data was done in close collaboration with Anna Hakobyan from the Dr. Jörg Menche lab, and Dr. Israel Tojal da Silva. The author, Mathilde Meyenberg, was involved at every step of the computational analysis and performed the interpretation and visualization of the results.

Collaborator Name	Affiliation
Anna Hakobyan	CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria

	Department of Structural and Computational Biology in the Max Perutz Laboratories of the University of Vienna, Austria
Dr. Jörg Menche	CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria Department of Structural and Computational Biology in the Max Perutz Laboratories of the University of Vienna, Austria Faculty of Mathematics, University of Vienna, Austria
Dr. Israel Tojal da Silva	Laboratory of Computational Biology and Bioinformatics, A.C. Camargo Cancer Center, São Paulo, Brazil

The review article 1 presented in Chapter 1 was published in frontiers in Genetics and was reprinted under the copyright rules of Frontiers in Genetics (see Appendix for copyright statement).

Meyenberg M, Ferreira da Silva J & Loizou JI (2021) Tissue Specific DNA Repair Outcomes Shape the Landscape of Genome Editing. *Front Genet* 12: 728520 https://doi.org/10.3389/fgene.2021.728520

The review article 2 presented in Chapter 1 was published in Trends in Genetics (Cell Press) and was reprinted under the creative commons license of Trends in Genetics (see Appendix for copyright statement).

Ferreira da Silva J, Meyenberg M & Loizou JI (2021) Tissue specificity of DNA repair: the CRISPR compass. *Trends Genet* 37: 958–962 https://doi.org/10.1016/j.tig.2021.07.010

The article presented in chapter 2 is under review with Scientific Reports.
All analysis and interpretation were performed by Mathilde Meyenberg in collaboration with Anna Hakobyan; both contributed equally to the work.

The author of this thesis, Mathilde Meyenberg, prepared the figures and wrote the manuscript with input from Anna Hakobyan, Dr. Israel Tojal da Silva, Dr. Christoph Binder, Dr. Bon-Kyoung Koo, Dr. Nikolina Papac-Milicevic, Dr. Jörg Menche, and Dr. Joanna Loizou.

All chapters of this thesis were authored by Mathilde Meyenberg with input from Dr. Jörg Menche and Dr. Joanna Loizou.

Table of Contents

List of Figures

List of Tables

Abstract

Cancer evolves from normal tissue through acquiring adaptive mutations which rewire the functionality of inherent cellular pathways. The process of DNA damage and repair (DDR) is at the center of many cancer hallmarks and at once is a barrier and an aid for cancer development. Unraveling the complexity of cancer etiologies requires a systematic understanding of how DDR processes lead to the development of genomic instability under certain conditions. The mathematical and computational framework of mutational signatures has enabled the systematic discovery of biologically meaningful signatures attributable to specific mutational processes which reveal the evolution of malignancies. The analysis of mutational signatures in large cancer cohorts has sparked the discovery of mutational signatures of various mutational types. While some signatures have an explainable etiology and are even clinically actionable, the cause of many other signatures remains unknown. Too elucidate the etiology of unknown signatures, bottom-up *in-vivo* or *in-vitro* studies have proven useful. Since DDR processes exert their effect in a tissue specific manner, it has furthermore become increasingly important to study DNA repair and mutagenesis in a tissue specific manner, which has been made possible by the development of organoid based culturing systems. We combined the recent technical advantages in both, ex-vivo culturing systems and mutational signature analysis, to focus on studying mutational processes in the development of obesity associated cancer. Epidemiologically obesity is a well-recognized risk factor increasing the chance of cancer development, especially for gastrointestinal organs. Although it has been shown that diet induced obesity changes the homeostasis of intestinal stem cells – the cell of origin for intestinal cancers – it remains unclear how specifically obesity contributes to the development of gastrointestinal cancers. In this study, we used a mouse model of diet induced obesity on a wild type C57/BL6 background to observe the mutational landscape of intestinal stem cells after a 48-week exposure to a high-fat diet. By clonally expanding single stem cells in organoid culture, and obtaining whole genome sequences, we found that single base substitution and insertion/deletion signatures present in the mice on the high-fat diet were similar to those on a standard diet and reflected normal processes of aging, cellular replication, and oxidative stress. Our study concludes that high fat diet alone, in the absence of other stressors such as chemical exposure or driver gene mutations, is not enough to induce an increase in genomic instability.

Zusammenfassung

Krebs entsteht aus normalem Gewebe, durch die Ansammlung von Mutationen, welche die Funktionalität der zellulären Signalwege verändern. Die Prozesse der DNA-Schädigung und -Reparatur (DDR) stehen im Mittelpunkt vieler Krebsmerkmale und stellen gleichzeitig ein Hindernis und eine Hilfe für die Krebsentwicklung dar. Um die Komplexität der Krebsentstehung zu begreifen, ist ein systematisches dafür Verständnis erforderlich, wie DDR-Prozesse unter bestimmten Bedingungen zur Entwicklung von genomischer Instabilität beitragen. Die mathematischen und computergestützten Konzepte der Mutationssignaturen hat die systematische Entdeckung biologisch aussagekräftiger Signaturen ermöglicht, welche spezifischen Mutationsprozessen zuzuschreiben sind und so den Entwicklungsprozess bösartiger Erkrankungen beschreiben. Die Analyse von Mutationssignaturen in DNA-Sequenz Daten von Krebspatienten hat die Entdeckung von Mutationssignaturen für verschiedene Mutationstypen vorangetrieben. Während einige Signaturen eine erklärbare Ätiologie haben und sogar im klinischen Kontext einsetzbar sind, bleibt die Ursache vieler anderer Signaturen unbekannt. Um die Ätiologie unbekannter Signaturen aufzuklären, haben sich Bottom-up *In-vivo* oder *In-vitro* Studien als nützlich erwiesen. Da DNA Reparatur Prozesse ihre Wirkung gewebespezifisch entfalten, ist es immer wichtiger geworden, DNA-Reparatur und Mutagenese auch gewebespezifisch zu untersuchen, was durch die Entwicklung von organoid-basierten Kultursystemen möglich geworden ist. Wir haben die neusten technischen Vorteile von Ex-vivo-Kultursystemen mit der Analyse von Mutationssignaturen kombiniert, um uns auf die Untersuchung von Mutationsprozessen bei der Entstehung von Adipositát bedingtem Krebs zu konzentrieren. Epidemiologisch gesehen ist Fettleibigkeit ein anerkannter Risikofaktor, der die Wahrscheinlichkeit der Entstehung von Krebs, insbesondere im Darmtrakt, erhöht. insbesondere für den Darmtrakt. Obwohl bereits erforscht wurde, dass Fettleibigkeit die Homöostase und Signalwege von Darmstammzellen - den Ursprungszellen für Darmkrebs - verändert, bleibt unklar, wie genau Adipositát zur Entwicklung von Magen-Darm-Krebs beiträgt. In dieser Studie haben wir ein Mausmodell für ernährungsbedingte Fettleibigkeit in einem genetischen Wildtyp-Hintergrund verwendet (C57/BL6). So konnten wir genomweite Mutationen in Darmstammzellen nach 48-wöchiger Exposition gegenüber einer fettreichen Ernährung zu beobachten. Durch die klonale Expansion einzelner Stammzellen in organoider Kultur und der Analyse der Genomsequenzen ermittelten wir, dass die Mutationssignaturen bei Mäusen mit fettreicher Ernährung denen einer Standarddiät ähneln und normale Prozesse der Alterung, der Zellreplikation und des oxidativen Stresses widerspiegeln. Unsere Studie kommt zu dem Schluss, dass eine fettreiche Ernährung allein, ohne andere Stressfaktoren wie chemische Exposition oder Treibermutationen, nicht ausreicht, um einen Anstieg der genomischen Instabilität zu bewirken.

Publications arising from this thesis

Meyenberg M, Ferreira da Silva J & Loizou JI (2021) Tissue Specific DNA Repair Outcomes Shape the Landscape of Genome Editing. *Front Genet* 12: 728520 https://doi.org/10.3389/fgene.2021.728520

Ferreira da Silva J, Meyenberg M & Loizou JI (2021) Tissue specificity of DNA repair: the CRISPR compass. *Trends Genet* 37: 958–962 https://doi.org/10.1016/j.tig.2021.07.010

Meyenberg M, Hakobyan A et al (2023), Mutational Landscape of Intestinal Stem Cells After Long-term In Vivo Exposure to High Fat Diet, *Under Review in Scientific Reports*

List of Abbreviations

Abbreviation	**Explanation**
GLOBOCAN	Global cancer observatory of the international agency for research on cancer
BMI	Body mass index
IARC	International agency for research on cancer
CRC	Colorectal cancer
APC	Adenomatous polyposis coli
Wnt	Wingless-Type MMTV Integration Site Family
LEF1	Lymphoid enhancer binding factor 1
TCF	T-cell factor
c-myc	Proto-Oncogene C-Myc (transcription factor)
CIN	Chromosomal instability
MSI	Microsatellite instability
CIMP	CpG island methylation pathway
KRAS	KRAS Proto-Oncogene
BRAF	B-Raf Proto-Oncogene
TP53	Tumor Protein P53
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
SMAD4	SMAD family member 4 (MADH4 - Mothers Against Decapentaplegic Homolog 4)
IGF	Insulin growth factor
IL-6	Interleukin 6
JAK	Janus kinase
STAT	Signal transducer and activator of transcription
MAPK	Mitogen activated protein kinase
PI3K pathway	Phosphoinositide 3-kinase pathway
ISC	Intestinal stem cells
LGR5	Leucine rich repeat containing G-protein coupled receptor 5
PPAR-∂	peroxisome proliferator-activated receptor delta
RAS	Ras-Proto Oncogene
RAF	Raf-Proto-Oncogene
MEK	Mitogen activated protein kinase
RB	retinoblastoma protein
TGFβ	transforming growth factor β
IGF-1R	Insulin Like Growth Factor 1 Receptor

IL-3	Interleukin 3
BAX	BCL2 Associated X Apoptosis Regulator
p53	Protein produced from TP53
bp	basepairs
ALT	alternative lengthening of telomeres
VEGF	vascular endothelial growth factor
HLA-I	Major Histocompatibility Complex Class IA
CD8+ T cells	T cells positive for Surface Glycoprotein CD8
TCR	T-cell receptor
PD-1	programmed death-1 (receptor)
PD-L1	programmed death-1 ligand 1
PD-L2	programmed death-1 ligand 2
EGF	Epidermal Growth Factor
FGF2	Fibroblast Growth Factor 2
DNA	Deoxyribonucleic acid
DDR	DNA damage and repair
UV	Ultraviolet (radiation)
ATM	ataxia telangiectasia mutated
ATR	Ataxia Telangiectasia And Rad3-Related Protein
MRN	MRE11-RAD50-NBS
CHK2	Checkpoint kinase 2 (cell cycle)
CHK1	Checkpoint kinase 1 (cell cycle)
911 complex	RAD9, RAD1, HUS1 (Checkpoint Clamp Components)
53BP1	Tumor Protein P53 Binding Protein 1
BRCA1	Breast Cancer Type 1 Susceptibility Protein
NHEJ	Non-homologous end joining
HR	Homologous recombination
MGMT	DNA methyltransferase (O-6-Methylguanine-DNA Methyltransferase)
POLB	DNA Polymerase Beta
POLQ	DNA Polymerase Theta
TLS	Translesion synthesis
REV1	REV1 DNA Directed Polymerase
REV3	REV3 Like DNA Directed Polymerase Zeta Catalytic Subunit
POLH	DNA Polymerase Eta
POLK	DNA Polymerase Kappa
BER	Base excision repair
ROS	Reactive oxygen species
NER	nucleotide excision repair
GG-NER	Global genome nucleotide excision repair
TC-NER	Transcription coupled nucleotide excision repair
RNA	Ribonucleic acid
CPDs	cyclobutene-pyrimidine dimers
6-4PPs	pyrimidine-(6-4)-pyrimidone photoproducts
BaP	benzo[a]pyrene
XPC	Xeroderma Pigmentosum Complementation Group C
RAD23B	UV Excision Repair Protein RAD23 Homolog B
DDB2	Damage Specific DNA Binding Protein 2
CSA	Cockayne Syndrome WD Repeat Protein CSA (ERCC Excision Repair 8)
CSB	Cockayne Syndrome Protein CSB (ERCC Excision Repair 6)
USP7	Ubiquitin Specific Peptidase 7
UVSSA	UV Stimulated Scaffold Protein A
XAB2	XPA Binding Protein 2
RNA-Pol2	RNA Polymerase II
XPA	Xeroderma Pigmentosum Complementation Group A
TFIIH complex	basal transcription factor 2 complex
XPB	Xeroderma Pigmentosum Complementation Group B

XPD	Xeroderma Pigmentosum Complementation Group D
XPG	Xeroderma Pigmentosum Complementation Group G
XPF	Xeroderma Pigmentosum Complementation Group F
ERCC1	ERCC Excision Repair 1
POLD	DNA Polymerase Delta
POLE	DNA Poymerase Epsilon
LIG1	DNA Ligase 1
LIG3	DNA Ligase 3
MMR	Mismatch repair
MSH2	MutS Homolog 2
MSH6	MutS Homolog 6
MSH3	MutS Homolog 3
ATP	Adenosine Triphosphate
MLH1	MutL Homolog 1
MLH3	MutL Homolog 3
PMS2	PMS2 Postmeiotic Segregation Increased 2
MutH	DNA glycosylase for removing mismatched adenine
PCNA	proliferating cell nuclear antigen
RPC	Replication factor c
EXO1	Exonuclease 1
RPA	replication protein A
ICL	Interstrand crosslinks
FA	Fanconi Anemia
FANC	Fanconi Anemia Complementation Group Genes
FAAP	Fanconi Anemia Core Complex Associated Proteins
MHF1	FANCM-Interacting Histone Fold Protein 1
MHF2	FANCM-Interacting Histone Fold Protein 2
UBE2T	Ubiquitin Conjugating Enzyme E2 T (FANCT)
FANCD2	Fanconi Anemia Complementation Group D2
FAN1	Fanconi-associated nuclease
ERCC4	ERCC Excision Repair 4 (XPF)
SSB	Single strand break
SSBR	Single strand break repair
TOP1	topoisomerase 1
DSB	double strand break
PARP1	Poly(ADP-Ribose) Polymerase 1
XRCC1	X-Ray Repair Cross Complementing 1
APE1	Apurinic/Apyrimidinic Endodeoxyribonuclease 1
TDP1	Tyrosyl-DNA Phosphodiesterase 1
APTX	Aprataxin
AMP	adenosine monophosphate
FEN1	Flap Structure-Specific Endonuclease 1
DSBR	Double strand break repair
V(D)J	Variable (V) diverse (D) joining (J) segments of immunoglobulin and T-cell receptor genes
SSA	single strand annealing
TMEJ	polymerase theta-mediated end joining
DNA2	DNA Replication Helicase/Nuclease 2
BLM	Bloom Syndrome RecQ Like Helicas
WRN	Werner Syndrome ATP-Dependent Helicase
CtIP	Retinoblastoma-Binding Protein 8
Ku70	X-Ray Repair Cross-Complementing Protein 6 (XRCC6)
Ku80	X-Ray Repair Cross-Complementing Protein 5 (XRCC5)
PNKP	Polynucleotide Kinase 3'-Phosphatase
ARTEMIS	DNA Cross-Link Repair 1C
LIG4	DNA Ligase 4
XRCC4	X-Ray Repair Cross Complementing Protein 4
XLF	Non-Homologous End-Joining Factor 1 (XRCC4-like factor)

BRCA2	Breast Cancer Type 2 Susceptibility Protein
RAD51	RAD51 Recombinase
RAD52	DNA Repair Protein RAD52 Homolog
MMEJ	microhomology-mediated end joining
CRISPR	clustered regularly interspaced short palindromic repeats
HNPCC	hereditary non-polyposis colon cancer
WGS	whole genome sequencing
ddNTP	dideoxunucleotide
SBS	Single base substitution signature
SNV	Single nucleotide variants
DBS	Double base substitution signature
indel	Insertion and deletion
ID	Indel signature
CN	Copy number
LOH	loss of heterozygosity
Het	heterozygosity status
TD	tandem duplication
RS	Rearrangement signature
NMF	Non-negative matrix factorization
BIC	Bayesian Information Criterion
KDM4A/B	Lysine Demethylase 4 A/B
TIP60	Lysine Acetyltransferase 5
NTHL1	Nth Like DNA Glycosylase 1
MUTYH	MutY DNA Glycosylase
TOP2A	topoisomerase2
APOBEC	Apolipoprotein B MRNA Editing Enzyme (Family)
AID	Activation Induced Cytidine Deaminase
POLD1	DNA Polymerase Delta Catalytic Subunit 1
CX	Chromosomal instability signature
SD	Standard diet
HFD	High fat diet
APCmin	Mouse model with a mutation in the APC gene (min = multiple intestinal neoplasia)
FAP	familial adenomatous polyposis
PLA2G2A	Phospholipase A2 Group IIA
dMMR	Mismatch repair deficiency

CHAPTER 1: INTRODUCTION

Cancer – The enemy from Within

A Reflection of the Normal Self

History

Our earliest ancestors of the hominid species have been befallen by tumors 1.7- 1.9 million years ago (Randolph-Quinney *et al*, 2016; Odes *et al*, 2016). What they believed to be the cause of their ailments will remain unknown but throughout the centuries many physicians and philosophers have speculated about the cause of the mysterious disease. While the ancient Greeks believed cancer to be a disease of natural origin, later writings from the Roman empire and medieval ages all heavily lean on Galen's humoral theory, which postulated that thick black bile is the cause of malignant cancers. At the end of the dark ages, French physicians Henri de Mondeville, Lanfranc, and Guy the Chauliac started a new era of cancer research by rejecting Galen's millennial old theories and beginning to map cancer and its anatomy systematically (Hajdu, 2011a).

The advent of the Renaissance brought many scientific inventions, such as the microscope, which allowed to study cancer not just superficially, building the foundation of modern pathology and oncology. The increased acceptance of postmortem examinations and autopsies allowed for the systematic discovery and description of cancers according to their primary anatomical site. While treatments ranged from compression to surgical excision to treatment with arsenic pastes, the underlying cause of the evolution of cancer remained in the dark. Around 1650, two prominent physicians, Nicholas Tulp and Zacutus Lusitani, independently concluded cancer to be contagious based on the observation that members of the same household often developed breast cancer (Hajdu, 2011b). While the heritable genetic cause of this observation about breast cancer would not be uncovered until centuries later, other observations about the origins of lung cancer began to form an idea of external causes of cancer. In his book "De Grandibus" from 1567, the swiss physician and chemist Paracelsus described incidences of lung cancer in miners and smelters of metal ores (Hajdu, 2011b). Two hundred years later (1761), English physician John Hill authored a warning about the use of snuff tobacco as the purported cause of nasal polyps, a pre-cancerous condition (Hajdu, 2011b).

Amidst all the case descriptions and observations, the idea of the origin of cancer began to shift away from mystic or contagious causes. Instead, cancer began to be understood as a disease involving the continuous and unhindered growth of normal bodily tissue. Cancer is therefore a disease like none other because it has multiple causes and originates from within. It is, as author and physician Siddhartha Mukherjee stated in his book "The Emperor of all Maladies – A Biography of Cancer": "*The very cellular processes and genes which allow normal cells to grow, divide, adapt, and heal, become distorted in cancer, allowing cancer to be an anamorphic reflection of our normal self*" (Mukherjee, 2010).

The Cancer Burden Today – a Global Perspective

The International Agency for Research on Cancer supports a global cancer surveillance program called the Global Cancer Observatory, which provides publicly available data on cancer statistics, allowing accurate estimates of the global cancer burden. According to the 2020 GLOBOCAN cancer statistics, nearly 20 million people worldwide were diagnosed with cancer, while almost 10 million people died prematurely due to cancer (Sung *et al*, 2021). The global cancer incidence is only expected to increase from 17 million in 2018 to nearly 34 million cases by 2070 (Soerjomataram & Bray, 2021). The expected increase in incidence is largely driven by demographic changes due to a shift in socio-economic factors, especially in low to middle income countries (Sung *et al*, 2021) (Figure 1).

As the average life expectancy increases with rising socio-economic status, the cancer incidence also increases (Soerjomataram & Bray, 2021). In a sense, cancer can be viewed as an aging related disease, and will thus influence the healthcare burden of the future. To successfully meet this current and future challenge, efforts in prevention, detection and diagnosis, and treatment must simultaneously advance. Consequently, it becomes paramount to understand the development and evolution of cancer intimately, on the cellular and molecular level. The state-of-the-art science today allows for personalized and large-scale study of tumor biology and puts us in a unique position to understand the ever more complex landscape of tumor evolution, from origin through treatment.

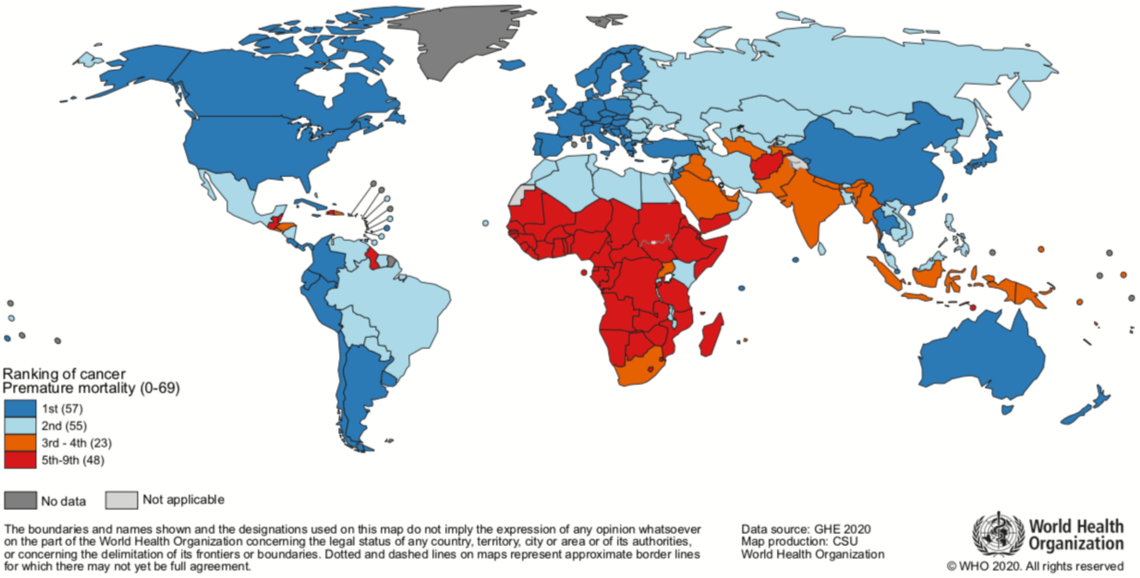

Ranking of cancer
Premature mortality (0-69)
1st (57)
2nd (55)
3rd - 4th (23)
5th-9th (48)
No data
Not applicable
The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.
Data source: GHE 2020
Map production: CSU
World Health Organization
World Health Organization
© WHO 2020. All rights reserved

Figure 1. Global ranking of cancer as a cause of premature death
The ranking of cancer as a cause of premature death (age < 70 years by country. Cancer ranks 1st and 2nd in high income and middle-income countries respectively, and 3rd - < 5th in middle- to low-, and low-income countries respectively. Figure taken from (Sung *et al*, 2021), Data Source (WHO, 2020), for licensing information see Appendix

Lifestyle factors affecting cancer risk

Apart from normal aging increasing the cancer burden, there are other, preventable, factors which increase cancer risk. A recent systematic analysis of cancers attributable to specific risk factors considered behavioral (*e.g.* smoking), environmental and occupational (*e.g.* exposure to hazardous chemicals or radiation), and metabolic risk factors (high body-mass index (BMI) and dietary factors such as high fat diet and diets low in fiber). Globally, all risk factors combined accounted for 4.45 million deaths (95% confidence interval), representing 44.4% of all cancer deaths worldwide (Tran *et al*, 2022). Thus, epidemiological evidence suggests that just under half of all cancer cases may be preventable.

In detail, the leading risk factors are smoking, followed by alcohol use (behavioral), and high BMI (metabolic). While behavioral and environmental risk factors are somewhat equally distributed worldwide, irrespective of income status of countries, cancers attributable to metabolic risk factors are markedly more common in high income countries (Tran *et al*, 2022). However, the cancer incidence rate of cases attributable to metabolic risks saw the greatest percentage increase between 2010 and 2019, also in low and low-middle income countries (Tran *et al*, 2022). Thus, metabolic risk factors such as high BMI are a growing concern globally.

Obesity and Metabolic Risks

Over the past 40 years, the global prevalence of obesity has been rising substantially (Jaacks *et al*, 2019). Obesity is associated with numerous comorbidities, such as hypertension, non-alcoholic fatty liver disease, and type II diabetes (Jaacks *et al*, 2019; Blüher, 2019). Importantly, body fat accumulation has been recognized by the International Agency of Research on Cancer (IARC) as an important risk factor in cancer development (Hopkins *et al*, 2016; Calle *et al*, 2009; Friedenreich *et al*, 2021; Avgerinos *et al*, 2019). Especially cancers of organs along the gastrointestinal tract are impacted by obesity, including the esophagus, stomach, colon and rectum, as well as the liver (Lauby-Secretan *et al*, 2016). Of those, the chance of developing colorectal cancer is particularly increased in patients with high BMI (Tran *et al*, 2022). Given the obvious link between a high BMI and a higher risk of CRC, learning more about the underlying disease etiology should help to inform preventive programs. However, even though the epidemiological association seems clear, the underlying molecular mechanisms still do not paint the full picture of obesity associated cancer development. However, there are multiple lines of evidence connecting the increase in body fat to cellular pathways, also commonly found altered in cancer.

Mechanisms of Obesity Driven Cancer Risk in Colorectal Cancer

The development of colorectal cancer is governed by a well-defined progression of mutations known as the adenoma-carcinoma sequence (Fearon & Vogelstein, 1990). One of the first mutations inactivates adenomatous polyposis coli (APC), which leads to constitutive Wnt-/ β-catenin signaling. Without APC controlling the levels of β-catenin, the phosphorylated transcriptional activator β-catenin translocates to the nucleus and activates downstream transcriptional targets such as lymphoid enhancer-binding factor 1 (LEF1) and T cell factor (TCF), which in turn activate the transcription of downstream genes involved in proliferation, differentiation, and migration. Notably, the transcription factor c-myc, a common proto-oncogene involved in cell proliferation is also activated by canonical Wnt-signaling (Matsui, 2016; Duchartre *et al*, 2016; Rim *et al*, 2022).

Beside increased proliferation, the development of genomic instability plays a key role in tumor development. Colorectal cancer arises via three distinct genetic pathways: the chromosomal instability pathway (CIN), the microsatellite instability pathway (MSI), and the CpG island methylation pathway (CIMP) (Bogaert & Prenen, 2014). Although the development of a tumor is heterogeneous and occasionally involves overlapping pathways, all three pathways are defined by an increase in genomic instability, which allows for the acquisition of additional mutations in a set of tumor suppressor and oncogenes, including KRAS and BRAF, which are often found to be mutually exclusive (Midthun *et al*, 2019). Additionally, TP53, PIK3CA, and SMAD4 are common genes mutated in CRC. Drost *et al* tested the combinatorial and additive effect of multiple gene mutations of the adenoma-carcinoma sequence in intestinal stem cells. Their study discovered that the simultaneous loss of APC and P53 is already sufficient to cause significant levels of chromosomal instability, which is typical for the CIN pathway (Drost *et al*, 2015). Therefore, the loss of genomic stability might present an early event in tumor development, enabling further alterations. How obesity affects the development of genomic instability is not fully explored, although multiple hypotheses exist on how the obesity condition affects related cellular pathways, which will be discussed briefly.

The expansion of adipose tissue leads to an increase in the secretion of multiple signaling molecules including insulin, insulin growth factor (IGF), hormones such as leptin, and cytokines such as interleukin 6 (IL-6), as well as the decrease in adiponectin. These soluble factors are transported through the blood and activate downstream signaling cascades through docking and activating their specific cell surface receptor (Hopkins *et al*, 2016) (Figure 2). Specifically, the JAK/STAT, MAPK, and PI3K pathways are activated, which collectively stimulate processes that drive cellular survival, growth and proliferation, and angiogenesis. Because the molecular targets of these pathways are overlapping, there may be some amplification of the resulting biological effects. Insulin signaling, for instance, acts through both the MAPK and PI3K signaling pathways, effecting cell growth via PI3K and proliferation via MAPK. This creates a metabolic environment which allows for increased glucose uptake and processing via glycolysis, thus providing increased amounts of molecular building blocks that can support enhanced proliferation (Hopkins *et al*, 2016). Taken together, these processes are thought to create an environment that lowers the barrier to oncogenic transformation.

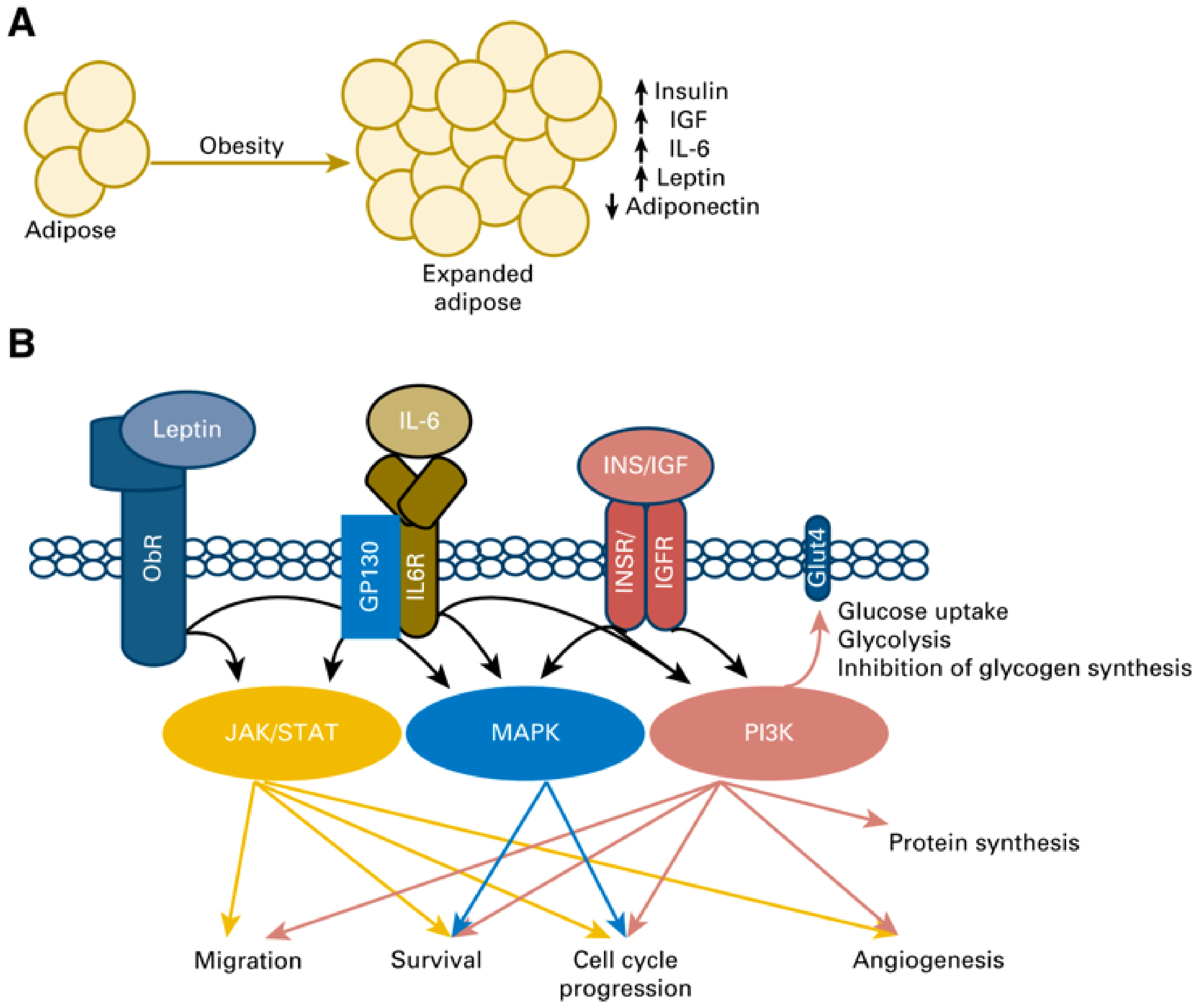


Figure 2. Signaling of Adipose Tissue and Downstream effects, figure taken from (Hopkins *et al*, 2016) *Reprinted with permission of Wolters Kluwer Health, Inc* (see Appendix)

Since the relative risk of developing cancer in response to obesity varies by organ, there is likely a tissue specific effect of these signaling pathways at play (Speakman & Goran, 2010; Pereira *et al*, 2021). Thus, for each tissue, the most relevant cell population should be considered when studying the impact of obesity. In the case of colorectal cancer, the cells of origin are intestinal stem cells (ISC), located at the bottom of the intestinal crypts (Sato *et al*, 2009). These actively cycling cells can be identified by their expression of the LGR5 protein (leucine rich repeat containing G protein coupled receptor 5), harvested and studied *ex-vivo* with organoid culturing techniques (Sato *et al*, 2011). The advancements of 3D culturing techniques allow the study of unaltered cells within their normal tissue architecture and signaling environment.

The metabolically highly active ICSs have been demonstrated to be linked to increased risk of cancer initiation though metabolic or dietary perturbation. Wang *et al* observed that increased availability of cholesterol, for instance through dietary supplementation, increases stem cell proliferation and the rate of tumor formation in an APC deficient background (Wang *et al*, 2018). Another mechanism of action is the activation of PPAR-∂ (peroxisome proliferator-activated receptor delta) signaling, which, via canonical Wnt-signaling, confers features of stemness on non-stem cell progenitors (Beyaz *et al*, 2016). This increases the pool of actively proliferations cells and thus raises the risk of one of these cells acquiring driving mutations and escaping proliferative control.

Since the DNA damage response always must balance proliferative control with genome stability, it is an interesting and unexplored question to study how dietary components or diet induced obesity impact genomic stability in intestinal stem cells. Before we can consider the complexity of how obesity affects the development of cancer, we need an in-depth discussion of the hallmarks of cancer and the role of genomic stability in the development of cancer, as well as the specific pathways involved in the DNA damage response. Additionally, we must consider the tissue specificity of DDR signaling pathways.

Hallmarks of cancer

Cancer is more accurately a collection of many different malignancies, originating from different cells and tissues throughout the body. Still, tumors of different origins all share common features which marks their aberrant functionalities compared to normal cells. Hanahan and Weinberg have proposed a logical framework, describing common biological principles in malignant transformation. The suggested hallmarks of cancer describe both, the individual cellular processes (Hanahan & Weinberg, 2000) and signaling networks between cells, known as the tumor microenvironment (Hanahan & Weinberg, 2011). In the following, I will discuss the hallmarks in detail, contemplating how each contributes to the development of malignancies.

Sustaining Proliferative Signaling

Each cell exists within a tightly regulated environment within the tissue and specifically the extracellular matrix. Within this context, cells integrate external signals via receptors, as well as internal signals to correctly balance proliferative signaling. Cancer cells have evolved numerous ways to become independent at each level of regulation. Instead of relying on outside signals, cancers can overexpress mitogenic signals in an autocrine manner, or upregulate the expression of cell surface growth factor receptors to amplify existing signals. The best-known example of pro-mitogenic signaling is the RAS-RAF-MEK signal path, which in about 25% of all human cancers is mutated in some form, resulting in pro-internal mitogenic signal activation, independent of extracellular ligands (Hanahan & Weinberg, 2000, 2011). Through these alterations, cancer cells escape the normal cellular fate of eventual quiescence. Once a tumor grows, the complex tumor microenvironment serves to further amplify this effect, as neighboring tumor and other cell types send and amplify proliferative signals (Bianchi *et al*, 2020).

Evading Growth Suppressors

The cellular balance between quiescence and proliferation does not solely rely on proliferative signals, but also the counterbalance of growth-suppressive signals. Mechanistically, growth suppression takes effect within the cell cycle at the G1 stage, the crossroad to G0, leading to quiescence or a differentiated post-mitotic state, which permanently suppresses continued proliferation. One prominent mechanism of control is the phosphorylation of retinoblastoma protein (Rb), which in phosphorylated form releases transcription factor E2F, responsible for the expression of genes favoring advancement of the cell cycle from G1 to S phase. The anti-growth molecule transforming growth factor β (TGFβ) acts through a variety of mechanisms to prevent the Rb phosphorylation, thus preventing cell cycle progression (Hanahan & Weinberg, 2000, 2011). Cancer cells adapt to evade these signals by lowering their TGFβ sensitivity, either through downregulation of TGFβ receptor expression or mutation. In this way, cancer cells not only signal continuous proliferation but also escape growth stop signals.

Resisting Cell Death

Any normal organism goes through a cycle of regulated overturn of developing and dying cells. Apoptosis describes the controlled mechanism of programmed cell death, which commences upon death signals the cell senses, and starts a mostly irreversible program of effectors within the cell. At the beginning of the signaling cascade, cell surface receptors receive and integrate survival signals (e.g. via the IGF-1R receptor or the IL-3 receptor) as well as death signals (e.g. TNFα via the TNF-R1 receptor) (Hanahan & Weinberg, 2000, 2011). Most signals result in the eventual release of cytochrome c from the mitochondria. Beside external signals, endogenous signals like the p53 signaling cascade can induce apoptosis after sensing DNA damage, through upregulation of the pro-apoptotic gene BAX, which in turn triggers cytochrome c release (Junttila & Evan, 2009).

Cancer cells evade apoptosis by both modulating pro-survival and pro-apoptotic signaling cascades. Common mechanisms include an upregulation of ligands and receptors involved in anti-apoptotic signaling or expressing truncated death-signal receptors to block pro-apoptotic signaling. With characteristic sustained proliferation, cancerous cells, through oncogene-induced replicative stress, encounter more DNA damage (Hills & Diffley, 2014), which further necessitates an escape from the normal apoptotic control. It is not surprising that p53, as the central signal integrating molecule, is the most frequently mutated gene in all human cancers (Hainaut & Pfeifer, 2016).

Enabling Replicative Immortality

Upon the fulfillment of the first three hallmarks, cancer cells have yet another obstacle to overcome. To truly replicate without restraint, cells must overcome the limitation of senescence, induced by excessive telomere shortening during the many cell divisions a malignant cell undergoes (Herbig *et al*, 2004). The ends of chromosomes contain long stretches of 6bp repeats, of which 10-100 bp are lost during each cell cycle because DNA polymerase is unable to completely replicate DNA to its final 3’ end (Hanahan & Weinberg, 2000). The presence of non-coding telomeric repeats thus protects from loss of vital DNA elements. At the same time, telomeric shortening constitutes a kind of cellular clock which limits the ultimate number of divisions possible. This is because exposed DNA at the end of chromosomes binds to other unprotected ends and engages in chromosomal fusion, leading to translocations, karyotypic aberrations, chromosomal breakage, and ultimately mitotic catastrophe and cell death (Counter *et al*, 1992).

To circumvent this problem, cancer cells have evolved to exploit naturally existing pathways of telomere maintenance. Many cancers upregulate the expression of the telomerase enzyme, which extends telomeres by adding hexanucleotide repeats. A different mechanism for achieving the same outcome is the ALT pathway (alternative lengthening of telomeres), which relies on recombination. Double strand breaks in telomeric regions lead to break induced replication. Owing to the high level of homology in telomeric sequences, the invading homologous strand presents a perfect primer for the synthesis and extension of telomeres (Zhao *et al*, 2019). Either mechanism provides cancer cells with the solution needed to overcome replicative senescence and enables continued replication and evolution of the malignant cell.

Inducing Angiogenesis

All cells require access to nutrients and oxygen, requiring even cancer cells to reside close to blood vessels. During the development of a tumor, this poses yet another challenge to cancer. In order to grow to macroscopic sizes, a tumor must produce not just cells but with them a network of blood vessels. The process of angiogenesis is tightly regulated by a balance of initiating and inhibiting factors such as VEGF (vascular endothelial growth factor), a promoter, or thrombospoindin-1, an inhibitor. Before gaining the ability to form a larger tumor, cancer cells must develop the ability to disturb the balance of regulating factors of angiogenesis. Often this is achieved by altering gene expression, upregulating and downregulating promoting and inhibiting factors respectively (Hanahan & Folkman, 1996).

Activating Invasion and Metastasis

The final stage in a cancer's evolution often is the migration and invasion of tissues at sites distant to the primary tumor. This process, termed metastasis is composed of several steps, including the local invasion of malignant cells into blood or lymphatic vessels, the travel to distant sites, and expansion within the distance sites (Suhail *et al*, 2019). One of the rate-limiting steps to this cellular mobility is the cell-to-cell adhesion within the tumor microenvironment. Normally, adhesion molecules, expressed on the surface of epithelial cells mediate the anchoring of cells. One such molecule, E-cadherin, is frequently lost in cancer through a variety of mechanisms, including loss of function mutations, proteolytic degradation, or transcriptional downregulation (Christofori & Semb, 1999). Through eliminating factors involved in cell-to-cell adhesion, cancer cells enable the chain of events involved in metastasis, also known as the epithelial to mesenchymal transition.

Deregulating Cellular Energetics

All hallmarks described above are critical to a cancer's ability to outgrow the normal constraints of cellular and tissue homeostasis. However, one challenge remains unaddressed: How cancerous cells can fulfil their immense need for energy and biomolecular building blocks while sustaining a high level of proliferation. This obstacle of cellular energetics is circumvented by cancer cells through rewiring the cellular metabolism, away from oxidative phosphorylation and toward glycolysis. The curious observation that cancer cells prefer glycolysis, even in the presence of oxygen, was first made by Otto Warburg in the 1930's (Hanahan & Weinberg, 2011). Despite the relative inefficiency of glycolysis over the citric acid cycle and the mitochondrial electron transport chain, there is a benefit to this metabolic switch, termed the 'Warburg effect'. Glycolysis is a central hub in the metabolic network, producing metabolic intermediates as basic building blocks for many other biosynthetic pathways. In this way, enhanced glycolysis provides the building blocks for increased production of nucleic- and amino-acids, required for the gain in proliferation. The deficiency in energy supply caused by this switch is partially mitigated by increased glucose import, often achieved by upregulated glucose transporters (DeBerardinis & Chandel, 2020).

Avoiding Immune Destruction and Tumor-promoting Inflammation

As cancer evolved, its dealings with the host immune system are a double-edged sword. On the one hand, malignant cells must avoid recognition and destruction by the immune system. However, on the flip side, advancing tumors also can utilize infiltrating immune cells and the resulting inflammation to drive tumor evolution forward (Hanahan & Weinberg, 2011). Given the antagonistic but complementary nature of these two hallmarks, they will be discussed together here.

That immune cells of both branches of the immune system, innate and adaptive, contribute the fight against cancer is made obvious by the observation that patients with primary immunodeficiencies have up to 1.6-fold higher relative risk to develop cancer (Kebudi *et al*, 2019). To avoid immune destruction, malignant cells evolve to trick the immune system in a variety of ways.

A great example of passive immune evasion is the modulation of HLA-I dependent antigen presentation. The human leukocyte antigen (HLA) is responsible for presenting antigen peptides on cell surfaces to CD8+ T cells, leading to their activation and subsequent killing of the antigen presenting cell. Tumor cells, especially those with a high mutational burden, often produce many tumor neo-antigens which could be recognized by CD8+ T cells (Jhunjhunwala *et al*, 2021). The successful recognition of tumor neoantigens depends on two steps. First, immune cells, often dendritic cells, must take up and present tumor neoantigens to prime naïve CD8+ T cells. Second, CD8+ T cells must find the tumor neoantigen at the surface of the cancer cell to recognize and destroy the target. Cancer cells have developed diverse strategies to interfere at both steps, via repressing dendritic cell function (Wculek *et al*, 2020) or downregulation of HLA-I to avoid immune recognition (Jhunjhunwala *et al*, 2021).

More actively, malignant cells can block T-cell functionality after immune recognition. After binding of the T-cell receptor (TCR) to the presented antigen, the induction of apoptosis in the target cell requires the PD-1 molecule (programmed death-1) on the surface of the T-cells. Cancer cells can evade their induced cell death by expressing PD-L1 or PD-L2, two ligands which bind PD-1 and inhibit the T-cell action (Iwai *et al*, 2002; Latchman *et al*, 2001).

As a tumor grows, it attracts a great variety of immune cells from both branches of the immune system. Ironically, immune cell infiltrates are not always a threat to the tumor but can instead even facilitate and support tumor growth. This is especially true of special subpopulations of macrophages, neutrophils and myeloid progenitors which are normally involved in inflammation and wound healing. These specialized cells have functions such as secreting inflammatory cytokines, growth factors (EGF), and proangiogenic factors like VEGF and FGF2, which tumor cells use to their advantage to grow and engineer a favorable tumor-microenvironment (Hanahan & Weinberg, 2011).

Other factors like epigenetic dysregulation and changes in the microbiome can also modulate systemic immune responses to cancer, shaping the tumor microenvironment and thus aiding or hindering immune detection and destruction of cancerous cells. Indeed, the authors of a more recent review propose changes in microbiome, altered nerve signaling, and epigenetic dysregulation as well as de- and trans-differentiation as new fundamental hallmarks (Senga & Grose, 2021). Considering however, how epigenetic changes and the resulting ability of cancer cells to switch lineages go hand in hand, epigenetics and lineage switching may also be regarded as complementary enabling hallmark.

Genome Instability and Mutation

When the concept of hallmarks was first introduced, the greatest value lay in examining many lines of evidence in a unifying framework of principles which conceptually organized the field of cancer biology. In 2000, Hanahan and Weinberg have described genomic instability as an “acquired characteristic which enables the evolution of cancer cells” toward the other hallmarks (Hanahan & Weinberg, 2000). A decade later, in the second hallmark-paper, genomic instability was discussed with a deeper embedding into its context, recognizing how genome maintenance integrates with epigenetics, telomere maintenance, and the alteration of many signaling pathways which are found at the core of the other hallmarks (Hanahan & Weinberg, 2011).

Genomic instability is not merely a side effect of cancer evolution, but can be viewed as precursor, requirement, and resulting effect at all once. At the center of this particular hallmark is the well-connected network of DNA damage and repair (DDR) genes, which orchestrate the detection and repair of damages, as well as signal integration during proliferation and apoptosis.

DNA Damage and Repair – A double edged sword in cancer development

DNA repair is an orchestrated response to cellular insults and DNA damage. The primary roles envelop surveillance of DNA sequence fidelity during replication and transcription, repair and maintenance of telomeres, and detection and adequate repair of lesions in DNA. Beyond repair capabilities, DNA repair genes are also involved in class switch recombination, creating genetic diversity in variable regions of antibodies (Shang & Meng, 2021), as well as recombination processes during meiosis (San Filippo *et al*, 2008).

The central theme of this discussion focusses on DDR as a barrier to and an aid in cancer development. As such, DDR is a central component in cancer development. When functioning correctly, it provides a barrier to mutagenesis, limitless replication, and evasion of apoptosis. Conversely, when DDR is altered in preneoplastic lesions or cancerous lesions, it can unwittingly aid cancer evolution by accelerating mutagenesis and thus altering the functionality of the very pathways which are altered across almost all hallmarks of cancer. At the same time, however, defective DNA repair pathways leave cancers with exploitable vulnerabilities for cytotoxic chemotherapeutics or targeted agents. To have an in-depth discussion about the role of DDR in cancer development, we need to consider the details of DNA repair mechanisms and DNA repair in context.

DNA Damage

DNA, like any other chemical structure faces limits of stability and can be chemically or physically altered. As the central blueprint of life, alterations at the DNA level can have far reaching consequences all the way from stalling cellular functions to altering a phenotype at the organismal level. The need to protect and preserve the correct sequence information in DNA is reflected in the fact that all domains of life have active forms of DNA repair (Friedberg *et al*, 2005). Therefore, the ability to detect and repair damage in DNA must be critical as it has emerged early in the evolutionary timeline (Aravind *et al*, 1999).

Types and Sources of DNA damage

Broadly, sources of DNA damage can be categorized into exogenous and endogenous sources (Figure 3). Exogenous sources of DNA damage can be physical such as UV-radiation, ionizing radiation, or chemical. The exposure to environmental toxins can alter the structure of the DNA, forming adducts, crosslinks within and between DNA strands, and oxidizing parts of the DNA molecule. Endogenous sources of DNA damage are far more complex than exogenous sources because lesions can arise not just from chemical reactions with compounds present in the cellular environment, *i.e.,* metabolites, but especially from ongoing cellular processes such as replication and transcription.

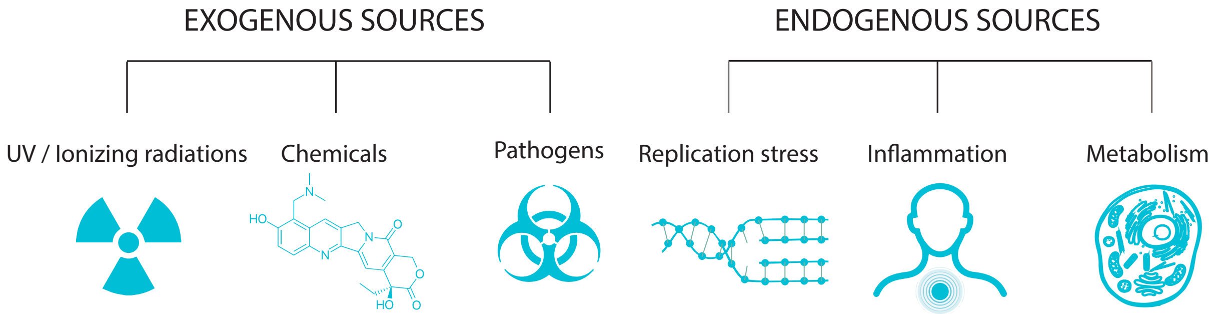


Figure 3. Classification and examples of exogenous and endogenous sources of DNA damage

Biological Impact of DNA Damage

If the structural integrity of DNA is compromised, this can lead to mutations which consequently alters the information content for building downstream RNA and protein molecules, potentially leading to altered functions and phenotypes. Although random mutations are the core engine of evolution, there is a limit to how many mutations a cell can incur without disrupting cellular homeostasis. The average cell encounters up to 100,000 lesions per day (Ciccia & Elledge, 2010; Hoeijmakers, 2009). Without an active DNA damage repair system, cells would very quickly become senescent, apoptotic, highly mutated, or simply cease to function.

DNA Repair

General

Upon DNA damage, the DDR machinery orchestrates an elaborate response which integrates signals and eventually leads to signaling cascades initiating repair action or apoptosis. The architecture of the DNA damage response includes sensors, transducers/ mediators, and effector proteins (Figure 4). These molecules work together to effectively form a chain of events capable of sensing the damage, removing damaged DNA (nicking and resection), synthesizing new DNA, and repairing the backbone incision (ligation) (Zhou & Elledge, 2000). Of the sensors the apical kinases ATM and ATR are best studied. ATM is recruited by the MRE11-RAD50-NBS (MRN) complex, which senses DNA double strand breaks. ATM then phosphorylates the MRN complex as well as other downstream proteins such as p53 and CHK2 to halt the cell cycle and give time to initiate repair (Harper & Elledge, 2007). The other well-studied sensor, ATR integrates damage sense signals from single stranded DNA coated with RPA to subsequently phosphorylate CHK1 launch repair through recruitment of the 911-complex (RAD9, RAD1, HUS1) (Harper & Elledge, 2007). After the initial detection and signaling of the apical kinases, the transducers and mediator proteins act to transduce and amplify the signal by recruiting more downstream repair proteins, or by acting as a physical anchoring platform onto which repair protein complexes can be assembled (Harper & Elledge, 2007). Well known examples of this class are 53BP1 and BRCA1, active in NHEJ and HR respectively. Finally, various effector proteins carry out the functions of the signaling cascade. Arguably, one of the most central effector proteins is the transcription factor p53, which relays signals to a multitude of targets, effecting changes in cell cycle state or inducing apoptosis. Other effectors have a more direct role, such as enzymes involved in DNA break resection, resynthesis of DNA, and ligation of the backbone. Beyond the obvious role in mending DNA lesions, effectors of the DNA damage response also modulate other ongoing cellular processes, including transcription and chromatin remodeling, metabolism, autophagy, and cellular energetics, as well as RNA processing. This allows to halt and accelerate processes which interfere or support DNA repair respectively (Harper & Elledge, 2007; Jackson & Bartek, 2009; Zhou & Elledge, 2000). In summary, the interplay between the immense number of DDR involved factors regulates a complete cellular response to insults to DNA integrity. To gain an appreciation of the diversity of processes involved in DNA repair, the individual signaling pathways dealing with specific types of DNA lesions will be discussed in more detail.

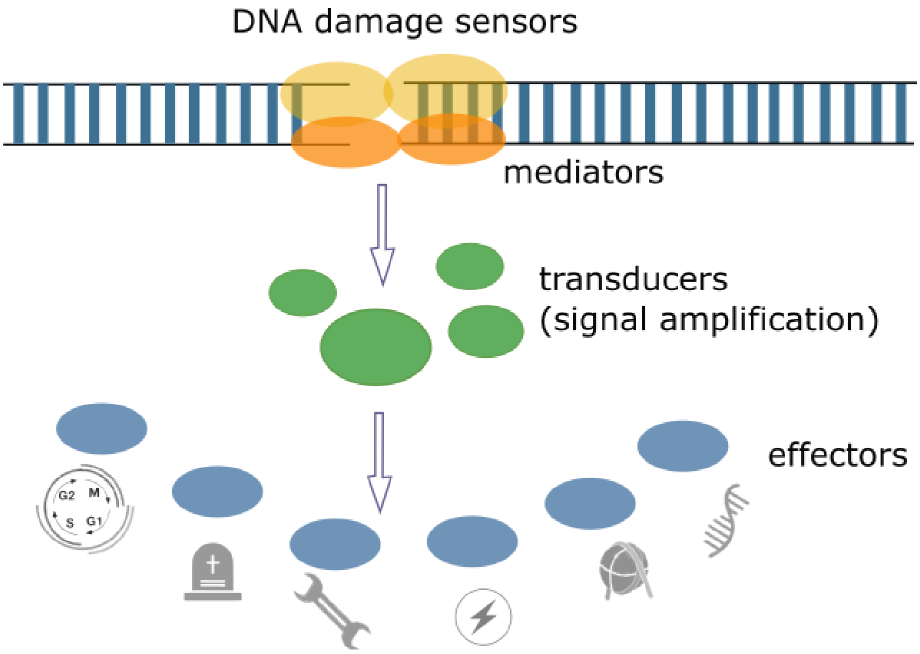


Figure 4. Overview of the general architecture of the DNA damage response

Direct Repair

The simplest form of DNA repair involved direct reversal of damage on DNA bases, foregoing incision into the backbone or replacing the damaged nucleotide in a single catalytic reaction involving one enzyme. The types of damages which can be repaired via this mechanism involve small non-bulky alkylations such as methylated guanine (6-O-methylguanine). This base alteration, if left unrepaired, leads to G:C > A:T mutations because 6-O-methylguanine falsely pairs with thymine and thus gets changed to an A in subsequent cell divisions. The enzyme methylguanine DNA methyltransferase (MGMT) can repair such lesions by covalently binding and removing the offending alkyl group in an irreversible reaction, inactivating the enzyme (Figure 5). Cancer cells can downregulate the expression of or mutate MGMT to increase the mutation rate. Conveniently, this opens a therapeutic vulnerability making cancer cells more susceptible to alkylating chemotherapy agents like temolzolomide (Kelley *et al*, 2014).

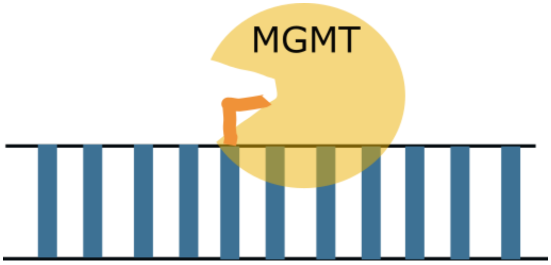


Figure 5. Schematic overview of MGMT acting in Direct repair

Translesion Synthesis

DNA repair proteins include specialized polymerases which synthesize new DNA during the process of repair (Figure 6). Some specialized polymerases function primarily within one distinct repair pathway, as is the case with polymerase beta (POLB), which functions in base excision repair. Another example is the polymerase theta (POLQ), which operates in double strand break repair. POLQ is error prone and possesses no proof-reading function and thus can directly bypass lesions that arise especially in the context of stalled replication forks. Often using the damaged DNA strand as template, this results in a highly mutagenic form of repair. Nonetheless, translesion synthesis (TLS) serves a valuable function by providing immediate repair allowing to keep cellular processes such as ongoing replication going (Maiorano *et al*, 2021). In cancerous cells this is yet another example of how a DNA repair factor is a double-edged sword in genomic stability. On the one hand, POLQ activity is mutagenic and helps the cancer evolve. On the other hand, in the context of enhanced replication, saving stalled replication forks from collapse or promoting repair after collapse, prevents excessive detrimental genomic stability in cancer cells (Maiorano *et al*, 2021). For this reason POLQ is also an attractive target for anti-cancer therapy, especially in cancers with a deficiency in homologous recombination, which are doubly dependent on POLQ for its TLS activity and for its role in double strand break repair (Schrempf *et al*, 2021).

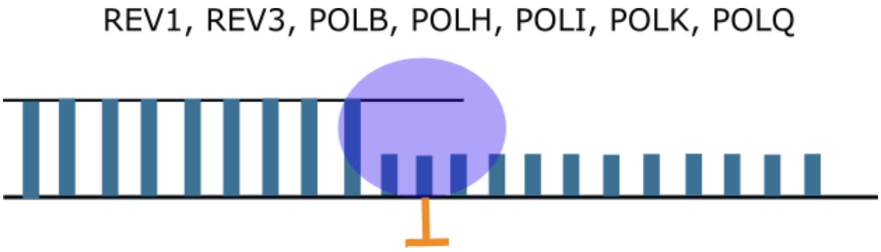


Figure 6. Schematic overview of proteins involved in Translesion synthesis

Base Excision Repair

DNA lesions which affect bases without disturbing the helical backbone of DNA are preferentially repaired via base excision repair (BER). The first step of repair involves the removal of the damaged base via glycosylases, leaving an abasic site behind. Next, specialized endonucleases create a nick in the DNA backbone to prepare for resynthesis of the missing nucleotides, using the complementary strand as template. If only one base was removed, this is termed short-patch BER, whereas if small stretches of nucleotides (2- ~10 bp) were removed, long-patch BER is employed. In the case of short-patch BER, DNA polymerases fill the gap (abasic site) and ligases seal the nick in the backbone. For long-patch repair, ligases such as POLB synthesize a stretch of DNA, reaching over the abasic site, displacing the original strand and leaving a 5'-overhang which needs to be hydrolytically cleaved by endonucleases before relegation of the backbone (Krokan & Bjørås, 2013) (Figure 7). Single base substitutions are the most abundant type of mutations in cancer, placing BER as a central player in preventing this type of mutagenesis (Pleasance *et al*, 2010). Due to the high replication rate in cancer, and the increased ROS damage, cancer cells become somewhat reliant on BER to limit excessive DNA damage and mutagenesis (Grundy & Parsons, 2020). At the same time the BER pathways ameliorates the effects of base-damaging chemotherapeutic agents and radiation therapy (Adhikari *et al*, 2012). Therefore, BER enzymes constitute attractive targets to maximize the response to chemotherapeutics by overwhelming the DNA damage response in malignant cells (Grundy & Parsons, 2020; Dianov, 2011; Kelley *et al*, 2014).

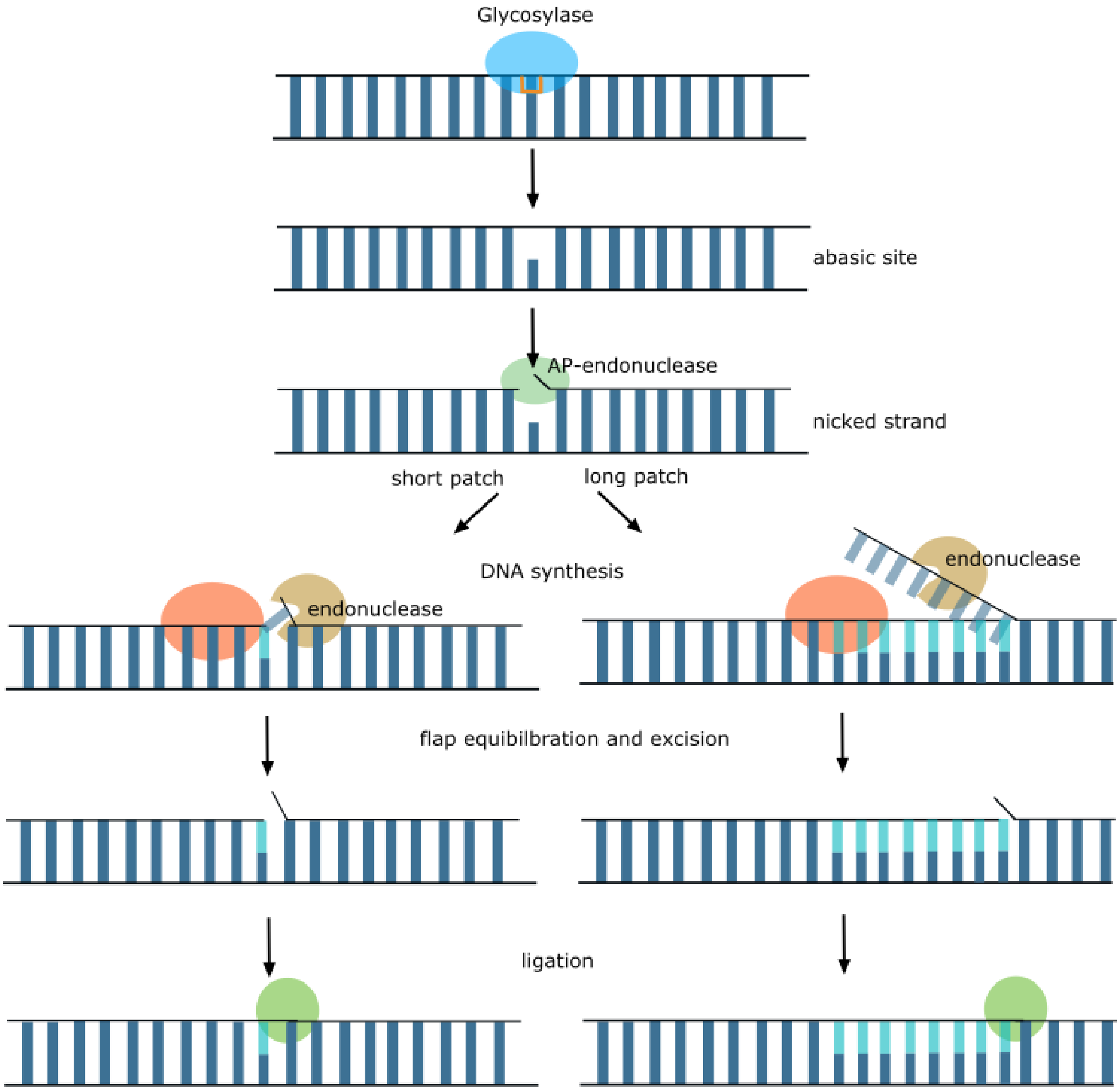


Figure 7. Schematic overview of base excision repair

Nucleotide Excision Repair

Bulky DNA adducts with helix distortion are recognized and cleared by nucleotide excision repair (NER) (Figure 8). NER can be subdivided into two pathways. Global genome NER (GG-NER) monitors the entire genome for lesions, whereas transcription coupled NER (TC-NER) specializes in detecting lesions which are encountered in the context of transcription when the RNA polymerase II complex stalls. In this context, lesions need to be removed for the transcription machinery to move forward. The type of damage recognized by both systems include ROS induced cyclopurenes, UV-mediated cyclobutene-pyrimidine dimers (CPDs) and pyrimidine-(6-4)-pyrimidone photoproducts (6-4PPs). Further types of damage include, bulky adducts caused by chemical agents, such as benzo[a]pyrene (BaP) or the chemotherapeutic drug cis-platin (Marteijn *et al*, 2014).

The damage sensor in GG-NER is XPC, aided by RAD23B and DDB2. In TC-NER, the same function is covered by CSA and CSB, which together with USP7, UVSSA, and XAB2 hold back the RNA-Pol2 complex to free the lesion for impending repair. After the initial damage recognition step, the sub pathways converge and continue with recruiting XPA and the TFIIH complex to the lesion. The TFIIH complex contains the helicases XPB and XPD which work together to unwind the DNA around the lesion in the 3'-5' and 5'-3' direction respectively. The stabilized unwound structure gives access to XPG (3'-incision) and the XPF/ERCC1 heterodimer for the 5' incision. Following the incision step, the lesion and surrounding nucleotides are removed, leaving the opposite DNA strand to serve as template for polymerases POLD, POLK, or POLE. Finally, the nick is sealed by LIG1 or LIG3 (Marteijn *et al*, 2014). Like the other DDR pathways, NER also takes on contradictory roles in cancer evolution. On the one hand, NER is a barrier to cancer development, because NER deficiency has been observed to predispose to a variety of conditions, including neurological abnormalities and skin cancer (Friedberg, 2001). Per contra, in the context of cancer, NER has been observed to act as a mechanism of resistance to cisplatin, helping cancerous cells to adapt and evolve (Duan *et al*, 2020).

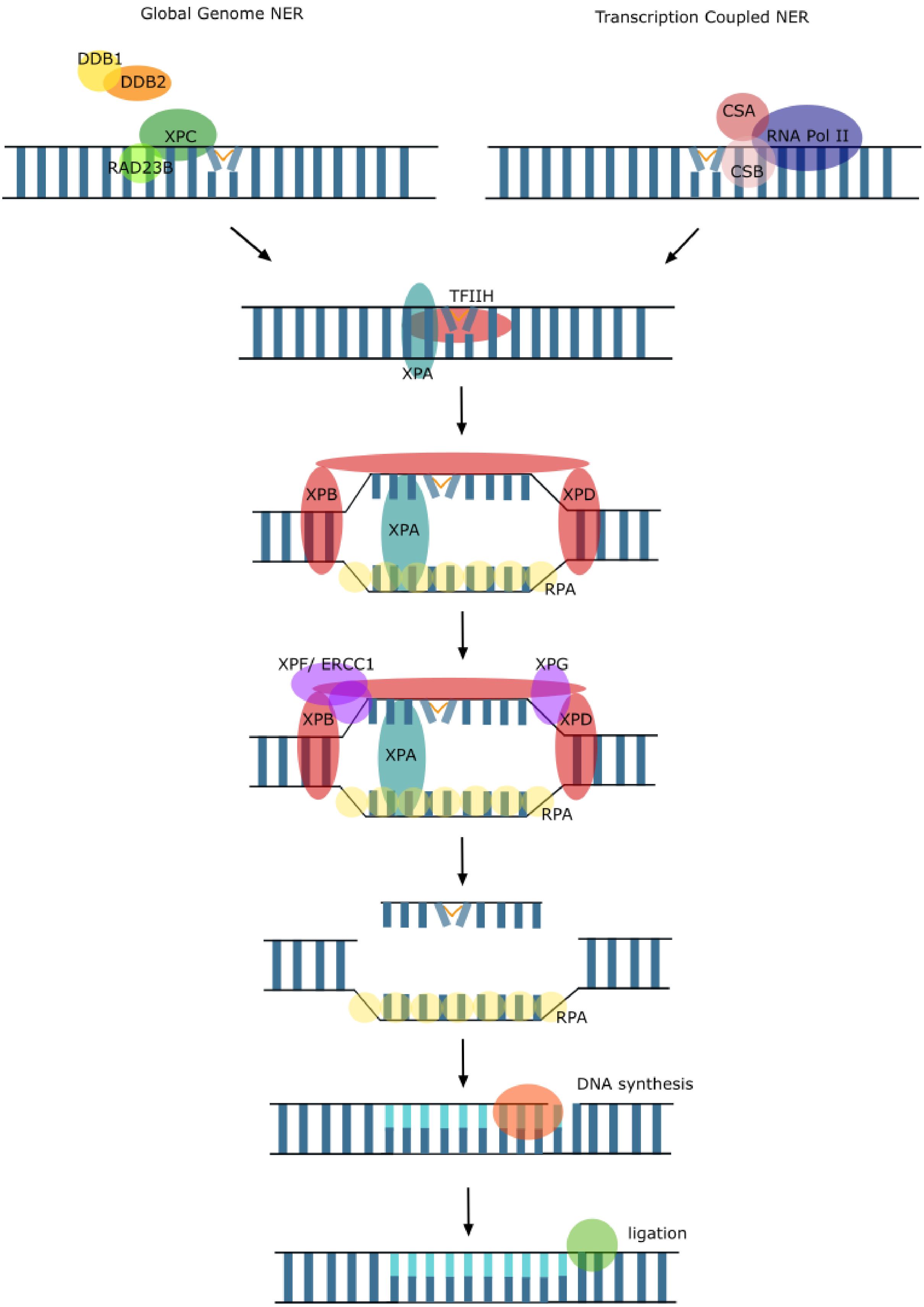


Figure 8. Schematic overview of Nucleotide excision repair

Mismatch Repair

During replication, recombination, and even repair processes, errors in base pairing may occur due to low fidelity of replication or translesion polymerases (Li, 2008). Furthermore, slippage of the replication machinery may also result in small insertions and deletions (+/- 1 bp), especially at stretches of repeat sequences. This generates a mismatch between the indel and the opposite strand. If left unrepaired, mismatches become fixed in the genome as single point mutations and small insertions and deletions. Not surprisingly, inactivation of mismatch repair is associated with increased mutagenesis, as is observed in microsatellite instable cancers (Li & Martin, 2016). Under normal conditions, the mismatch sensor is the MutS complex, a heterodimer of MSH2 and MSH6. This complex is specialized to recognize single base-pair mismatches. Longer mismatches, i.e. small indels up to 10 bp, are better recognized by the MutSβ complex, formed by MSH2 and MSH3 (Li, 2008). Upon encountering a mismatch, the detection complexes undergo an ATP-dependent change in conformation, initiating the recruitment of downstream factors MLH1, MLH3, and PMS2. In human cells, MutH is thought to play a crucial role in strand discrimination, using a strand specific nick to separate the damaged strand from the template strand (Li, 2008; Kunkel & Erie, 2005). Together with proliferating cell nuclear antigen (PCNA) and replication factor C (RPC), the strand containing the mismatch is nicked and thus marked for degradation and subsequent repair. The exonuclease EXO1 is responsible for removing nucleotides along the nicked strand, in both the 3’-5’ and 5’-3’ direction, depending on where the mismatch was detected (Li, 2008). The nicked strand is excised up to the mismatch and slightly further, leaving a stretch of single stranded DNA, which is coated and protected by replication protein A (RPA). Finally, the gap is filled with newly synthesized DNA by POLD and the nick is ligated by LIG1 (Li, 2008) (Figure 9).

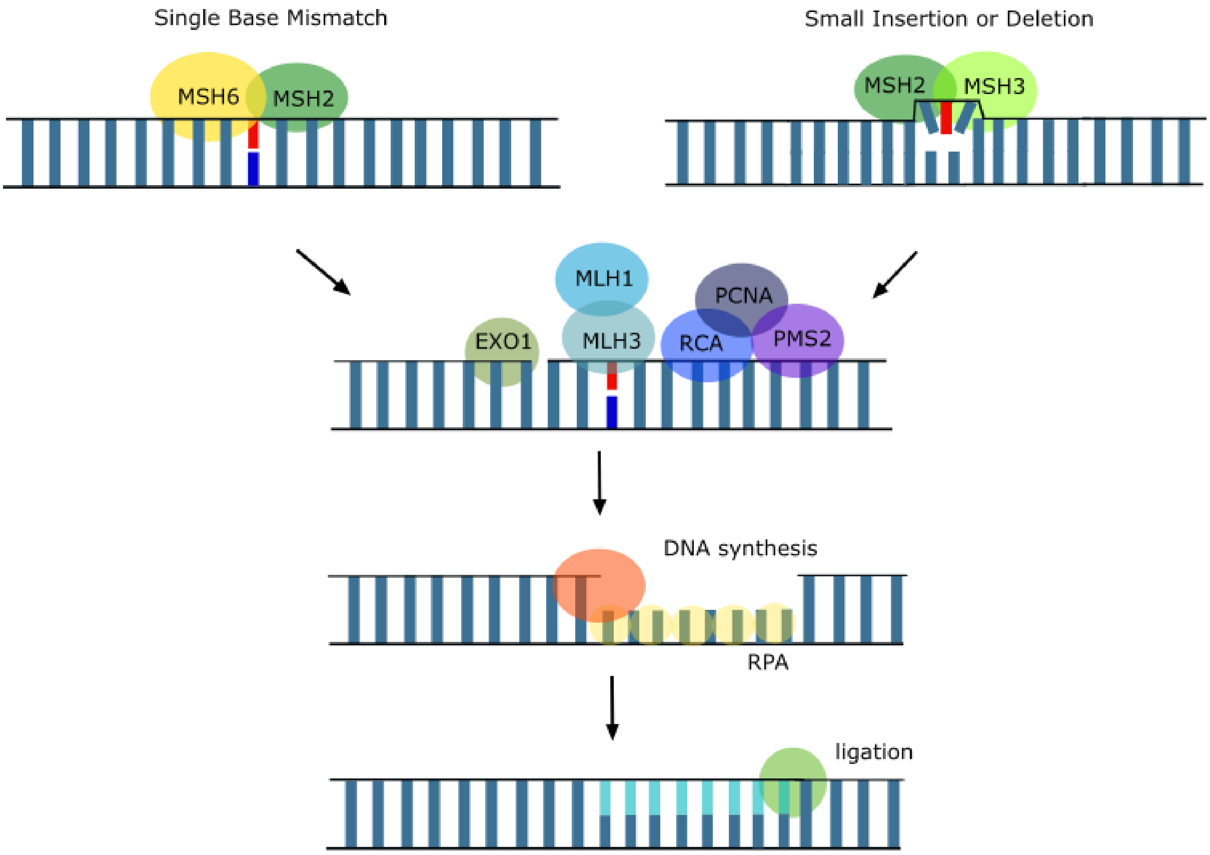


Figure 9. Schematic overview of mismatch repair

Crosslink Repair (FA)

Interstrand crosslinks (ICL) are induced by exogenous chemicals such as platinum compounds with two alkylating functional groups. Endogenous sources of these lesions include metabolic byproducts such as formaldehyde and acetylaldehyde (Langevin *et al*, 2011; Rosado *et al*, 2011). Interstrand crosslinks are extremely harmful because they prevent transcription and replication. If left unrepaired, crosslinks can lead to subsequent double strand breaks, mitotic failure, and apoptosis (Ceccaldi *et al*, 2016). The Fanconi Anemia pathway (FA) has evolved to take care of such lesions, specifically in the context of replication during S-phase. Stalling of replication forks at ICLs triggers detection of the lesion by the FANCM-FAAP24-MHF1-MHF2 complex, which recruits the Fanconi anemia core complex (FANCA, FANCB, FANCC, FANCF, FANCG, FANCL, FAAP100, FAAP20 and FANCM). After the initial steps, FANCL and UBE2T serve to ubiquitylate the FANCD2-FANCI heterodimer, together referred to as the ID-complex. The ID complex is a recruiting platform for other repair factors, including the Fanconi-associated nuclease (FAN1) and the NER associated endonuclease ERCC4. The interaction between these proteins allows for nucleolytic incision of the backbone near the lesion and subsequent untangling of the previously crosslinked DNA-heteroduplex. This process leaves behind two different intermediate damage products, one stretch of double stranded DNA with a gap in one strand (intermediate product a), and another stretch of double stranded DNA with a break in both strands (intermediate product b). Consequently, different strategies for downstream repair are adopted. The intermediate product a is resolved using translesion polymerases to fill the gap. Intermediate product b, instead, needs to be resolved using double strand break repair, preferably homologous recombination, using the sister chromatic as a template, which in this case is the just repaired product a (Ceccaldi *et al*, 2016) (Figure 10).

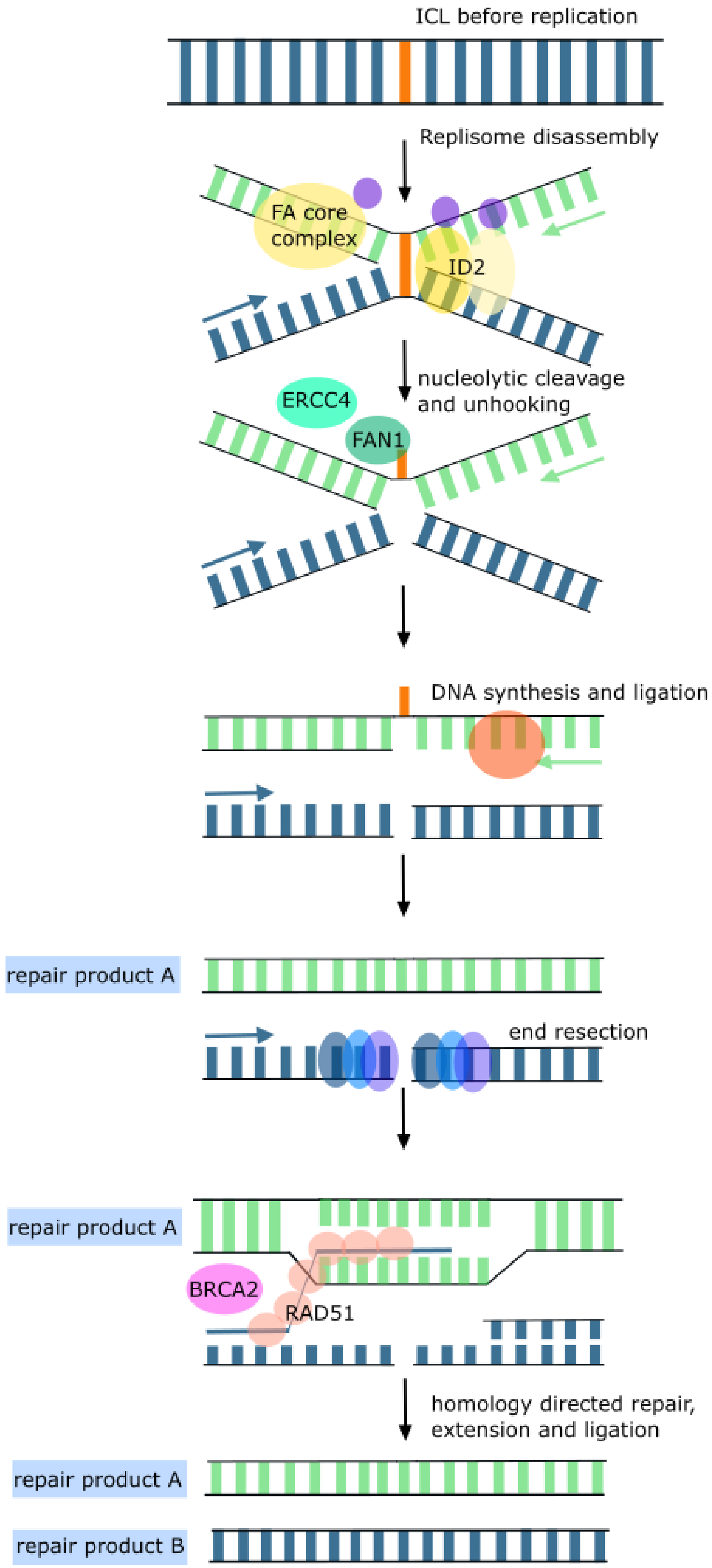


Figure 10. Schematic overview of crosslink repair (Fanconi Anemia pathway)

Single Strand Break Repair

Single strand breaks (SSBs) are one of the most common lesions in cells. They are formed as a consequence of oxidative stress from metabolic activity, spontaneous decay of the sugar backbone, and abandoned DNA repair intermediates, or even faulty activity of DNA topoisomerase 1 (TOP1) (Caldecott, 2008). Unrepaired SSBs can quickly become toxic if left unrepaired. Single stranded DNA stalls replication forks, leading to fork collapse and likely induction of double strand breaks (DSBs). Despite multiple available double strand break repair pathways, a sharp increase in DSBs due to unrepaired SSBs can lead to an overload of the available repair avenues, eventually leading to genomic instability or apoptosis (Caldecott, 2008). SSBs are mainly detected by PARP1, which binds to SSBs and signals for the recruitment of other repair proteins through branched chains of poly(ADP-ribose), a process also called PARylation. The break site is then stabilized mainly with the help of XRCC1. SSBs are characterized by damaged nuclei flanking the 3’ and 5’ of the broken strand. Depending on the type of damage on the 3’/5’-termini, different processing enzymes resect and restore the 3’-hydroxy and 5’-phosphate moieties before gap filling and ligation can commence. Most notably, APE1 serves this function for damaged 3’ends, whereas PNKP can restore both 3’ and 5’ flanking nucleotides. Other specialized enzymes such as TDP1 can remove stalled TOP1 from 3’ ends, whereas APTX can remove AMP from 5’ ends, left over from aborted ligase activity (Mei *et al*, 2020). Once the flanking ends are restored, the gap is filled by POLB, or sometimes POLD or POLE. During this process, the gap filling can replace only the missing nucleotides or extend further than the initial gap. In the latter case, FEN1 is needed to help remove the displaced nucleotides in the 5’-direction. The final ligation step is performed by either LIG1 or LIG3 (Caldecott, 2008) (Figure 11).

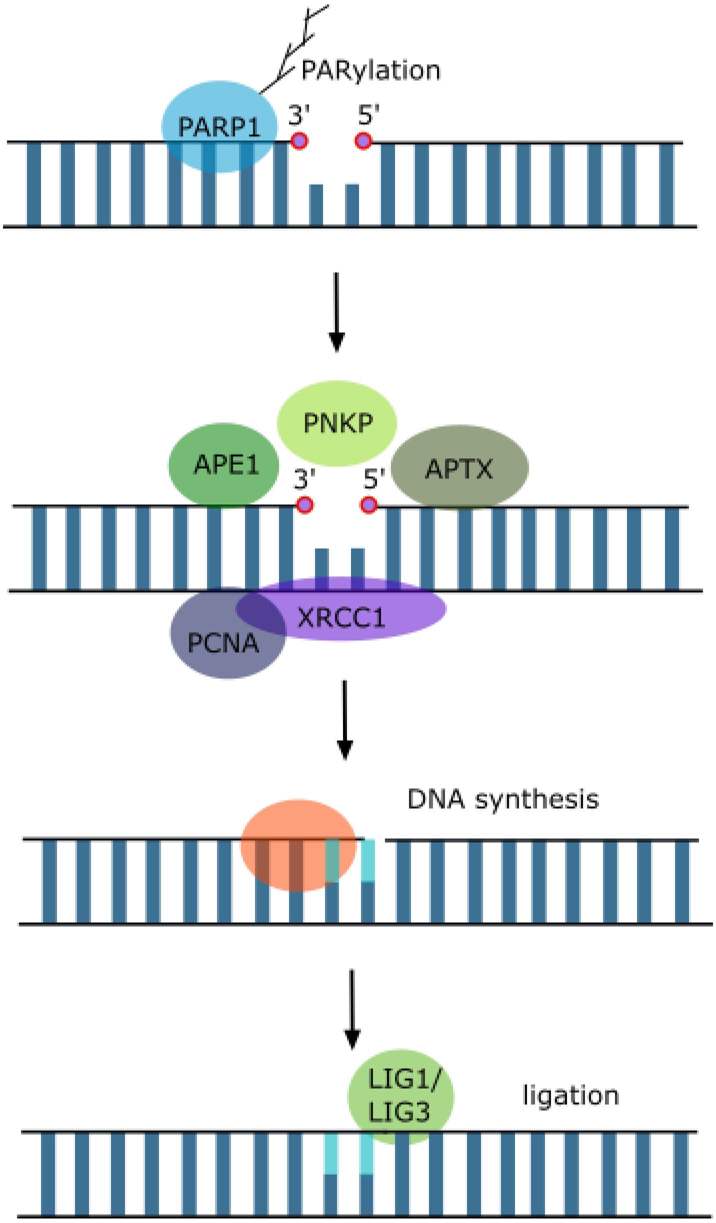


Figure 11. Schematic overview of single strand break repair

Double Strand Break Repair

Double strand breaks (DSBs) are among the most toxic lesions a cell can encounter and often lead to cell death if left unrepaired (Friedberg *et al*, 2005). Furthermore, DSBs create opportunity for mutations (indels) and translocations, accelerating genomic instability. Exogenous sources of DSBs include ionizing radiation or chemical agents such as topoisomerase inhibitors camptothecin and etoposide, or radiomimetic drugs like bleomycin (Mehta & Haber, 2014). Endogenously, metabolic activity, reactive nitrogen and oxygen species, and degraded or unrepaired DNA repair intermediates, can lead to the formation of DSBs. This is especially true in the context of replication stress, where unrepaired lesions lead to the collapse of the replication fork, often leading to DSB formation (Nickoloff *et al*, 2021). Importantly, there are also programmed cellular processes which require the formation and accurate repair of DNA double strand breaks. These include meiosis and recombination, DNA unwinding with topoisomerase 2, and V(D)J- and class switch recombination (Shang & Meng, 2021; San Filippo *et al*, 2008). Eukaryotic cells have developed a set of partly overlapping but closely integrated repair pathways to clear these types of lesions. The choice of pathway depends on cell cycle, cell type, and chemical configuration of the lesion site (Symington & Gautier, 2011; Shrivastav *et al*, 2008). Mainly, after the initial step of lesion recognition, the choice of pathway depends on the regulation of end resection of the broken DNA strands (Figure 12). Non-homologous end joining (NHEJ) is the pathway of choice for unresected compatible DNA ends. This pathway is functional throughout the cell cycle as it requires no repair template but rather joins the blunt DNA ends in a non-conservative way, creating small indels. Different from NHEJ, there are three other pathways which require end resection to proceed: homologous recombination (HR), single strand annealing (SSA), and polymerase theta-mediated end joining (TMEJ) (Chang *et al*, 2017). HR is the most faithful pathway, accurately repairing lesions during S and G2 phases, where a sister chromatid is available as repair template. SSA and TMEJ, like NHEJ are mutagenic pathways, leading to variable length deletions and insertions. SSA uses stretches of repeats along the resected DNA ends to anneal the DNA strands, resulting in long deletions between the homologous repeat sequences. Similarly, TMEJ uses sequences of microhomology to join the resected DNA strands but uses translesion synthesis polymerase POLQ to fill the gaps to the left and right of the newly joined strands, creating small microhomology flanked deletions.

Since the end resection is a major determining factor in repair pathway choice, this process is heavily regulated. Generally, end resection happens in two stages. First, the DNA ends are clipped, also called limited end resection, which is performed by the MRN-complex. This step makes regions of microhomology accessible, clearing the way for TMEJ, but precluding the progression of NHEJ. Second, DNA helicases and nucleases like DNA2, BLM, WRN, CtIP, and EXO1 synergize to complete further end resection, preparing the lesion site for commencement of either SSA or HR, depending on the cell cycle stage (Shrivastav *et al*, 2008; Symington & Gautier, 2011).

Double Strand Break Repair – Non-homologous End-Joining

Since NHEJ requires a blunt DSB, the first step in the process is the recruitment of break end protectors Ku70/Ku80, together with their co-factor DNA-PKcs. Together, these factors tether the broken DNA ends and provide a platform for other DNA repair factors to act. If required, endonucleases such as PNKP or ARTEMIS (SNM1C) can perform minimal end resection to restore compatibility of DNA ends before joining. Finally, ligation is performed by LIG4, XRCC4, and XLF (Lieber, 2010) (Figure 12). Although NHEJ is a mutagenic repair process, it serves important biological functions, such as the generation of immunoglobulin variety (Shang & Meng, 2021). Furthermore, NHEJ offers fast repair of DSBs throughout the cell cycle, limiting excessive genomic instability at the price of only small indels.

Double Strand Break Repair – Homologous Recombination

HR requires extensive end resection up to 1000 bp long, achieved by DNA2, EXO1, and BLM (San Filippo *et al*, 2008). The otherwise fragile ssDNA is covered and protected by RPA2, which helps to avoid chemical modification or formation of secondary DNA loops. Following this, BRCA1 and BRCA2 recruit RAD51, which is critical in forming a search filament that is able to invade the sister chromatid, finding the correct homologous sequence for templated repair (Wright *et al*, 2018). Once the correct sequence is found, the invading 3'-strand serves as a primer for DNA synthesis. The crossover construct, termed Holliday junction, needs to be resolved nucleolytically, separating the restored strands (Wright *et al*, 2018). The final step encompasses the ligation of newly resynthesized strand (Figure 12). HR is a high-fidelity pathway which is mainly active in the G2 and S phases of the cell cycle, where the presence of a sister chromatid template allows for accurate restoring of the sequence at the break site. Unsurprisingly, defects in this integral DNA repair pathway increase the risk of developing certain cancers, notably breast and endometrial cancers, but also prostate and pancreatic cancer (Nguyen *et al*, 2020; Mekonnen *et al*, 2022). Thus, HR acts as a barrier to cancer development.

Double Strand Break Repair – Single Strand Annealing

After extensive end resection, long stretches of homologous sequences are uncovered, often longer than 200 bp. In SSA, the broken ends of the DNA are brought together by overlapping the flanking homology strands, which are annealed by RAD52. The unpaired (non-homologous) single stranded flaps to either side of the match are removed by the ERCC1-XPF dimer (Figure 12). The factors involved in the downstream filling of the gaps and ligation of the backbone, remain unknown so far (Chang *et al*, 2017). SSA is a highly mutagenic repair process, resulting in large deletions. Furthermore, if DSBs occur at different sites in the genome simultaneously, the process of annealing homologous regions can lead to translocations, driving gross chromosomal aberrations. SSA with its canonical mutagenic function has been implied in the etiology of different cancers and may even play a role in modulating the response to DSB inducing chemotherapy (Blasiak, 2021).

Double Strand Break Repair – Theta Mediated End-Joining

In TMEJ, instead of committing to extensive end resection, the pathway makes use of short microhomologies unveiled upon the initial short-range end resection (Figure 12). Because of this, TMEJ has also been referred to as microhomology-mediated end joining (MMEJ) in the literature. Opposed to SSA, the microhomology regions are much shorter and annealing is not RAD52 dependent. Consequently, this means resulting deletions are much shorter than in SSA. The final steps in this repair pathway consist of gap filling, mediated by POLQ, flap removal by XPF/ERCC1 and ligation of the backbone by LIG1 and LIG3 (Schrempf *et al*, 2021; Seol *et al*, 2018). Beside the short deletions characteristic for TMEJ activity, POLQ is also able to use break overhangs as template, synthesize stretches of new DNA and anneal this newly synthesized DNA before ligation. This effectively creates stretches of de-novo DNA, which is identical in sequence to the original overhangs, called templated insertions (Schimmel *et al*, 2019).

It seems counterintuitive for evolution to have generated highly mutagenic repair pathways such as SSA and TMEJ. Still, the evolutionary conservation of this pathway speaks to the necessity of its existence. A standing hypothesis in the field is that TMEJ is preferentially active at sites of persistently stalled replication forks, or where the sister chromatid cannot be used as a repair template by HR, due to breaks, crosslinks, or other unrepaired lesions present in the sister strand (Schimmel *et al*, 2019). Conventionally, collapsed replication forks and failure of HR would lead to cell death. However, TMEJ offers another alternative to repair the damage at the cost of incurring indels. Therefore, in the context of oncogene induced replication stress and increased DNA damage, TMEJ may act as a mitigator of deadly genomic instability. The pathway at once protects from apoptosis inducing DNA damage and enables tumor evolution through mutagenic repair activity, again illustrating the double-edged sword DDR pathways play in cancer development.

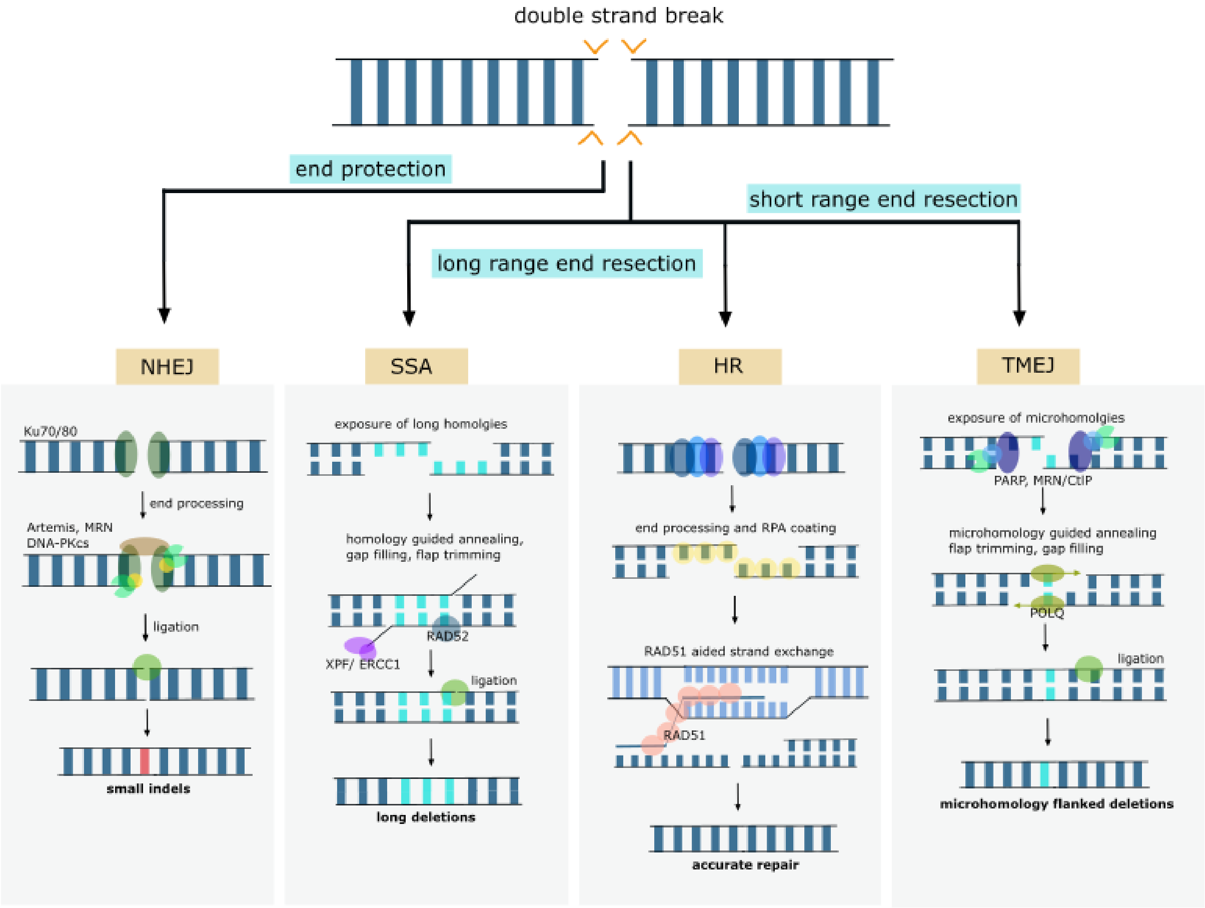

double strand break
end protection
short range end resection
long range end resection
NHEJ
SSA
HR
TMEJ
Ku70/80
end processing
Artemis, MRN
DNA-PKcs
ligation
small indels
exposure of long homolgies
homology guided annealing,
gap filling, flap trimming
RAD52
XPF/ ERCC1
ligation
long deletions
end processing and RPA coating
RAD51 aided strand exchange
RAD51
accurate repair
exposure of microhomolgies
PARP, MRN/CtIP
microhomology guided annealing
flap trimming, gap filling
POLQ
ligation
microhomology flanked deletions

Figure 12. Schematic overview of the competing pathways in double strand break repair.

The DNA damage response in context

Although it aids understanding to think of a common architecture and individual DNA repair pathways that form linear top to bottom signaling cascades, the reality of DNA repair is more akin to a tightly interwoven network with much crosstalk and overlap. DNA damage and repair pathways which overlap might even compete for the same lesion substrate. Evolutionarily, the crosstalk between pathways ensures multiple chances to repair a given lesion. However, for a functional DDR machinery, tight regulation is required. Both, the cell cycle regulation, as well as signaling feedback loops provide this level of regulation. Furthermore, cell identity and tissue architecture impose additional constraints on the interplay between the response to damage and repair. How different cells in their respective tissue contexts integrate these signals decides whether apoptosis is entered, differentiation induced, or repair is attempted. The latter option is then further governed by cell cycle stage, metabolic profile, and of course the type and structure of the genetic lesion.

Double Strand Breaks as a Model to Understand DNA Repair in Context of Cell Cycle Regulation and Tissue Specificity

To understand the true intricacy of the DDR, a more in-depth perspective is needed. An ideal model system to study the complexity of DNA repair pathway choice in context is the cellular response to breaks induced by the Cas9 endonuclease used in the CRISPR based genome editing system. First, depending on the variant of the nuclease used, the type of lesion and exact location is known, which allows to sequence and characterize all possible editing outcomes (Hussmann *et al*, 2021) giving insight into the complex regulation of the DDR in context. The following two reviews discuss DNA repair in context with a specific focus on the classical lesion induced by Cas9: DNA double strand breaks.

frontiers in Genetics

REVIEW
published: 03 September 2021
doi: 10.3389/fgene.2021.728520

Tissue Specific DNA Repair Outcomes Shape the Landscape of Genome Editing

Mathilde Meyenberg[1,2†], *Joana Ferreira da Silva*[1,2†] *and Joanna I. Loizou*[1,2*]

[1]*CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria,* [2]*Institute of Cancer Research, Department of Medicine I, Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria*

Edited by:
Sylvie M. Noordermeer,
Leiden University Medical Center,
Netherlands

Reviewed by:
Michael Aregger,
National Cancer Institute (NCI),
United States
Chris Richardson,
University of California,
Santa Barbara, United States

Correspondence:
Joanna I. Loizou
joanna.loizou@meduniwien.ac.at

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *21 June 2021*
Accepted: *05 August 2021*
Published: *03 September 2021*

Citation:
Meyenberg M, Ferreira da Silva J and Loizou JI (2021) Tissue Specific DNA Repair Outcomes Shape the Landscape of Genome Editing. Front. Genet. 12:728520. doi: 10.3389/fgene.2021.728520

The use of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 has moved from bench to bedside in less than 10 years, realising the vision of correcting disease through genome editing. The accuracy and safety of this approach relies on the precise control of DNA damage and repair processes to achieve the desired editing outcomes. Strategies for modulating pathway choice for repairing CRISPR-mediated DNA double-strand breaks (DSBs) have advanced the genome editing field. However, the promise of correcting genetic diseases with CRISPR-Cas9 based therapies is restrained by a lack of insight into controlling desired editing outcomes in cells of different tissue origin. Here, we review recent developments and urge for a greater understanding of tissue specific DNA repair processes of CRISPR-induced DNA breaks. We propose that integrated mapping of tissue specific DNA repair processes will fundamentally empower the implementation of precise and safe genome editing therapies for a larger variety of diseases.

Keywords: CRISPR-Cas9, genome editing, DNA double-strand break, homology directed repair, non-homologous end-joining, microhomology mediated end-joining, tissue specific DNA repair, tissue stem cells

DNA DOUBLE-STRAND BREAK REPAIR: THE FOUNDATION FOR GENOME EDITING

Genome stability is constantly challenged by endogenous and exogenous factors that threaten the integrity of DNA. If DNA damage is incorrectly repaired, this leads to mutations or widespread genome aberrations that impair cell function and survival. Intracellular reactive oxygen species (ROS) and reactive nitrogen species (RNS), reactive metabolites, and replication stress synergise with exogenous genotoxic sources of damage, such as radiation, chemical exposure, viral, or bacterial infections to challenge genomic stability. In order to protect genome integrity, cells have evolved sophisticated mechanisms to detect, signal, and repair diverse DNA lesions, known as the DNA damage response.

Biological Significance of DNA Double-Strand Breaks

DNA double-strand breaks (DSBs) are amongst the most toxic lesions cells can encounter, as both DNA ends become topologically separated. For this reason, DSBs are induced in cancer therapy, either through ionising radiation or by preventing their repair *via* topoisomerase inhibition. In contrast, formation of endogenous DSBs is an integral part of fundamental cellular processes, such as the generation of immune receptor diversity, meiosis, and ageing (Jackson and Bartek, 2009). Therefore, DSB repair is an essential and vital cellular process. Overall, DSBs are repaired in two ways: re-ligation of the DNA ends through pathways such as non-homologous end-joining (NHEJ) and microhomology-mediated end-joining (MMEJ), or templated repair from a separate donor DNA molecule, through a process called homology directed repair (HDR; Yeh et al., 2019). A key aspect in the repair of DSBs in human cells is the competition between these two types of repair, with end-joining pathways being favoured over templated repair, in a cell-cycle dependent manner.

Cas9-Induced DNA Double-Strand Breaks: The Genome Editing Revolution

During the early 2000s, site-specific DSB generation, induced by engineered endonucleases, became an increasingly useful approach to edit the genome. Zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) have been successfully used as genome editing tools in mammalian cells (Miller et al., 2011; Hossain et al., 2015). However, inherent difficulties with protein design, synthesis, and validation remained a challenge to the widespread implementation of these nuclease-based editing technologies. This limitation was solved upon the discovery of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), a breakthrough that revolutionised the field of genome editing (Jinek et al., 2012). CRISPR and the associated Cas9 endonuclease (CRISPR-Cas9) were initially identified as an antiviral defence mechanism in prokaryotes, but rapidly became a powerful genome editing tool in eukaryotic cells (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013). The CRISPR-Cas9 system, guided by a single-guide RNA (sgRNA), targets a particular region of the genome, generating a DNA DSB that subsequently activates the cellular DNA repair machinery. The considerable ease of manipulating the sgRNA, compared to ZFNs and TALENs, has served an important role in the CRISPR revolution, creating the possibility to edit a wide variety of cell types and organisms, with unprecedent precision and efficiency. Importantly, besides being a powerful approach for functional genetic studies, CRISPR-Cas9 approaches hold great promise for the correction of genetic disorders caused by specific alterations in the genome, with recent clinical trials reporting promising results (Wang et al., 2020; Frangoul et al., 2021). However, most current clinical applications are still based on the disruption of a genetic sequence, rather than a precise edit. Moreover, the safety and efficiency of CRISPR-based therapies still need to be closely addressed and an important step is the fundamental understanding of the tissue specific DNA repair pathway choice, following a Cas9-induced DSB. The focus of this review will be on the DSB-dependent genome editing technologies which make use of *Streptococcus pyogenes* Cas9 (SpCas9), generating a blunt end at a targeted genomic site. We direct readers to the following additional technical advances that have expanded the CRISPR-toolbox and fall outside the focus of this review: engineered Cas9 nucleases with higher fidelity (Kleinstiver et al., 2016) and broader specificity (Kleinstiver et al., 2015; Walton et al., 2020), DSB-independent applications that increase the range of possible editing outcomes, such as DNA base editors (Komor et al., 2016; Gaudelli et al., 2017) and prime editing (Anzalone et al., 2019), CRISPR-mediated regulation of gene expression (Gilbert et al., 2013; Qi et al., 2013; Nuñez et al., 2021), and new CRISPR nucleases repurposed for genome editing (Zetsche et al., 2015).

REPAIR OF Cas9-INDUCED DNA DOUBLE-STRAND BREAKS

Cell Cycle Regulates DNA Double-Strand Break Repair Pathway Choice

After a Cas9-induced DSB, repair pathway choice is a crucial factor in determining the editing outcome. The blunt ends of the DNA break can be protected by the Ku70/80 heterodimer, fating the lesion for repair by NHEJ. Conversely, 5'–3' resection of DNA ends reveals sequence homologies that direct repair toward HDR or MMEJ (Yeh et al., 2019). Therefore, the processing of DSB ends from blunt ends to overhangs, *via* end-resection, is the major factor dictating repair pathway choice. Although HDR faithfully repairs lesions, the end-joining pathways are preferentially upregulated through several mechanisms following DSB formation. This is because NHEJ is active throughout all phases of the cell cycle, predominating in G0 and G1 (Shrivastav et al., 2008), whereas factors that promote extensive end-resection are more active during S and G2 phases, favouring HDR when a sister chromatid is present (Chang et al., 2017). The balance between HDR and NHEJ is further regulated by reciprocal inhibition between these two pathways. While 53BP1 and RIF1 mostly promote NHEJ by blocking end-resection, BRCA1 and CtIP direct break processing toward HDR or MMEJ (Escribano-Díaz et al., 2013).

End-Joining Repair

In the absence of a repair template, a Cas9-induced DSB is predominantly repaired in an error-prone manner, resulting in insertions and deletions (indels) within the targeted genomic sequence. If these indels give rise to frameshift mutations, they result in loss-of-function alleles. This type of repair outcome has been largely attributed to the use of NHEJ, which directly ligates the two DNA ends following cleavage, leading to the generation of small indels (<10 bp; Bothmer et al., 2017). More recently, MMEJ has been shown to contribute to a large fraction of the edited alleles observed after genome editing (Shen et al., 2018). The MMEJ-mediated repair of Cas9-induced DSBs is characterised by a distinct indel profile where larger deletions are the predominant outcome (>10 bp; Ferreira da Silva et al., 2019; **Figure 1A**). Similar to NHEJ, MMEJ ligates the DNA ends in the absence of an exogenous repair template but, unlike NHEJ, MMEJ requires initial and short-distance DSB end-resection to reveal regions of microhomology (Seol et al., 2018). The initial resection (5–25

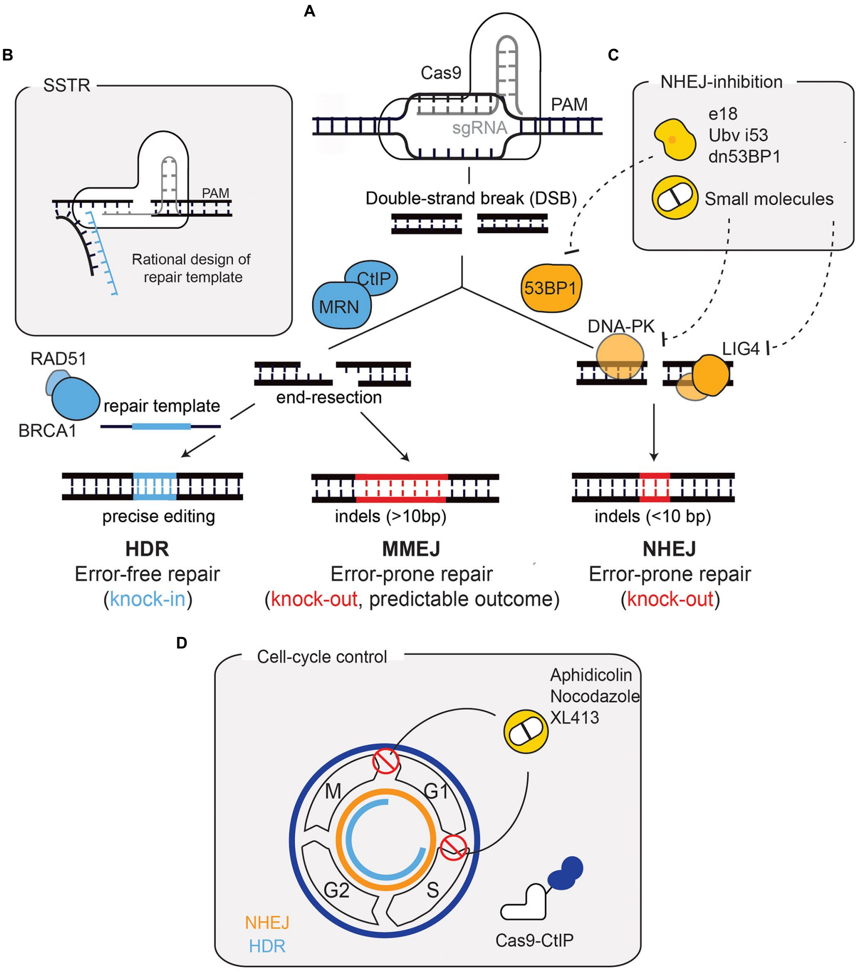


FIGURE 1 | Repair outcomes after a Cas9-induced DNA double-strand break (DSB) and strategies for enhancing precise repair. **(A)** Cas9, targeted by a sgRNA, induces a DSB in a precise region of the genome. Non-homologous end-joining (NHEJ), promoted by 53BP1, is the default repair pathway. Through the coordinated action of factors such as DNA-PK and LIG4, NHEJ repairs the DSB by re-joining the DNA-ends in an error-prone manner. This results in small insertions and deletions (indels) that can generate a loss-of-function allele if a frameshift is generated. If end-resection occurs [mediated by CtIP and MRE11-Rad50-NBS1 (MRN)], microhomology-mediated end-joining (MMEJ), or homology-directed repair (HDR) function. The repair outcome following MMEJ-mediated repair can vary, although this can be predicted since it depends on regions of microhomology and leads to larger indels. HDR, mediated by factors such as BRCA1 and RAD51, relies on a repair template and hence is error-free, leading to precise genomic alterations. **(B)** The use of ssDNA oligonucleotides (ssODN) as donor templates has also been developed to harness HDR. This process is called single-stranded templated repair (SSTR). SSTR is generally more efficient due to the asymmetry of the Cas9-DNA complex, which leads to the release of the PAM-distal non-target strand. Therefore, a rational design of the ssODN donor template complementary to the strand that is first released improves precise editing. **(C)** The inhibition of NHEJ has been used to improve precise repair following Cas9-breaks. 53BP1 inhibition through ubiquitin variants, dominant negative forms, or expression of factors that displace 53BP1, has proven useful. Small molecule inhibitors against DNA-PKcs and LIG4 have also been used. **(D)** Cell cycle manipulation has also proved useful for enhancing HDR. HDR (depicted in blue) is only active in S/G2/M phases, contrary to NHEJ (depicted in orange), which is active throughout the cell cycle. Strategies to improve HDR have included the use of compounds (such as XL413, aphidicolin, and nocodazole) to block cells in HDR-permissive phases. A Cas9-CtIP fusion allows end-resection (and subsequently HDR) to occur throughout the entire cell-cycle. PAM, protospacer adjacent-motif.

base pairs) is performed by the MRE11-Rad50-NBS1 (MRN) complex, which is activated in a cell-cycle dependent manner by CtIP (Truong et al., 2013). This exposes microhomologies on opposite strands that anneal to one another. DNA polymerase θ (POLQ) stabilises the annealed single-stranded DNA and fills the gaps, *via* templated synthesis. The early resection steps that occur in MMEJ are shared with HDR. However, annealing and extension of overhanging ends during MMEJ function to prevent HDR. Moreover, HDR requires extended end-resection, which depends on additional factors, such as the helicase Bloom syndrome protein (BLM) and Exonuclease 1 (EXO1; Truong et al., 2013).

Albeit being generally considered as an alternative pathway, studies based on the pharmacological and genetic ablation of NHEJ have shown that MMEJ can fully compensate for the absence of NHEJ in the repair of Cas9-induced DSBs (Brinkman et al., 2018; Ferreira da Silva et al., 2019). Despite the error-prone nature of end-joining pathways, there is mounting evidence indicating that the pattern of DNA repair following a Cas9-induced DSB is not stochastic (van Overbeek et al., 2016; Shou et al., 2018). Based on this observation, several studies have systematically analysed how sequences flanking the DSB impact repair outcome, leading to the important conclusion that template-free Cas9 editing can be predicted and applied to achieve a specific outcome (Allen et al., 2018; Shen et al., 2018).

Homology-Directed Repair

In contrast to the end-joining pathways, and within the context of genome editing, HDR depends on an exogenous repair template, allowing cells to integrate specific and precise alterations in their genome (**Figure 1A**), thus making it more relevant for therapeutic applications. HDR efficiency, however, remains a challenge and several approaches have been developed to overcome this limitation. Biochemical modelling of the Cas9-DNA interaction has been fundamental to prove that the efficiency of HDR can be improved through rational design of the repair template, concluding that the use of single-stranded DNA (i.e., synthetic oligonucleotides) as a repair template improves HDR (Richardson et al., 2016; Aird et al., 2018). This sub-type of HDR is commonly called single-stranded templated repair (SSTR; **Figure 1B**).

Importantly, transcriptional and genetic differences impact the efficiency of CRISPR-Cas9 editing and therefore the effectiveness of genome editing approaches. Screens performed in human cancer cell lines have shown that the Fanconi anaemia (FA) pathway diverts repair toward SSTR, playing an important role in HDR efficiency (Richardson et al., 2018). The Fanconi anaemia group D2 protein (FANCD2) has been shown to have a direct role on genome editing, by physically localising to Cas9-induced DSBs. This finding has important therapeutic implications for future genome editing applications in FA patients. Moreover, the involvement of FA, a pathway that repairs interstrand cross-links, on the repair of Cas9-mediated DSBs highlights how little is known about the interplay between DNA repair pathways in the context of different CRISPR-mediated technologies.

Rewiring DNA Double-Strand Break Repair Towards Homology-Directed Recombination

The importance of DNA repair for genome editing applications is further illustrated by the different approaches that modulate DNA repair pathways to improve HDR efficiency. For example, since NHEJ is the default pathway in human cells, its inhibition has been exploited to favour HDR. This has been achieved through the use of small-molecules targeting LIG4 or DNA-PKcs (Robert et al., 2015; Riesenberg and Maricic, 2018), ubiquitin-variants targeting 53BP1 (Canny et al., 2017), expression of factors that displace 53BP1 from DSBs (Nambiar et al., 2019), or 53BP1 dominant negative forms (Paulsen et al., 2017; **Figure 1C**). Another strategy to promote HDR is through cell cycle modulation, thereby increasing precise editing and minimising undesirable indels (**Figure 1D**). One of such strategies makes use of a Cas9 fused with the protein CtIP (Charpentier et al., 2018). This construct bypasses the requirement for cell cycle dependent activation of CtIP (by CDK1/2), necessary for end-resection and subsequent HDR. Pharmacological cell cycle arrests in HDR-permissive phases (S/G2) with aphidicolin, nocodazole, or the small molecule XL413, can also improve the efficiency of precise editing (Lin et al., 2014; Wienert et al., 2020). Overall, the modulation of DNA repair pathway choice, either through direct inhibition of NHEJ or cell-cycle regulation, comprises a potent strategy to boost precise editing.

CRISPR-Cas9 Editing Outcomes Are Shaped by DNA Repair Processes

The DNA damage response is a highly interconnected signalling network, which is modulated by cell cycle stage, gene expression changes, chromatin states, differentiation status, and cell type (Blanpain et al., 2011; Fortini et al., 2013; Klement and Goodarzi, 2014; Polak et al., 2015; Hustedt and Durocher, 2017; Weeden and Asselin-Labat, 2018; Yimit et al., 2019).

In the pursuit of safe and precise genome editing, next generation sequencing (NGS) technologies have empowered researchers to look for off-target effects beyond commonly predicted sites, enabling high standards for quality control of *ex vivo* edited cell populations (Li et al., 2019). Even in the near absence of off-target editing, the challenge of achieving precise editing outcomes at the desired target site remains. Investigating CRISPR-Cas9 outcomes in mouse embryonic stem cells, mouse hematopoietic progenitors, and differentiated human cells lines with intact DNA repair, Kosicki et al. (2018) found frequent large-scale deletions around the cut site, as well as crossover events with distant sites. Notwithstanding the advanced technologies to limit off-target effects, these surprising results revealed that more research is required to understand possible editing outcomes and how to avoid unwanted on-target effects.

A recently developed approach termed Repair-Seq was used to systematically map DNA repair outcomes, and hence editing outcomes, after Cas9 and Cas12a mediated genomic editing across several loci (Hussmann et al., 2021). This revealed that genetic dependencies driving repair outcomes are determined by the exact type of DNA lesion present. Predicting editing outcome is thus dependent on the understanding of lesion conformation and its interplay with DNA repair factors.

In summary, recent insights into the complex interplay between DNA break configuration and DNA repair factors, highlighted how the landscape of genome editing outcomes remains underexplored. The studies discussed above made their observations in a few cellular models but found a surprising variety of lesions and repair outcomes generated. The level of complexity further increases when one takes cell type and tissue specific effects of DNA repair into consideration. It becomes apparent that the full control of CRISPR-mediated genome editing is only possible with full understanding of the intricacy of endonuclease generated lesion conformation in combination with DNA repair regulation in a tissue dependent context.

SUCCESS OF CRISPR-BASED THERAPIES DEPENDS ON UNDERSTANDING TISSUE SPECIFIC DNA REPAIR

DNA Repair Outcomes Are Tissue Specific

Outside the CRISPR field, it has long been noted that the balance between the type of DNA lesion and DNA repair activity determines tissue specific repair outcome. Germline mutations in DNA repair genes cause disease phenotypes, which often manifest in a tissue specific manner. A classic example are *BRCA1/2* mutations, which cause a defect in HDR, yet predispose primarily to breast and endometrial cancers. Similarly, defects in DNA single strand break repair (SSBR), predominantly affect neuronal cell types, while, for instance defects in crosslink repair (Fanconi anaemia pathway) precipitate bone marrow failure and neurological degeneration (Tiwari and Wilson, 2019). The differential effect certain DNA repair defects have on specific cell types cannot be fully explained. Part of the explanation may be tissue specific differences in terms of which type of DNA damage is encountered, for instance, due to differential cellular metabolism or hormone levels (Langevin et al., 2011; Garaycoechea et al., 2012; Singh and Yu, 2020). However, DNA damage is only one side of the coin, while DNA repair is the other. Indeed, different cell types, even within tissues, have been found to show divergent propensity for DNA repair. Differential sensitivity to DSBs, for instance, has been observed among human hematopoietic stem cells (HSCs) and progenitor cell populations (Milyavsky et al., 2010). Compared to progenitor populations, HSCs showed delayed repair kinetics and higher levels of p53 activation, leading to increased apoptosis after DSB induction.

How the cell type affects the specificity of DNA repair outcomes across tissues is thus another level of consideration for designing CIRSPR applications. Although the intricate tissue specific response to DNA DSBs complicates design of gene editing therapies, in-depth characterization of tissue specific DNA repair mechanisms is key for developing safe and efficient therapies. We discuss recent insights which advanced the understanding of underlying mechanisms effectuating tissue specificity of DNA repair, and how this might influence CRISPR applicability.

Tissue Specific Cell Cycle Effects

Since cell cycle stage impacts repair pathway choice, only actively cycling cells have full accessibility to NHEJ, MMEJ, and HDR. Other cells, quiescent or post-mitotic, must re-enter the cell cycle to access DSB repair and other repair pathways (Nouspikel and Hanawalt, 2000; Shin et al., 2020). Upon exit of G0, NHEJ is the predominant repair pathway for DSBs, increasing the possibility of mutagenic repair (Mohrin et al., 2010; Shin et al., 2020). The inaccessibility of HDR coupled with the preference for NHEJ in some cell types poses a problem for the utility of CRISPR therapeutics. To achieve a long-lasting therapeutic effect, targeting long-lived stem cell populations offers the best strategy. However, many somatic stem cells across tissues are quiescent and therefore HDR-based therapies aimed at introducing specific edits are challenging and might limit the applicability of CRISPR technology in the clinics. A recent study, however, has demonstrated that detailed knowledge of DNA repair and cell cycle regulation can significantly increase the HDR-editability of the target cell population. Shin et al. demonstrated that quiescent HSCs can be edited with HDR up to an overall efficiency of 30% if they are stimulated to enter the cell cycle before commencing editing.

Tissue Specific Effects of Differentiation and Chromatin Status

It has been established that many different cell lineages across tissues exhibit slower rates of DNA repair and generally have reduced capacity to maintain their genome. This can be seen as an adaptive advantage, as highly differentiated cells do not spend energy on whole genome maintenance and instead focus on the conservation of actively transcribed genes (Nouspikel and Hanawalt, 2002). Most terminally differentiated cells are not of interest for CRISPR therapeutics, apart from long-lived differentiated cells such as neurons and intermittently mitotic hepatocytes. For the most part, tissue specific stem cells will be the target for clinical CRISPR applications by virtue of their ability to populate the tissue with gene-edited cells. Because DNA repair, from signalling to pathway choice, is tightly interconnected with epigenetic regulation, it must be appreciated that the distinct chromatin profiles of differentiated and non-differentiated cells might influence how a DNA lesion is repaired. HDR, in contrast to NHEJ, requires end-resection, which happens more effectively in open chromatin regions. Consequently, HDR is favoured in genomic regions with open chromatin conformation, marked by H4 acetylation and HeK36me3. NHEJ, on the other hand, is preferred in heterochromatic regions and at sites where H4 is demethylated at lysine 20 (H4K20me2; Karakaidos et al., 2020). Recently, the pathway balance between NHEJ and MMEJ as influenced by chromatin configuration has also been mapped

(Schep et al., 2021). This study showed that MMEJ is more active than NHEJ in specific heterochromatin contexts, namely late replicating regions, lamina associated regions, and at H3K9me2 sites. Moreover, MMEJ was shown to compete with SSTR (Schep et al., 2021). Therefore, systematically mapping chromatin environments across cell types can inform avenues for regulation to successfully install CRISPR edits which rely on the incorporation of repair templates.

The advances in mapping and understanding intrinsic differences in DNA repair regulation across cell types will undoubtedly promote design of more efficient CRISPR therapies, which can be applied *ex vivo* using induced pluripotent stem cells (iPSCs) and organoid-based approaches (Schwank et al., 2013; Xie et al., 2014; Li et al., 2015), while keeping unwanted on-target effects to a minimum. Especially when targeting long lived and actively dividing stem cells, *ex vivo* editing offers a safer route over *in vivo* editing, because edited cells can be thoroughly investigated and selected for the desired editing outcome, prior to transplantation into the patient. However, some diseases may require *in vivo* editing due to the plurality of tissues and cell types affected, adding another layer of complexity, since tissue context must be considered as well.

Editing Outcomes Are Influenced by Tissue Architecture

One disease in which *in vivo* editing would likely be necessary is cystic fibrosis, which is caused by mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. The function of this chloride/bicarbonate channel is to regulate the exchange of electrolytes and thus the hydration levels of secretory epithelia. Loss or reduction of function in this protein leads to cycles of mucus accumulation, inflammation, and infection in the lung, progressively destroying the airway epithelium (Ensinck et al., 2021).

With 360 reported pathogenic mutations, editing strategies for cystic fibrosis need to be tailored to each patient and draw on an integrated understanding of DNA repair. In order to achieve a long-term cure, the resident tissue stem cells, i.e., basal cells, must not only be studied in terms of their response to CRISPR-induced DNA breaks and subsequent repair, but also where they are situated within their host tissue. This is especially relevant because, within the lung, an intra-tissue variance in response to DNA damage exists. Along the airway epithelium of the trachea and larger bronchi, basal stem cells are responsible for renewing the epithelium, giving rise to ciliated and club cells (Rock et al., 2009; Asselin-Labat and Filby, 2012; Hogan et al., 2014). It should be noted that basal cells are the most active stem cell pool along the trachea, whereas in the bronchi, club cells have also been shown to self-renew and give rise to ciliated cells (Rawlins et al., 2009). Within the lung tissue, there is also the highly specialised alveolar epithelium, which consists of elongated type 1 cells and secretory type 2 cells (alveolar type 2 = AT2), the latter being the resident stem cell (Barkauskas et al., 2013; Yamamoto et al., 2020). Surprisingly, it has been observed that basal stem cells exhibit a greater capacity for repair of DSBs compared to AT2 cells. Basal cells utilise NHEJ more efficiently than AT2 cells, allowing them to resist apoptosis and to begin proliferation. In the disease context, the pathologic changes and inflammatory environment of the tissue also play a role in how efficient CRISPR editing might function. Hence, to avoid a mixture of editing outcomes across different cell types within one tissue, the utilisation of DNA repair pathways and their relative efficiency in the target cells must be taken into consideration for CRISPR-Cas9 editing.

As the CRISPR field advances, it has become ever increasingly interwoven with the DNA repair field, because it is recognised that genome editing is dependent on the activity of the cellular DNA repair machinery. We focused on CRISPR-Cas9 technologies, which depend on DSB repair pathways and reviewed the emerging research on the complexity of tissue specificity of DNA repair. The outcome of a genomic edit builds upon the complex interplay of the DNA repair machinery, which is specific to the type of lesion generated, and differs across cell types and within tissue environments, owing to cell cycle effects, differentiation status, and chromatin configurations. The power to translate genome editing to the clinic increases with a progressive understanding of all aspects of DNA repair.

CRISPR IN THE CLINICS: CHALLENGES AND LIMITATIONS DUE TO DNA REPAIR TISSUE SPECIFICITY

With ever improving CRISPR-based technologies, gene-editing treatment has become a reality in the clinics. The dream to cure diseases by correcting the causative mutations is far simpler than its implementation. For a few applications, including engineering T-cells for cancer therapy, inborn blood disorders, transthyretin (TTR) amyloidosis, and heritable blindness, CRISPR-therapies have become available to patients. We review recent achievements in clinical trials and consider the applicability of tissue specific DNA repair.

CRISPR in Cancer Therapy

Recently concluded clinical trials have successfully shown delivery of CRISPR-Cas9-based *ex vivo* therapies to patients and demonstrated safety and feasibility of these treatments. Yet, these trials have also demonstrated that the mere reduction of off-target editing is not sufficient to achieve the desired outcome. One trial (NCT02793856) studying the therapeutic effect of knocking out the programmed cell death protein 1 (PD-1) in patient derived T-cells *via* NHEJ in refractory non-small-cell lung cancer, found a good ratio of 48.7 of on-target over off-target editing. Even so, 28.8% of all on-target edits did not match the predicted outcome (Lu et al., 2020). Another trial (NCT03399448), also focused on enhancing anti-tumor immunity of T-cells, set out to simultaneously edit four loci encoding for the endogenous T-cell receptor (TCR), and PD-1, while introducing a transgene (NY-ESO-1), which is more efficient at recognising tumor cells than the TCR. While off-target editing events were rare, simultaneous editing of multiple loci led to translocations and large deletions. Of 12 possible translocation events, the most abundant rearrangement caused a 9.3 kb deletion, which was evident in all edited samples and remained detectable in patients up to 170 days post-transfusion (Stadtmauer et al., 2020). While all observed translocations persisted in peripheral blood, the frequency of detected rearrangements declined with time, indicating no specific growth advantage introduced by the unintended edits.

In summary, both trials demonstrated the utility of CRISPR-Cas9 based treatment approaches in patients, in addition to moderate clinical benefit. The editing strategy in both trials minimised off-target effects, while still introducing unwanted on-target effects. For transient cell populations such as engineered T-cells, this might be acceptable. However, for clinical applications which require precise editing of resident stem cell populations, better control over editing outcome is needed.

CRISPR for Hereditary Disease Therapy

Targeting Tissue Stem Cells

An important milestone in the development of therapeutic genome editing was reached in two CRISPR-based trials for β-thalassemia and sickle cell anemia (NCT03655678 and NCT03745287, respectively). Targeting CD45-positive hematopoietic stem and progenitor cells, the *ex vivo* editing strategy relied on error prone NHEJ to achieve gene knockout of *BCL11A*, a transcriptional repressor of foetal hemoglobin (Frangoul et al., 2021). Precise correction of the causative point mutations for these diseases seems like a more obvious choice compared to disrupting a transcription factor (**Figure 2A**). However, considering the relative ineffectiveness of HDR in the target cells and their propensity to utilize NHEJ, deliberate indel generation offers a more effective editing strategy. Both trials proved that minimising off-target effects, while carefully predicting and evaluating indels generated at the on-target

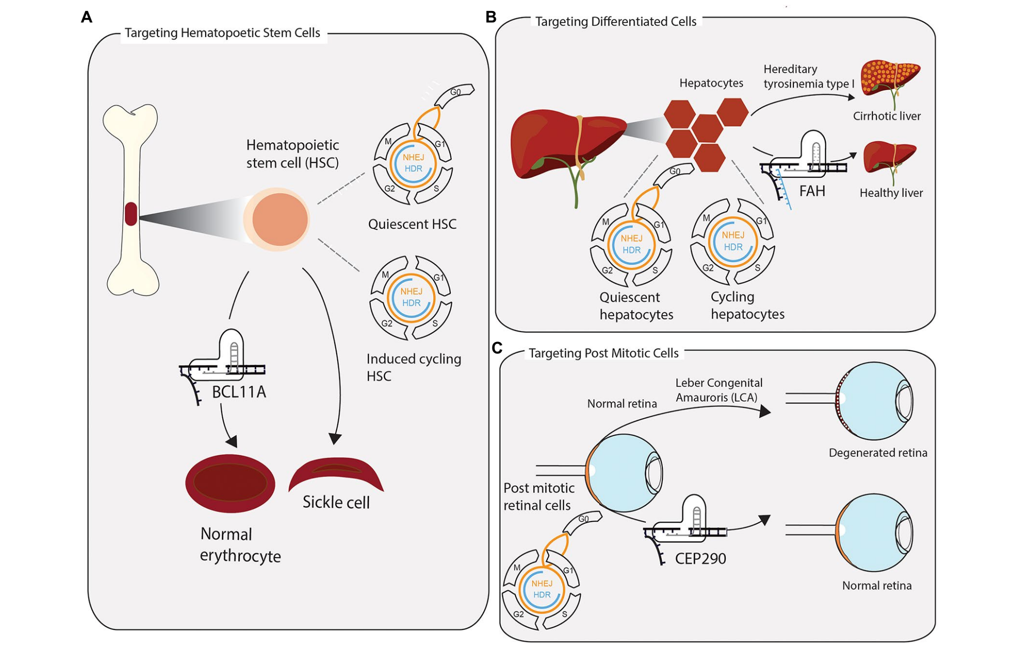


FIGURE 2 | Successful clustered regularly interspaced short palindromic repeats (CRISPR) applications require consideration of tissue-specific DNA repair and repair pathway accessibility. **(A)** Hematopoietic stem cells (HSCs) are extracted from the bone marrow and edited *ex vivo* for the treatment of sickle cell anaemia and β-thalassemia. Without stimulating cells to enter the cell cycle once before CRISPR-Cas9 editing, quiescent HSCs rely on error prone NHEJ to repair induced DSB. In a clinical application, the preference for NHEJ is leveraged to disrupt the transcription factor BCL11A, which represses the expression of foetal hemoglobin. The re-expression of foetal hemoglobin allows for the formation of normally shaped erythrocytes. **(B)** The toxic accumulation of fumarylacetoacetate in fatal hereditary tyrosinemia type I (HTI) leads to liver cirrhosis and liver failure due to a mutation in the fumarylacetoacetate hydrolase gene (FAH). Highly differentiated quiescent cells can be stimulated to re-enter the cell cycle upon DNA, or tissue, damage. Provided with a single-stranded repair template, few cycling hepatocytes have access to repair DSB *via* homologous directed repair. Precisely edited hepatocytes have a growth advantage over non-edited cells and reconstitute tissue homeostasis.
(C) Leber congenital amauroris (LCA) is the first disease treated with an *in vivo* CRISPR approach. The post mitotic light sensitive cells in the retina degenerate with age, leading to impaired vision early on in life. Appropriating the propensity of post mitotic cells to repair DSBs *via* NHEJ, the therapy aims to disrupt an aberrant splicing site in exon 26 of CEP290, maintaining a functional retina.

site, are valid strategies to utilise NHEJ for safe editing of stem cells. Edited cells engrafted in patients' bone marrow, demonstrating the feasibility of editing long lived stem cells and replenishing stem cell compartments of interest with corrected cells (Frangoul et al., 2021). In future applications, which require precise editing, controlling quiescent and cycling states of HSCs might prove useful to increase HDR (Shin et al., 2020).

Targeting Differentiated Cells

Integrated knowledge of tissue architecture and DNA repair outcomes can help designing better CRISPR therapies. A prime example of this is the fatal genetic disease hereditary tyrosinemia type I (HTI). HTI is caused by a G>A point mutation in the fumarylacetoacetate hydrolase (FAH) gene, which causes skipping of exon 8, leading to a dysfunctional protein and accumulation of the toxic metabolite fumarylacetoacetate in hepatocytes, ultimately leading to cirrhosis, acute liver failure, and increased risk of hepatocellular carcinoma (Yin et al., 2014; King et al., 2017). The liver consists largely of highly differentiated hepatocytes, while the population of hepatic progenitor cells (HPCs) is considerably smaller. Although fully differentiated, in response to disturbances to homeostasis, quiescent hepatocytes can enter the cell cycle and begin proliferating to repair tissue injury (**Figure 2B**; Kiseleva et al., 2021). In their study on HTI, Yin and colleagues demonstrated that precise correction of the mutation can be achieved in mice *via* delivering CRISPR-Cas9 along with a single-stranded DNA repair template into hepatocytes, using hydrodynamic tail vein injection. Once stimulated to proliferate, actively cycling hepatocytes can utilize HDR to make the edit of interest. Although only one in 250 liver cells were successfully edited, corrected cells have a selective advantage and begin to outgrow unedited cells and repopulate the liver, effectively ameliorating the disease. Therefore, considering tissue architecture along with DNA repair pathway choice, results in a therapy which is more effective than the initial editing efficiency.

Gene editing of hepatocytes has recently found application in a clinical trial using *in vivo* editing (Gillmore et al., 2021). TTR amyloidosis (ATTR) is a progressive fatal disease, which may be inherited in an autosomal dominant manner through inheritance of one of more than 100 recognised pathogenic mutations in the TTR protein. Misfolding of mutant TTR promotes the accumulation of insoluble protein fibers, which are deposited predominantly in heart and nervous tissue, leading to cardiomyopathies and polyneuropathies. TTR has normal, but dispensable, functions in vitamin A transport and is almost exclusively produced in the liver. Thus, targeted knockout of the TTR gene in hepatocytes, coupled with vitamin A supplementation, is a viable treatment strategy to reduce systemic levels of TTR and curb the deposition of pathogenic TTR fibers (Gertz et al., 2015).

Gillmore et al. (2021) describe the intermediate results of an ongoing clinical study seeking to reduce TTR protein level in patients with hereditary ATTR (Gillmore et al., 2021). Extensive pre-clinical screening for off-target effects was conducted to allow for the optimal selection of an efficient sgRNA and the formulation of the editing drug "NTLA-2001." The CRISPR editing machinery, encoded in mRNA, and the TTR sgRNA was delivered encapsulated in lipid nanoparticles with liver tropism. Patients showed a dose dependent effect of TTR serum level reduction after 28 days, between 47–56 and 80–96% for the lower and higher dose of NTLA-2001, respectively. Thus far, patients have not exhibited serious adverse effects. Long-term monitoring of protein level reduction, side effects, and outcomes on disease progression and mortality will show the safety and applicability of this therapy. The liver is an optimal target organ for the first *in vivo* therapy targeting differentiated cells. It consists mostly of intermittently mitotic hepatocytes, which at once reduces the risk of pathogenic outgrowth, compared to consistently cycling cells, and simplifies the complexity of having to consider many cell types in the design of the editing strategy. Aside from the rarity of hereditary ATTR, pathogenic accumulation of wild type TTR fibers in the heart is also observed in patients and has been recognised as a cause for cardiomyopathy and eventual heart failure (Gertz et al., 2015). Hence, a successful CRISPR therapy for transthyretin amyloidosis may be the first to find broad application beyond rare diseases.

Targeting Post Mitotic Cells

Since specificity of editing outcomes and safety are still major technological hurdles, there are currently few ongoing clinical trials utilising *in vivo* CRISPR Cas9 editing. One trial is seeking to treat Leber congenital amaurosis (LCA; ClinicalTrials.gov, 2019). LCA manifests in degeneration of the retina and is caused by mutations in more than 25 genes (Daich Varela et al., 2021). The CRISPR-based drug, EDIT-101, targets a heterozygous mutation in intron 26 of the LCA gene CEP290 to remove an aberrant splicing site *via* generating an indel through NHEJ (**Figure 2C**; Maeder et al., 2019). While it is exciting that *in vivo* CRISPR editing begins to move into the clinic, it is pertinent to keep in mind that LCA constitutes an ideal model disease for this approach. The post-mitotic nature of the targeted cells ensures a greater propensity for utilising NHEJ to repair the induced break and reduces the risk of selective pathogenic outgrowth of edited cells, when compared to actively cycling somatic stem cells. Furthermore, there is reduced risk of inflammation or adverse reactions to introduction of Cas9, due to the immunoprivileged status of the eye.

The examples above illustrate the potential and versatility of CRISPR-based therapies. The success of such approaches, however, relies on careful consideration about the biology of targeted cells and a deep understanding about the tissue specific mechanisms of DNA damage signalling and repair.

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The successful implementation of CRISPR-Cas9 technologies in a clinical setting relies on a deeper understanding of the DNA repair mechanisms and pathways responsible for genetic replacement outcomes, as well as the activity and accessibility of these pathways in specific cell types and tissues. Following the generation of a DSB, cell cycle regulation, and DNA repair pathway choice play major roles in determining the editing outcome. Therefore, genome editing approaches have begun to harness DNA repair control and modulation for more efficient and predictable outcomes.

Overall, the genome and transcriptome of target cells impact the effectiveness of genome editing approaches. Moreover, cell identity and tissue context are important considerations in designing effective editing strategies. While *ex vivo* editing strategies allow for extensive quality control, *in vivo* editing strategies could target multiple cell types at once, but must be safe and accurate, especially when targeting long-lived somatic stem cells. Recent successes in therapeutic editing achieved in β-thalassemia and sickle cell anemia demonstrated the feasibility of utilising CRISPR-Cas9 editing in stem cells to alleviate disease. While these reports are encouraging, there is a large margin for improving treatment strategies for diseases which require editing of multiple loci or precise editing of one locus across multiple tissues. CRISPR technologies that do not rely on the generation of DSBs, such as DNA base editors and prime editing, are promising avenues for future precision medicine. These technologies are independent of cell cycle stage and hence have the potential to correct multiple cell types. However, both base editors and prime editing introduce unique types of DNA damage products, such as DNA single-strand breaks and base mismatches, to facilitate genome editing. Hence these approaches rely on other DNA repair pathways that must be understood, in tissue-specific contexts, for further expansion and improvement of these technologies (Gu et al., 2021).

The expansion of the tools available to understand and control the CRISPR-Cas9 system has continuously fuelled the development of new therapeutic strategies and has brought a fundamental discovery into the clinics in less than a decade. The implications for personalised medicine are immense. However, for this steep trajectory to continue and to broaden the applicability and impact of these technologies, the focus of future developments must shift to include the investigation of tissue specific DNA repair. Knowledge of the underlying mechanisms of how the DNA repair machinery reacts to a CRISPR break within a distinct cellular context is a key to mapping the landscape of genome editing.

AUTHOR CONTRIBUTIONS

MM and JF wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

MM and JF are supported by DOC fellowships from the Austrian Academy of Sciences (ÖAW25757 and ÖAW25035, respectively). The Loizou lab is funded by an ERC Synergy Grant (DDREAMM grant agreement ID: 855741) and an FWF Stand Alone Grant (P33024). CeMM is funded by the Austrian Academy of Sciences.

ACKNOWLEDGMENTS

We apologise to the authors whose work we could not cite due to space limitations. We thank Amandine Moretton, Christof Gaunt, Gonçalo Oliveira, and Jana Slyskova from the Loizou lab for constructive comments on our manuscript.

REFERENCES

Aird, E. J., Lovendahl, K. N., St. Martin, A., Harris, R. S., and Gordon, W. R. (2018). Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun. Biol.* 1:54. doi: 10.1038/s42003-018-0054-2

Allen, F., Crepaldi, L., Alsinet, C., Strong, A. J., Kleshchevnikov, V., De Angeli, P., et al. (2018). Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* 37, 64–72. doi: 10.1038/nbt.4317

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157. doi: 10.1038/s41586-019-1711-4

Asselin-Labat, M. L., and Filby, C. E. (2012). Adult lung stem cells and their contribution to lung tumourigenesis. *Open Biol.* 2:120094. doi: 10.1098/rsob.120094

Barkauskas, C. E., Cronce, M. J., Rackley, C. R., Bowie, E. J., Keene, D. R., Stripp, B. R., et al. (2013). Type 2 alveolar cells are stem cells in adult lung. *J. Clin. Invest.* 123, 3025–3036. doi: 10.1172/JCI68782

Blanpain, C., Mohrin, M., Sotiropoulou, P. A., and Passegué, E. (2011). DNA-damage response in tissue-specific and cancer stem cells. *Cell Stem Cell* 8, 16–29. doi: 10.1016/j.stem.2010.12.012

Bothmer, A., Phadke, T., Barrera, L. A., Margulies, C. M., Lee, C. S., Buquicchio, F., et al. (2017). Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat. Commun.* 8:13905. doi: 10.1038/ncomms13905

Brinkman, E. K., Chen, T., de Haas, M., Holland, H. A., Akhtar, W., and van Steensel, B. (2018). Kinetics and fidelity of the repair of Cas9-induced double-strand DNA breaks. *Mol. Cell* 70, 801.e6–813.e6. doi: 10.1016/j.molcel.2018.04.016

Canny, M. D., Moatti, N., Wan, L. C. K., Fradet-Turcotte, A., Krasner, D., Mateos-Gomez, P. A., et al. (2017). Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR–Cas9 genome-editing efficiency. *Nat. Biotechnol.* 36, 95–102. doi: 10.1038/nbt.4021

Chang, H. H. Y., Pannunzio, N. R., Adachi, N., and Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* 18, 495–506. doi: 10.1038/nrm.2017.48

Charpentier, M., Khedher, A. H. Y., Menoret, S., Brion, A., Lamribet, K., Dardillac, E., et al. (2018). CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat. Commun.* 9:1133. doi: 10.1038/s41467-018-03475-7

ClinicalTrials.gov (2019). Single ascending dose study in participants with LCA10 – full text view – ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT03872479?term=NCT03872479&draw=2&rank=1 (Accessed March 26, 2021).

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143

Daich Varela, M., Cabral De Guimaraes, T. A., Georgiou, M., and Michaelides, M. (2021). Leber congenital amaurosis/early-onset severe retinal dystrophy:

current management and clinical trials. *Br. J. Ophthalmol.* doi: 10.1136/bjophthalmol-2020-318483 [Epub ahead of print]

Ensinck, M., Mottais, A., Detry, C., Leal, T., and Carlon, M. S. (2021). On the corner of models and cure: gene editing in cystic fibrosis. *Front. Pharmacol.* 12:662110. doi: 10.3389/fphar.2021.662110

Escribano-Díaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J. T. F., Tkáč, J., et al. (2013). A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol. Cell* 49, 872–883. doi: 10.1016/j.molcel.2013.01.001

Ferreira da Silva, J., Salic, S., Wiedner, M., Datlinger, P., Essletzbichler, P., Hanzl, A., et al. (2019). Genome-scale CRISPR screens are efficient in non-homologous end-joining deficient cells. *Sci. Rep.* 9:15751. doi: 10.1038/s41598-019-52078-9

Fortini, P., Ferretti, C., and Dogliotti, E. (2013). The response to DNA damage during differentiation: pathways and consequences. *Mutat. Res.* 743–744, 160–168. doi: 10.1016/j.mrfmmm.2013.03.004

Frangoul, H., Altshuler, D., Cappellini, M. D., Chen, Y.-S., Domm, J., Eustace, B. K., et al. (2021). CRISPR-Cas9 gene editing for sickle cell disease and β-thalassemia. *N. Engl. J. Med.* 384, 252–260. doi: 10.1056/NEJMoa2031054

Garaycoechea, J. I., Crossan, G. P., Langevin, F., Daly, M., Arends, M. J., and Patel, K. J. (2012). Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature* 489, 571–575. doi: 10.1038/nature11368

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., et al. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551, 464–471. doi: 10.1038/nature24644

Gertz, M. A., Benson, M. D., Dyck, P. J., Grogan, M., Coelho, T., Cruz, M., et al. (2015). Diagnosis, prognosis, and therapy of transthyretin amyloidosis. *J. Am. Coll. Cardiol.* 66, 2451–2466. doi: 10.1016/J.JACC.2015.09.075

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154:442. doi: 10.1016/j.cell.2013.06.044

Gillmore, J. D., Gane, E., Taubel, J., Kao, J., Fontana, M., Maitland, M. L., et al. (2021). CRISPR-Cas9 in vivo gene editing for transthyretin amyloidosis. *N. Engl. J. Med.* 385, 493–502. doi: 10.1056/NEJMoa2107454

Gu, S., Bodai, Z., Cowan, Q. T., and Komor, A. C. (2021). Base editors: expanding the types of DNA damage products harnessed for genome editing. *Gene Genome Ed.* 1:100005. doi: 10.1016/J.GGEDIT.2021.100005

Hogan, B. L. M., Barkauskas, C. E., Chapman, H. A., Epstein, J. A., Jain, R., Hsia, C. C. W., et al. (2014). Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell* 15, 123–138. doi: 10.1016/j.stem.2014.07.012

Hossain, M. A., Barrow, J. J., Shen, Y., Haq, M. I., and Bungert, J. (2015). Artificial zinc finger DNA binding domains: versatile tools for genome engineering and modulation of gene expression. *J. Cell. Biochem.* 116, 2435–2444. doi: 10.1002/jcb.25226

Hussmann, J. A., Ling, J., Ravisankar, P., Yan, J., Cirincione, A., Xu, A., et al. (2021). Mapping the genetic landscape of DNA double-strand break repair. BioRxiv [Preprint]. doi:10.1101/2021.06.14.448344

Hustedt, N., and Durocher, D. (2017). The control of DNA repair by the cell cycle. *Nat. Cell Biol.* 19, 1–9. doi: 10.1038/ncb3452

Jackson, S. P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature* 461, 1071–1078. doi: 10.1038/nature08467

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. *elife* 2:e00471. doi: 10.7554/eLife.00471

Karakaidos, P., Karagiannis, D., and Rampias, T. (2020). Resolving DNA damage: epigenetic regulation of DNA repair. *Molecules* 25:2496. doi: 10.3390/molecules25112496

King, L. S., Trahms, C., and Scott, C. R. (2017). "Tyrosinemia type I," in *Encycl. Mol. Mech. Dis.* 2132–2133. Available at: https://www.ncbi.nlm.nih.gov/books/NBK1515/ (Accessed August 16, 2021).

Kiseleva, Y. V., Antonyan, S. Z., Zharikova, T. S., Zharikov, Y. O., Tupikin, K. A., and Kalinin, D. V. (2021). Molecular pathways of liver regeneration: a comprehensive review. *World J. Hepatol.* 13, 270–290. doi: 10.4254/wjh.v13.i3.270

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., et al. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490–495. doi: 10.1038/nature16526

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., et al. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523, 481–485. doi: 10.1038/nature14592

Klement, K., and Goodarzi, A. A. (2014). DNA double strand break responses and chromatin alterations within the aging cell. *Exp. Cell Res.* 329, 42–52. doi: 10.1016/j.yexcr.2014.09.003

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424. doi: 10.1038/nature17946

Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* 36, 765–771. doi: 10.1038/nbt.4192

Langevin, F., Crossan, G. P., Rosado, I. V., Arends, M. J., and Patel, K. J. (2011). Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature* 475, 53–58. doi: 10.1038/nature10192

Li, H. L., Fujimoto, N., Sasakawa, N., Shirai, S., Ohkame, T., Sakuma, T., et al. (2015). Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. *Stem Cell Rep.* 4, 143–154. doi: 10.1016/j.stemcr.2014.10.013

Li, J., Hong, S., Chen, W., Zuo, E., and Yang, H. (2019). Advances in detecting and reducing off-target effects generated by CRISPR-mediated genome editing. *J. Genet. Genomics* 46, 513–521. doi: 10.1016/j.jgg.2019.11.002

Lin, S., Staahl, B. T., Alla, R. K., and Doudna, J. A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *elife* 3:e04766. doi: 10.7554/eLife.04766

Lu, Y., Xue, J., Deng, T., Zhou, X., Yu, K., Deng, L., et al. (2020). Safety and feasibility of CRISPR-edited T cells in patients with refractory non-small-cell lung cancer. *Nat. Med.* 26, 732–740. doi: 10.1038/s41591-020-0840-5

Maeder, M. L., Stefanidakis, M., Wilson, C. J., Baral, R., Barrera, L. A., Bounoutas, G. S., et al. (2019). Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat. Med.* 25, 229–233. doi: 10.1038/s41591-018-0327-9

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826. doi: 10.1126/science.1232033

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., et al. (2011). A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* 29, 143–150. doi: 10.1038/nbt.1755

Milyavsky, M., Gan, O. I., Trottier, M., Komosa, M., Tabach, O., Notta, F., et al. (2010). A distinctive DNA damage response in human hematopoietic stem cells reveals an apoptosis-independent role for p53 in self-renewal. *Cell Stem Cell* 7, 186–197. doi: 10.1016/j.stem.2010.05.016

Mohrin, M., Bourke, E., Alexander, D., Warr, M. R., Barry-Holson, K., Le Beau, M. M., et al. (2010). Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis. *Cell Stem Cell* 7, 174–185. doi: 10.1016/j.stem.2010.06.014

Nambiar, T. S., Billon, P., Diedenhofen, G., Hayward, S. B., Taglialatela, A., Cai, K., et al. (2019). Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant. *Nat. Commun.* 10:3395. doi: 10.1038/s41467-019-11105-z

Nouspikel, T., and Hanawalt, P. C. (2000). Terminally differentiated human neurons repair transcribed genes but display attenuated global DNA repair and modulation of repair gene expression. *Mol. Cell. Biol.* 20:1562. doi: 10.1128/MCB.20.5.1562-1570.2000

Nouspikel, T., and Hanawalt, P. C. (2002). DNA repair in terminally differentiated cells. *DNA Repair* 1, 59–75. doi: 10.1016/S1568-7864(01)00005-2

Nuñez, J. K., Chen, J., Pommier, G. C., Cogan, J. Z., Replogle, J. M., Adriaens, C., et al. (2021). Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* 184, 2503.e17–2519.e17. doi: 10.1016/j.cell.2021.03.025

Paulsen, B. S., Mandal, P. K., Frock, R. L., Boyraz, B., Yadav, R., Upadhyayula, S., et al. (2017). Ectopic expression of RAD52 and dn53BP1 improves homology-directed repair during CRISPR-Cas9 genome editing. *Nat. Biomed. Eng.* 1, 878–888. doi: 10.1038/s41551-017-0145-2

Polak, P., Karlic, R., Koren, A., Thurman, R., Sandstrom, R., Lawrence, M. S., et al. (2015). Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* 518, 360–364. doi: 10.1038/nature14221

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., et al. (2013). Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183. doi: 10.1016/j.cell.2013.02.022

Rawlins, E. L., Okubo, T., Xue, Y., Brass, D. M., Auten, R. L., Hasegawa, H., et al. (2009). The role of Scgb1a1+ clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell Stem Cell* 4, 525–534. doi: 10.1016/j.stem.2009.04.002

Richardson, C. D., Kazane, K. R., Feng, S. J., Zelin, E., Bray, N. L., Schäfer, A. J., et al. (2018). CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat. Genet.* 50, 1132–1139. doi: 10.1038/s41588-018-0174-0

Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L., and Corn, J. E. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* 34, 339–344. doi: 10.1038/nbt.3481

Riesenberg, S., and Maricic, T. (2018). Targeting repair pathways with small molecules increases precise genome editing in pluripotent stem cells. *Nat. Commun.* 9:2164. doi: 10.1038/s41467-018-04609-7

Robert, F., Barbeau, M., Éthier, S., Dostie, J., and Pelletier, J. (2015). Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome Med.* 7:93. doi: 10.1186/s13073-015-0215-6

Rock, J. R., Onaitis, M. W., Rawlins, E. L., Lu, Y., Clark, C. P., Xue, Y., et al. (2009). Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12771–12775. doi: 10.1073/pnas.0906850106

Schep, R., Brinkman, E. K., Leemans, C., Vergara, X., van der Weide, R. H., Morris, B., et al. (2021). Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Mol. Cell* 81:2216. e10–2230.e10. doi: 10.1016/j.molcel.2021.03.032

Schwank, G., Koo, B. K., Sasselli, V., Dekkers, J. F., Heo, I., Demircan, T., et al. (2013). Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* 13, 653–658. doi: 10.1016/j.stem.2013.11.002

Seol, J. H., Shim, E. Y., and Lee, S. E. (2018). Microhomology-mediated end joining: good, bad and ugly. *Mutat. Res. Fundam. Mol. Mech. Mutagen.* 809, 81–87. doi: 10.1016/j.mrfmmm.2017.07.002

Shen, M. W., Arbab, M., Hsu, J. Y., Worstell, D., Culbertson, S. J., Krabbe, O., et al. (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* 563, 646–651. doi: 10.1038/s41586-018-0686-x

Shin, J. J., Schröder, M. S., Caiado, F., Wyman, S. K., Bray, N. L., Bordi, M., et al. (2020). Controlled cycling and quiescence enables efficient HDR in engraftment-enriched adult hematopoietic stem and progenitor cells. *Cell Rep.* 32:108093. doi: 10.1016/j.celrep.2020.108093

Shou, J., Li, J., Liu, Y., and Wu, Q. (2018). Precise and predictable CRISPR chromosomal rearrangements reveal principles of Cas9-mediated nucleotide insertion. *Mol. Cell* 71, 498.e4–509.e4. doi: 10.1016/j.molcel.2018.06.021

Shrivastav, M., De Haro, L. P., and Nickoloff, J. A. (2008). Regulation of DNA double-strand break repair pathway choice. *Cell Res.* 18, 134–147. doi: 10.1038/cr.2007.111

Singh, A. K., and Yu, X. (2020). Tissue-specific carcinogens as soil to seed BRCA1/2-mutant hereditary cancers. *Trends Cancer* 6, 559–568. doi: 10.1016/j.trecan.2020.03.004

Stadtmauer, E. A., Fraietta, J. A., Davis, M. M., Cohen, A. D., Weber, K. L., Lancaster, E., et al. (2020). CRISPR-engineered T cells in patients with refractory cancer. *Science* 367:eaba7365. doi: 10.1126/science.aba7365

Tiwari, V., and Wilson, D. M. III (2019). DNA damage and associated DNA repair defects in disease and premature aging. *Am. J. Hum. Genet.* 105:237. doi: 10.1016/j.ajhg.2019.06.005

Truong, L. N., Li, Y., Shi, L. Z., Hwang, P. Y. H., He, J., Wang, H., et al. (2013). Microhomology-mediated end joining and homologous recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 110, 7720–7725. doi: 10.1073/pnas.1213431110

van Overbeek, M., Capurso, D., Carter, M. M., Thompson, M. S., Frias, E., Russ, C., et al. (2016). DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Mol. Cell* 63, 633–646. doi: 10.1016/j.molcel.2016.06.037

Walton, R. T., Christie, K. A., Whittaker, M. N., and Kleinstiver, B. P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* 368, 290–296. doi: 10.1126/science.aba8853

Wang, D., Zhang, F., and Gao, G. (2020). CRISPR-based therapeutic genome editing: strategies and in vivo delivery by AAV vectors. *Cell* 181, 136–150. doi: 10.1016/j.cell.2020.03.023

Weeden, C. E., and Asselin-Labat, M. L. (2018). Mechanisms of DNA damage repair in adult stem cells and implications for cancer formation. *Biochim. Biophys. Acta Mol. basis Dis.* 1864, 89–101. doi: 10.1016/j.bbadis.2017.10.015

Wienert, B., Nguyen, D. N., Guenther, A., Feng, S. J., Locke, M. N., Wyman, S. K., et al. (2020). Timed inhibition of CDC7 increases CRISPR-Cas9 mediated templated repair. *Nat. Commun.* 11:2109. doi: 10.1038/s41467-020-15845-1

Xie, F., Ye, L., Chang, J. C., Beyer, A. I., Wang, J., Muench, M. O., et al. (2014). Seamless gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome Res.* 24, 1526–1533. doi: 10.1101/gr.173427.114

Yamamoto, Y., Korogi, Y., Hirai, T., and Gotoh, S. (2020). A method of generating alveolar organoids using human pluripotent stem cells. *Methods Cell Biol.* 159, 115–141. doi: 10.1016/bs.mcb.2020.02.004

Yeh, C. D., Richardson, C. D., and Corn, J. E. (2019). Advances in genome editing through control of DNA repair pathways. *Nat. Cell Biol.* 21, 1468–1478. doi: 10.1038/s41556-019-0425-z

Yimit, A., Adebali, O., Sancar, A., and Jiang, Y. (2019). Differential damage and repair of DNA-adducts induced by anti-cancer drug cisplatin across mouse organs. *Nat. Commun.* 10:309. doi: 10.1038/s41467-019-08290-2

Yin, H., Xue, W., Chen, S., Bogorad, R. L., Benedetti, E., Grompe, M., et al. (2014). Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat. Biotechnol.* 32, 551–553. doi: 10.1038/nbt.2884

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771. doi: 10.1016/j.cell.2015.09.038

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Forum

Tissue specificity of DNA repair: the CRISPR compass

Joana Ferreira da Silva,[1,2,3] Mathilde Meyenberg,[1,2,3] and Joanna I. Loizou[1,2,*]

CRISPR-Cas9-mediated genome editing holds great promise for the correction of pathogenic variants in humans. However, its therapeutic implementation is hampered due to unwanted editing outcomes. A better understanding of cell type- and tissue-specific DNA repair processes will ultimately enable precise control of editing outcomes for safer and effective therapies.

Introduction

Clustered regularly interspaced short palindromic repeats (CRISPR) and the associated Cas9 endonuclease (CRISPR-Cas9) are a powerful genome editing tool [1]. Guided by a single-guide RNA (sgRNA), CRISPR-Cas9 targets a specific region of the genome, generating a DNA double-strand break (DSB) that activates the cellular DNA repair machinery. The DNA repair signaling network is dependent on cell cycle stage as well as other cell intrinsic factors, such as signaling networks, gene expression, 3D genome organization, and cell identity [2,3]. The highly specialized nature of different cells across the body results in differential responses to DNA damage and, ultimately, DNA repair outcomes, which in turn determine whether a lesion is precisely repaired or produces a mutation [4]. While **non-homologous end-joining (NHEJ)** (see Glossary) and **microhomology-mediated end-joining (MMEJ)** are predominantly error-prone pathways, leading to gene disruption, **homology-directed repair (HDR)** relies on a repair template, allowing precise alterations to the genome. Since DNA repair is tissue specific, this has far-reaching implications for CRISPR-Cas9-based applications and understanding this is paramount to the success of CRISPR-based therapeutics.

Cell type matters

Which DNA repair pathways are utilized upon CRISPR-Cas9-mediated DSBs depends heavily on tissue context and whether a cell is terminally differentiated, quiescent, or actively cycling. Fully differentiated, long-lived postmitotic cells, such as those found in neuronal tissues, reside in G0, making HDR inaccessible to repair lesions. Instead, postmitotic neurons respond to DNA DSBs by briefly re-entering the cell cycle, upon which NHEJ is the pathway of choice for DSB repair [5] (Figure 1A). Hence, postmitotic cells are difficult to target for precise genome editing by HDR. Instead, homology-independent repair may be exploited for therapeutic applications [6] (Box 1).

Quiescent cells exist in a state of metabolic downregulation and do not actively divide, but can become activated upon DNA damage and tissue injury to reinstate tissue homeostasis [7] (Figure 1B). Upon cell cycle re-entry, NHEJ is the predominant pathway for repairing DSBs, which limits possible editing applications and bears the risk of introducing mutations [8]. With adult tissue resident stem cells, incurred mutations will then be passed on to progenitor cells and fixed within the stem cell pool, thus increasing the risk for future malignancies. This is of particular concern for quiescent hematopoietic stem cells (HSCs), which, upon incurring DSBs, exhibit shorter p53 activation, utilize NHEJ exclusively [9], and resist apoptosis, at the cost of long-term genome stability [2,10]. Likewise, bulge stem cells of hair follicles escape elimination through upregulation of the antiapoptotic protein BCL-2 [2].

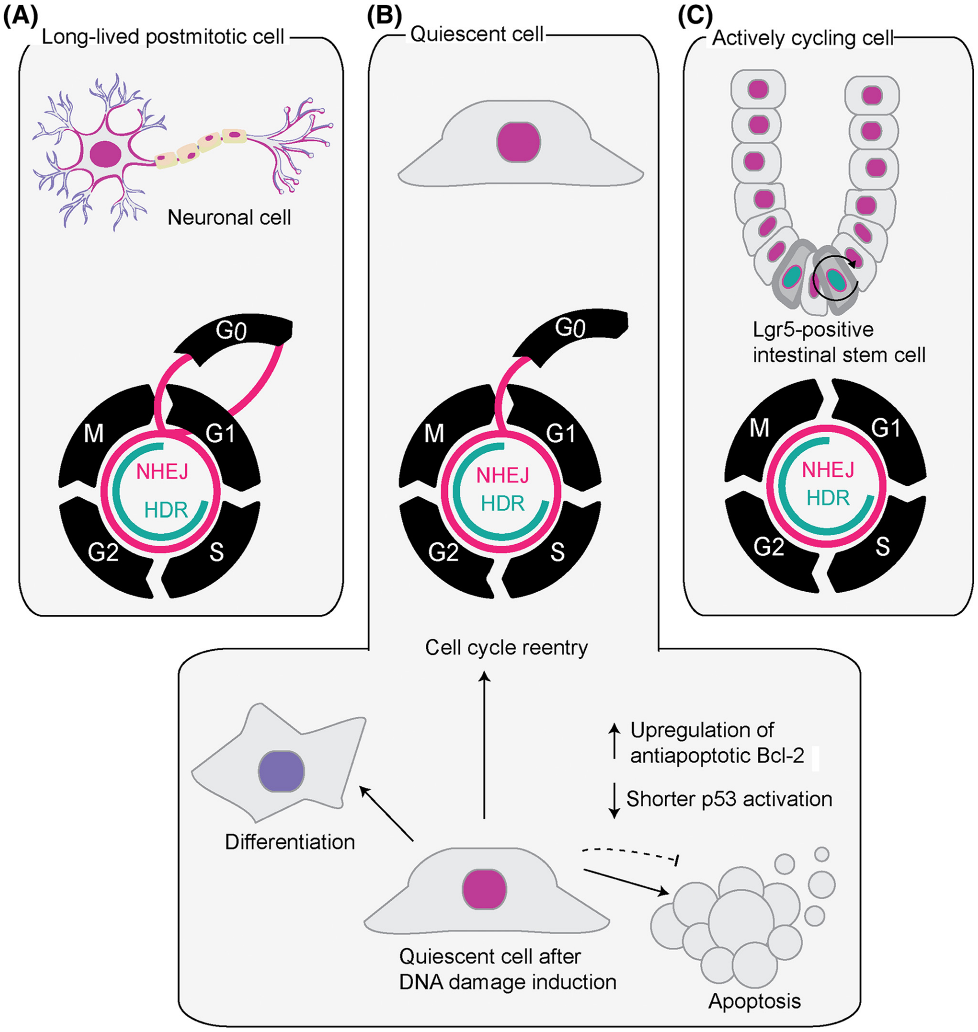
Glossary

Homology-directed repair (HDR): a pathway that accurately repairs DNA double-strand breaks, by utilizing either double- or single-stranded homologous DNA as a template.
Interhomolog repair (IHR): a subtype of HDR, where a homologous chromosome is used as template.
Loss-of-heterozygosity (LOH): a cross chromosomal event that results in loss of the entire gene and the surrounding chromosomal region.
Microhomology-mediated end-joining (MMEJ): a pathway that repairs DNA double-strand breaks through the use sequences with microhomologies, during the alignment of broken ends.
Non-homologous end-joining (NHEJ): a DNA double-strand break repair pathway that, contrary to HDR, does not utilize a homologous DNA template. NHEJ directly joins the broken ends after minimal processing.

On the contrary, other tissue stem cells, such as melanocyte stem cells, which give rise to mature pigmented melanocytes, are predominantly pushed into terminal differentiation upon DNA damage, leading to depletion of the stem cell pool [2] (Figure 1B). Overall, the differential response to DNA damage requires careful consideration for applicability of CRISPR-therapeutics.

Tissue context matters

Designing safe CRISPR therapies is dependent on understanding biology of individual stem cell pools and their context within tissues. Somatic stem cells exist in specialized niches and their response to DNA damage and tissue injury frequently depends on back-up stem cells existent within the same tissue [7]. For instance, in the intestine, aside from the main stem cell pool of actively cycling LGR5-positive cells (Figure 1C), quiescent BMI1-positive stem cells also exist. Unlike BMI1-positive cells, LGR5-positive cells are more sensitive to DNA damage and undergo apoptosis after encountering DSBs [11]. Thus, BMI1-positive cells can become activated to reconstitute LGR5-positive stem cells and the intestinal epithelium [11]. Predicting the repair outcome in response to DNA damage *in vivo* is thus further complicated



Trends in Genetics

Figure 1. Differential responses to DNA double-strand breaks of actively cycling, quiescent, and long-lived postmitotic cells. (A) Long-lived postmitotic cells, such as neuronal cells, cannot repair DNA double-strand breaks (DSBs) without briefly re-entering the cell cycle before returning to G0 and are thus restricted to utilize non-homologous end-joining (NHEJ). (B) Quiescent cells exist in a state of metabolic downregulation and reside within tissues at cell cycle stage G0. Upon tissue damage, these cells become activated and begin proliferating to mend tissue injury. The response to DNA damage varies among cell types and tissue of origin. If quiescent cells re-enter the cell cycle, NHEJ is the predominant repair pathway for DSBs. Hematopoietic stem cells and hair bulge stem cells actively resist apoptosis at the cost of long-term genomic stability via shortened duration of p53 activation and upregulation of antiapoptotic protein BCL-2, respectively. Terminal differentiation in response to DNA damage is an alternative to cell cycle re-entry but leads to depletion of the stem cell pool, as with melanocyte stem cells. (C) Actively cycling cells, such as intestinal LGR5-positive stem cells, may utilize NHEJ (depicted in pink) or homology-directed repair (HDR) (depicted in green), depending on cell cycle stage.

Box 1. Overcoming editing challenges

The genome editing field has recognized the challenges imposed by DSB-mediated editing and responded by developing mitigation strategies aimed at overcoming these challenges. Even though these strategies are currently not as widely adopted as Cas9-mediated CRISPR editing, they hold great promise for clinical applications of genome editing.

(i) DSB-independent strategies

The activation of signaling networks following a DSB has raised concerns that potentially limit the applicability of CRISPR technologies in the clinics. To this end, DSB-independent strategies, such as base editors [22,23] and prime editing [24], have been developed. Moreover, applications that modulate gene expression at the transcription level have also been successfully developed [25–27]. It is relevant to mention that even though these strategies do not rely on the generation of a DSB, they often create a DNA single-strand break at the DNA target site, which activates other signaling pathways.

(ii) Temporal delivery of CRISPR components

Cell cycle modulation is a strategy that can be used to promote precise editing, whilst minimizing undesirable indels. Timing the delivery of CRISPR components to HDR-permissive phases (S/G2) has been shown to improve the efficiency of precise editing. This has been achieved, for example, through the use of aphidicolin, nocodazole, or the small molecule XL413 [28,29]. Cas9 fusions have also been used to overcome the cell cycle limitations of HDR. One such strategy makes use of a Cas9 fused with the protein CtIP, bypassing the requirement for cell cycle-dependent activation of CtIP, which is necessary for HDR [30]. A further strategy fused Cas9 with an anti-CRISPR protein, a natural inhibitor of CRISPR-Cas systems, that is specifically degraded in HDR-permissive cell cycle phases [31].

In the context of quiescent stem cells, the development of a strategy to transiently exit quiescence, making HDR accessible during CRISPR-Cas9 editing before re-establishing quiescence, has expanded the applicability of genome editing in HSCs [9].

(iii) Homology-independent repair

The difficulty inherent to using HDR in postmitotic cells has prompted the development of homology-independent targeted integration strategies [32]. These strategies often rely on end-joining pathways for precise editing, making them sensitive to indels. This problem has been overcome by the development of other methods that generate DSBs in noncoding regions flanking the region of interest [6].

by the activation of quiescent backup stem cells, which upon cell cycle re-entry will also favor NHEJ to repair DNA breaks, possibly leading to unintended editing outcomes across different stem cell pools, including mosaicism and unintended events at the target site. Regarding the latter, while it has been demonstrated that off-target effects can be controlled in editing intestinal stem cells [12], recent studies suggest that repair pathway choice can lead to unintended on-target effects. These effects, observed in human induced pluripotent cell lines (iPSCs), include large deletions and **loss-of-heterozygosity (LOH)** around the target site in up to 40% of HDR-edited and 50% of NHEJ-edited iPSCs [13]. Undesired editing outcomes at the target site depend on the repair pathway choice and therefore might differ among different stem cell compartments, such as found in the intestine, possibly leading to mosaicism, further complicating the prediction of editing outcomes for *in vivo* applications.

In summary, even though DNA repair pathways are often portrayed as linear, they operate within a dense signaling network, influenced by cell type and state. Designing CRISPR therapeutics for adult stem cells is thus challenging for three reasons. Firstly, genome editing strategies must overcome the inaccessibility of HDR in the quiescent state and consider the propensity of NHEJ in these cells. Secondly, editing outcomes must be precise and predictable to ensure long-term genome stability. Finally, it is important to understand stem cell function and DNA repair within tissue context to target the most appropriate stem cell population while leveraging its unique response to DNA damage to achieve the desired editing outcome.

Germline editing

The precision and efficiency inherent to the CRISPR system raise the possibility of human germline editing. However, the informed discussion on the safety of these approaches relies on understanding the fundamental cellular processes that occur during embryogenesis, such as cell cycle control and DNA damage repair, as highlighted by several lines of research.

A study aimed at correcting a heterozygous mutation in the causative gene for hypertrophic cardiomyopathy (*MYBPC3*), in human preimplantation embryos, reported high levels of CRISPR-Cas9 correction [14]. The mechanism that led to correctly edited embryos in this study was suggested not to be attributed to HDR, through the exogenously provided repair template, but rather to the utilization of the wild type maternal allele as a homolog for repair. This conclusion has, however, been largely disputed because there may be alternative explanations, including the generation of large deletions and rearrangements that were undetected [15–17]. One of the main arguments against **interhomolog repair (IHR)** is the fact that maternal and paternal genomes undergo distinct developmental programs and localize in separate nuclei before the first mitotic cell division [15]. In contrast to this observation, it is worth noting that a consistent overlap between maternal and paternal genomes has been described following nuclear envelope breakdown [18].

In early mouse embryos, IHR has been shown to be a common DSB repair mechanism, enhanced by RAD51 activity [18]. In line with this observation, a study that induced DSBs within a mutant allele in heterozygous preimplantation human embryos showed that IHR and NHEJ are the two main repair mechanisms involved, while HDR using the exogenous DNA template was highly inefficient [19]. The authors showed that NHEJ and IHR often compete within the same cell, frequently leading to embryos carrying identical indel mutations on both loci. Additionally, repair from the maternal chromosome by IHR led to extensive LOH. While IHR could be applicable for gene correction, the high incidence of NHEJ as well as the extensive LOH are safety concerns.

Other recent studies have reported unexpected repair outcomes of CRISPR-Cas9-mediated editing in human embryos. Zuccaro *et al.* edited a blindness-causing frameshift mutation in the *EYS* locus in the paternal chromosome [20]. Here, the vast majority of repair outcomes were small deletions, consistent with end-joining mediated repair of the Cas9-induced DSBs. Contrary to the studies described earlier [14,18,19], Zuccaro and colleagues showed no evidence for IHR [20]. By analyzing genome-wide chromosome content, the authors report that the LOH observed following CRISPR-Cas9 mediated editing is due to the loss of the paternal chromosome arm, or even the entire chromosome [20]. An independent study, targeting the gene *OCT4*, reached similar conclusions [21].

The discrepancies between these studies could be due to locus-specific differences as, for example, proximity to telomeres considerably increases the possibility of chromosome arm truncations [21]. Nonetheless, to resolve the issues around IHR in human embryos, an evaluation of the extent of on-target mutagenesis is necessary. This would require the development of an approach to enrich for the region of interest, followed by deep, long-read sequencing [21].

Discussions around human germline editing have raised important questions that affect the future direction of the CRISPR-Cas9 technology, but also ethical and societal issues that accompany such applications. The latest studies using CRISPR-Cas9 to edit embryos have challenged our understanding of the DNA repair mechanisms that occur in these cells and illustrate the importance of using human embryos to study DNA repair outcomes [18–21]. Unexpected editing outcomes, including the loss of the targeted chromosome, highlight how important it is to fundamentally understand these mechanisms to enable precise genome modification.

Therapeutic genome editing has the potential to have a profound impact on patients. However, the successful implementation of CRISPR-Cas9 technologies in clinical settings relies on a deeper understanding of the DNA repair mechanisms and pathways responsible for genetic replacement outcomes, as well as the activity of these pathways in specific cell types and tissues. Moreover, an in-depth investigation of outcomes at the target site in different tissue contexts, using relevant models, such as human embryos, is warranted. This combined knowledge is critical for genome editing implementation and may limit or determine the application of specific genome-editing tools.

Acknowledgments

We apologize to the authors whose work we could not cite due to space limitations. We thank Drs Amandine Moretton, Christof Gaunt, Gonçalo Oliveira, and Jana Slyskova from the Loizou lab for constructive comments on our manuscript. J.F.daS. and M.M. are supported by DOC fellowships from the Austrian Academy of Sciences (ÖAW25035 and ÖAW25757, respectively). The Loizou lab is funded by an ERC Synergy Grant (DDREAMM Grant agreement ID 855741). CeMM is funded by the Austrian Academy of Sciences.

Declaration of interests

The authors declare no commercial or financial relationships that could be construed as a potential conflict of interest.

[1] CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria

[2] Institute of Cancer Research, Department of Medicine I, Comprehensive Cancer Center, Medical University of Vienna, 1090 Vienna, Austria

[3] These authors contributed equally to this work

*Correspondence:
joanna.loizou@meduniwien.ac.at (J.I. Loizou).

https://doi.org/10.1016/j.tig.2021.07.010

References

1. Jinek, M. *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821
2. Blanpain, C. *et al.* (2011) DNA-damage response in tissue-specific and cancer stem cells. *Cell Stem Cell* 8, 16–29
3. Polak, P. *et al.* (2015) Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* 518, 360–364
4. Blokzijl, F. *et al.* (2016) Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* 538, 260–264
5. Fortini, P. *et al.* (2013) The response to DNA damage during differentiation: pathways and consequences. *Mutat. Res. Fundam. Mol. Mech. Mutagen.* 743–744, 160–168
6. Fang, H. *et al.* (2021) An optimized CRISPR/Cas9 approach for precise genome editing in neurons. *eLife* 10, e65202
7. Fuchs, E. and Blau, H.M. (2020) Tissue stem cells: architects of their niches. *Cell Stem Cell* 27, 532–556
8. Beerman, I. *et al.* (2014) Quiescent hematopoietic stem cells accumulate DNA damage during aging that is repaired upon entry into cell cycle. *Cell Stem Cell* 15, 37–50
9. Shin, J.J. *et al.* (2020) Controlled cycling and quiescence enables efficient HDR in engraftment-enriched adult hematopoietic stem and progenitor cells. *Cell Rep.* 32, 108093
10. Sun, S. *et al.* (2019) Tissue specificity of DNA damage response and tumorigenesis. *Cancer Biol. Med.* 16, 396–414
11. Yan, K.S. *et al.* (2012) The intestinal stem cell markers Bmi1 and Lgr5 identify two functionally distinct populations. *Proc. Natl. Acad. Sci. U. S. A.* 109, 466–471
12. Schwank, G. *et al.* (2013) Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* 13, 653–658
13. Weisheit, I. *et al.* (2020) Detection of deleterious on-target effects after HDR-mediated CRISPR editing. *Cell Rep.* 31, 107689
14. Ma, H. *et al.* (2017) Correction of a pathogenic gene mutation in human embryos. *Nature* 548, 413–419
15. Egli, D. *et al.* (2018) Inter-homologue repair in fertilized human eggs? *Nature* 560, E5–E7
16. Kosicki, M. *et al.* (2018) Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* 36, 765–771
17. Adikusuma, F. *et al.* (2018) Large deletions induced by Cas9 cleavage. *Nature* 560, E8–E9
18. Wilde, J.J. *et al.* (2021) Efficient embryonic homozygous gene conversion via RAD51-enhanced interhomolog repair. *Cell* 184, 3267–3280
19. Liang, D. *et al.* (2020) Frequent gene conversion in human embryos induced by double strand breaks. *bioRxiv* Published online June 20, 2020. https://doi.org/10.1101/2020.06.19.162214
20. Zuccaro, M.V. *et al.* (2020) Allele-specific chromosome removal after Cas9 cleavage in human embryos. *Cell* 183, 1650–1664

21. Alanis-Lobato, G. *et al.* (2021) Frequent loss-of-heterozygosity in CRISPR-Cas9-edited early human embryos. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2004832117

22. Komor, A.C. *et al.* (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424

23. Gaudelli, N.M. *et al.* (2017) Programmable base editing o A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551, 464–471

24. Anzalone, A.V. *et al.* (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157

25. Gilbert, L.A. *et al.* (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442

26. Qi, L.S. *et al.* (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183

27. Nuñez, J.K. *et al.* (2021) Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* 184, 2503–2519

28. Wienert, B. *et al.* (2020) Timed inhibition of CDC7 increases CRISPR-Cas9 mediated templated repair. *Nat. Commun.* 11, 2109

29. Lin, S. *et al.* (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* 3, e04766

30. Charpentier, M. *et al.* (2018) CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat. Commun.* 9, 1–11

31. Matsumoto, D. *et al.* (2020) A cell cycle-dependent CRISPR-Cas9 activation system based on an anti-CRISPR protein shows improved genome editing accuracy. *Commun. Biol.* 3, 1–10

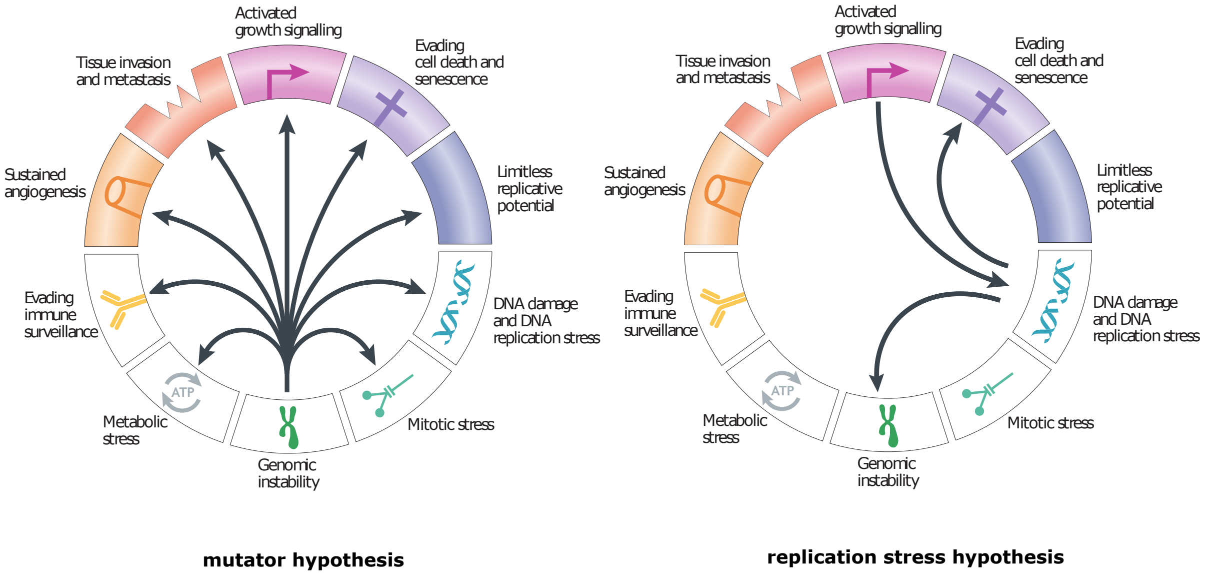
32. Suzuki, K. *et al.* (2016) In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* 540, 144–149

DDR as Protector and Promoter in Tumorigenesis

To understand the central role of DDR in cancer development, two standing theories in the field must be discussed. The first is the so called “mutator hypothesis”, which assumes that genomic instability is a preexisting condition in pre-neoplastic lesion and eventually drives tumor progression (Loeb, 1991; Nowell, 1976; Kinzler & Vogelstein, 1997). The second theory is known as the “oncogene induced replication stress model”, which contextualizes genomic stress as a consequence of breakage and failure of DDR, consequent to oncogene driven replication. According to the replication stress hypothesis, activation of oncogenes, and subsequent increases in proliferation lead to replication stress and frequent replication fork collapse. This in turn results in DNA breaks in such abundance that even normally functioning DDR cannot repair all damages (Halazonetis *et al*, 2008; Gorgoulis *et al*, 2005; Bartkova *et al*, 2005, 2006; Di Micco *et al*, 2006). By comparing and contrasting both hypotheses, it becomes clear that both offer a good explanation for some observations but neither model offers a universal fit that would explain the double-edged role DDR plays in cancer development.

The mutator hypothesis is often cited in the context of hereditary cancers, such as Lynch Syndrome, also known as hereditary non-polyposis colon cancer (HNPCC). This condition is defined by mutations in mismatch repair genes, leading to increased mutation burden and microsatellite instability. Other well-known examples include hereditary breast cancer, caused by mutations in double strand break repair genes BRCA1 and BRCA2, and skin cancer, caused by germline defects in nucleotide excision repair genes (Negrini *et al*, 2010). The observed frequency of cancer incidence in patients with germline mutations in DDR genes makes it obvious that DDR acts as a barrier to cancer development and that the normal functioning of these pathways is required to hinder the evolution of cancerous cells. Another strong argument for the mutator hypothesis is the frequent mutation of TP53 and ATM (ataxia telangiectasia mutated) in human cancers. As the central hub of integrating and relaying DNA damage, repair, and apoptosis signals, p53 as well as ATM underlie selective pressure to become inactivated in order to facilitate multiple cancer hallmarks, including sustained proliferation and evasion of apoptosis (Negrini *et al*, 2010) (Figure 13). Taken together, these observations argue that genomic instability is a pre-cancerous condition, enabling and driving tumorigenesis in hereditary cancer, and to some extent also in sporadic cancers, via mutating apex DDR signaling molecules like TP53 and ATM.

Paradoxically, despite the clear role DDR plays in the prevention of tumorigenesis, DDR genes are not significantly more frequently mutated in most sporadic human cancers. A possible explanation lies in the recessive nature of DDR mutations, meaning mutation of both alleles is less likely than mutation of the remaining single functional allele in the case of hereditary DDR defects. Thus, the central argument of the mutator hypothesis does not hold, because in most sporadic cancers, genomic instability does not seem to arise from mutations in DDR genes. Instead, the replication stress hypothesis offers an explanation for the observed genomic instability in the absence of DDR mutations. The oncogene induced proliferation leads to stress and collapse of replication forks, leading to DNA breakage so frequent, it overwhelms the DNA damage and repair machinery. Coupled with defects in cell cycle checkpoints, this quickly leads to an accumulation of unrepaired lesions, mutations, and chromosomal aberrations classically defining the hallmark of genomic instability (Negrini *et al*, 2010) (Figure 13). Thus, the resulting genomic instability could be viewed as a secondary effect rather than a primary driver of tumorigenesis. However, defining DDR defects solely by the presence of mutations in DDR genes underestimates the complexity of the DDR signaling network.


Activated growth signalling
Evading cell death and senescence
Limitless replicative potential
DNA damage and DNA replication stress
Mitotic stress
Genomic instability
Metabolic stress
ATP
Evading immune surveillance
Sustained angiogenesis
Tissue invasion and metastasis

mutator hypothesis

replication stress hypothesis

Figure 13. Sequence of cancer hallmark evolution according to mutator hypothesis and replication stress hypothesis (adapted from Negrini *et al*)

Cancer is as much a signaling disease as it can be defined via the mutations recorded in the genome (Yaffe, 2019). Without integrating information from multiple layers of biological organization, the full phenotypic effect of genetic changes cannot be appreciated. Even synonymous mutations in coding regions can function as cancer drivers, as is the case with TP53 (Supek *et al*, 2014), where synonymous mutations adjacent to splice sites disrupt the function of the downstream gene product. The intricacy of signaling from DNA sequence, via transcription, splicing, translations, post-translational modification, and protein structure offers limitless possibilities to alter signaling in favor of carcinogenesis.

Even with the rise of rapid multi-omics techniques, assessing every level of biological regulation remains a large challenge, both, in terms of cost and bottlenecks in experimentation and data analysis. However, with the rise of next generation sequencing technologies it has become more feasible to obtain whole genome sequencing (WGS) for many samples, including clinical samples, which provide limited analysis material. Due to the accessible nature of sequencing information, understanding cancer development through the lens of genomics has made great advances. Nevertheless, single mutations alone cannot capture all biological complexity and the maxim that the whole is more than the sum of its parts holds true here. Excitingly, new computational approaches offer new solutions to analyze the information contained in WGS. Mutational signatures are an analytical framework which allow to analyze the sequencing data beyond the impact of individual mutations and instead map the genomic imprints of function or dysfunction of pathways. Therefore, mutational signatures allow access to a higher level of biological organization while retaining single mutation resolution. Mutational signatures aid the understanding of cancer etiology, tumor evolution, and offer insights into targeted treatment options for specific patient populations. The technological and analytical advances which enable these exciting developments are discussed in detail in the next section.

Cancer Genomes – Understanding Cancer Development with Omics Data

Rise of high Throughput Data Generation

DNA Sequencing

The earliest sequencing methods which were able to infer exact nucleotide position, and thus sequence, were developed by Frederick Sanger and Alan Coulson, and concurrently by Allan Maxam and Walter Gilbert in the mid 1970s (Heather & Chain, 2016). Sanger and Coulson's so called plus-and-minus technique relied on extension of the DNA template by DNA polymerase and incorporation of radiolabeled nucleotides. The plus reaction would contain only one type of nucleotide, terminating the extension of all fragments at that nucleotide. The minus reaction would contain the three other nucleotide types. By performing these reactions in parallel and separating all resulting fragments on a polyacrylamide gel, the original sequence could be inferred. While relying on the same principle of puzzling fragments together, Maxam and Gilbert developed a different approach. Instead of starting with natural DNA and incorporating labelled nucleotides by synthesis, they chemically broke apart radiolabeled DNA fragments at specific positions and could thus infer the position of specific nucleotides on the gel (Heather & Chain, 2016). What both methods have in common is a labor-intensive protocol and a requirement for radioactive labelling material. Nevertheless, the development of these techniques would lay the foundation for modern sequencing as we know it.

Only a few years later, in 1977, Frederick Sanger developed the breakthrough dideoxy 'chain termination' method, which would become known as first generation sequencing or simply, Sanger sequencing (Sanger *et al*, 1977). Based on the same principle of stopping the DNA polymerase reaction to 'sequence' the identity of the nucleotide, Sanger sequencing build upon the simple chemical principle of DNA elongation. By using small concentrations of chemically altered nucleotides lacking the 3'-hydroxyl group required for DNA synthesis, the DNA polymerase would stall at each point where a dideoxunucleotide (ddNTP) is incorporated and terminate the DNA chain. This new method simplifies the workflow of the earlier method by using 4 radiolabeled ddNTPs, performing 4 parallel reactions, and then running 4 lanes on a gel to infer the sequence of the fragments. The simplified protocol was widely adapted and is still in use today, albeit with modern modifications. Substituting radiolabeling with 4 unique fluorescence-based labels allows to even perform all 4 parallel reactions in one vial. Automated sequencers using Sanger sequencing were the technology of choice when the first draft of the human genome project was produced (Heather & Chain, 2016).

Later developments, so called second-generation, or next-generation sequencing, would push the data generation forward by massively parallelizing DNA sequencing (Heather & Chain, 2016). Throughput increased while cost decreased in an exponential fashion, driving not only the completion of the human genome project but also increased use of sequencing in the clinical context. The idea of understanding every nuance of life if only we could read the complete genetic code was tantalizing. Now, in the post-genomic era, we marvel at the complexity of biological life, which cannot be fully understood through only one level of biological regulation. Nevertheless, the decoding of the entire human genome was a milestone achievement and marked the start to a new era in research focused on data (Tsui & Scherer, 2008). The wealth of cancer data available now allowed for a fresh perspective on old problems. How does cancer evolve? What drives the evolution of the cancer? How does cancer-heterogeneity emerge? How can we predict treatment sensibilities?

Using the available data and developing new mathematical frameworks to systematically understand cancer genome data facilitated some impactful discoveries. In the following the theoretical background and application of mutational signature analysis are discussed in detail.

Mutational Signatures – A Framework to Functional Understanding of Cancer Genomes

General Framework

For a long time, cancer genetics research has focused on identifying relevant driver mutations, largely ignoring passenger mutations, in part because of their uninterpretable nature. Seemingly random scatterings of mutations across the genome, passenger mutations in fact hold a large amount of information about the origins and evolution of a tumor(Nik-Zainal & Morganella, 2017). Upon closer inspection, it is crucial to note that mutation generation across the genome is not a random process, but rather the consequence of the interplay between cellular processes. A lesion to DNA is followed by an attempt to repair and the outcome of such attempt relies on the type of lesion, cellular and tissue context, and the complex inter-pathway signaling between DNA repair mechanisms as described above. The collection of resulting mutations is thus far from random and can be interpreted as a lifetime record of normally and aberrantly functioning cellular processes, inscribed in DNA (Koh *et al*, 2021a).

The mutational signatures field started with examining single base substitution signatures but has since evolved. Now mutational signatures of different classes of mutations can be described, including double base substitutions, insertion and deletions, as well as rearrangements. Each type of mutation gives rise to unique profiles, thus, developing a signature framework for different mutation categories allows further insight into the mechanisms generating these signatures. In the following each class of mutational signature is described in more detail.

Single Base Substitution Signatures

The simplest mutations involve the change of a single base, broadly categorized in two mutation classes. Transition mutations describe the change within the purine and pyrimidine classes respectively, meaning mutations A <-> G and C <-> T. Transversions, on the other hand describe point mutations involving changes from purine to pyrimidine bases or vice versa.

A structured way to list all possible point mutation types is to list them, by convention, pyrimidine bases first C>A, C>G, C>T, T>A, T>C, and T>G. These 6 types of single base mutations are more informative when viewed within their sequence context. For this reason, all 16 possible combinations of preceding and proceeding nucleotides are applied to the 6 mutation classes, which yields a matrix with 96 positions. The total number of mutations recorded for each position in the matrix gives a 96-channel profile that constitutes the overall mutational profile of a sample (Alexandrov *et al*, 2013; Nik-Zainal *et al*, 2012a) (Figure 14). The single nucleotide substitution framework can be further extended by considering a pentanucleotide context. Extraction and analysis of the complete mutational profile is performed using mathematical methods and algorithms further discussed later.

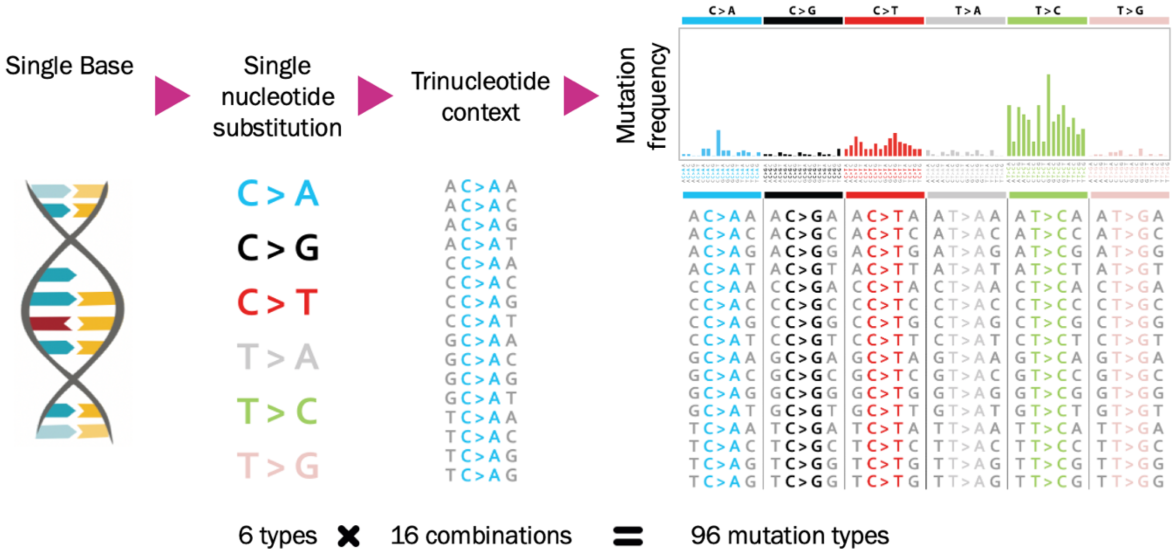


Figure 14. Single base substitutions are denoted in six distinct classes (written pyrimidine first by convention). Embedding the six mutation classes in the trinucleotide context gives rise to sixteen possible combinations for each of the six basic mutation types. Completing the combinatorics for all 6x16 combinations yields a matrix with 96 possible mutations in their respective trinucleotide context. The 96 channels represent the total frequency of single base mutations in their trinucleotide context in the entire genome.

Double Base Substitution Signatures

If two adjacent bases mutate together, a total of 10 source mutation types can be listed. Source doublets describe the two original bases mutated to any two other bases, *e.g.* AC>NN. For each source doublet, all 9 combinations for target doublets are considered, *e.g.* AC> CA, AC>CG, AC>CT, AC>GA, AC>GG, AC>GT, AC>TA, AC>TG, and AC>TT. The combinatorial process yields 90 different source to target doublet mutations. Subtracting the 12 reverse complement combinations then returns a final matrix with 78 possible doublet base substitutions (Alexandrov *et al*, 2020). The mutation frequency of each of the 78 channels can be visualized similarly to SBS signatures (Figure 15).

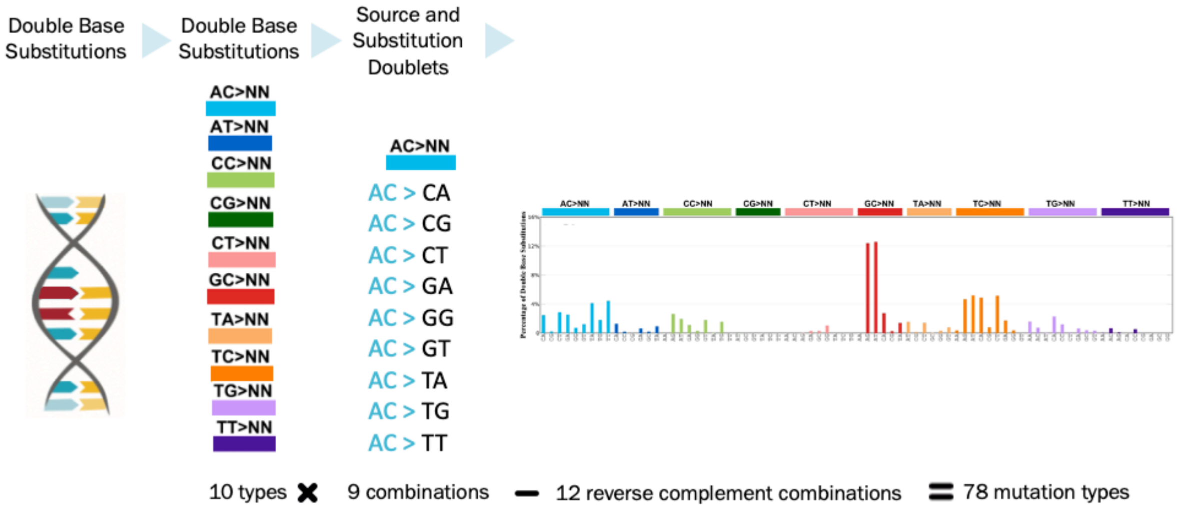


Figure 15. Consensus notation for 78 channels of double nucleotide substitution profile.

Insertion – Deletion Signatures

The removal or insertion of one or multiple bases results in indels. This class of mutations is difficult to fit into a defined matrix of all possible mutations because the combinations of different base indels of different lengths and sequence context are near endless. Instead, a framework of 83 curated indel types was developed, which generalizes common types of indels (Alexandrov *et al*, 2020). This includes 1 base pair (bp) C and T (G, A respectively) insertions and deletions in sequence contexts of varying length. Furthermore, 2bp-5bp insertions and deletions are listed in repetitive regions, as well as deletions between 2-5 bp with microhomology regions (Figure 16).

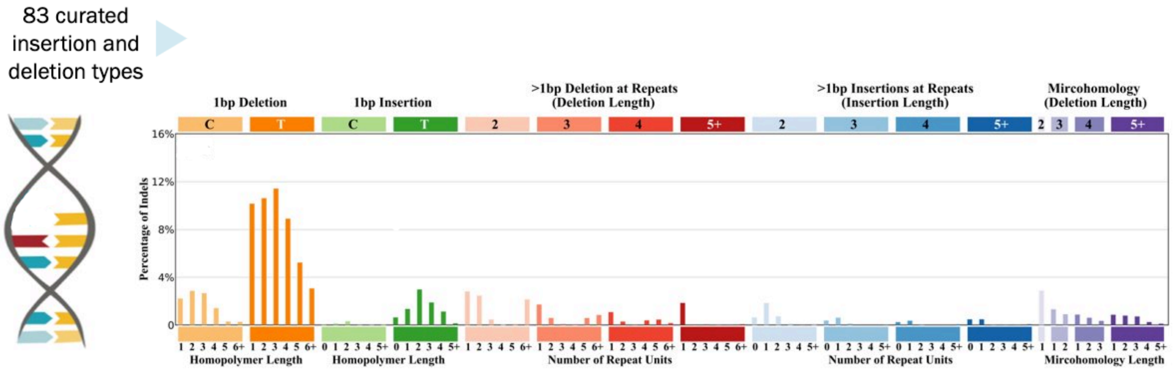


Figure 16. Consensus notation for 83 channels of curated indel profile.

Copy Number Variations and Complex Rearrangements

Lastly, copy number variations and rearrangements are types of mutations often encountered in cancers with high levels of genomic instability. The standardized way to catalogue these types of mutations includes binning the observed alterations by type, length and context. For copy number signatures, there are 48 defined channels, tracking the loss of heterozygosity (LOH) and heterozygosity status (Het). For each of these categories, the total copy number variation is accounted for 0 to more than 9 for LOH, and 2 to more than 9 for Het. Additionally, each bin with a given copy number alteration is further divided by the size of the affected fragments (Steele *et al*, 2022b, 2022a) (Figure 17).

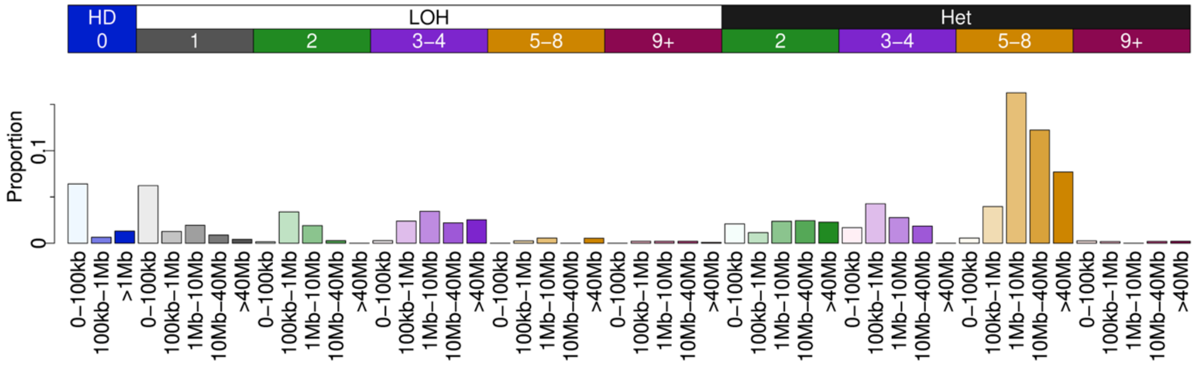


Figure 17. Consensus notation for Copy Number Signatures. Signature profile from (COSMIC, 2020), see Appendix for licensing information

Similarly for complex rearrangement signatures, the types of rearrangements are classified into deletions, tandem duplications (TD), inversions, and translocations. These 4 categories are further described by subdividing occurrences into clustered rearrangements and non-clustered rearrangements, as well as binning observations by fragment size (Nik-Zainal *et al*, 2016) (Figure 18).

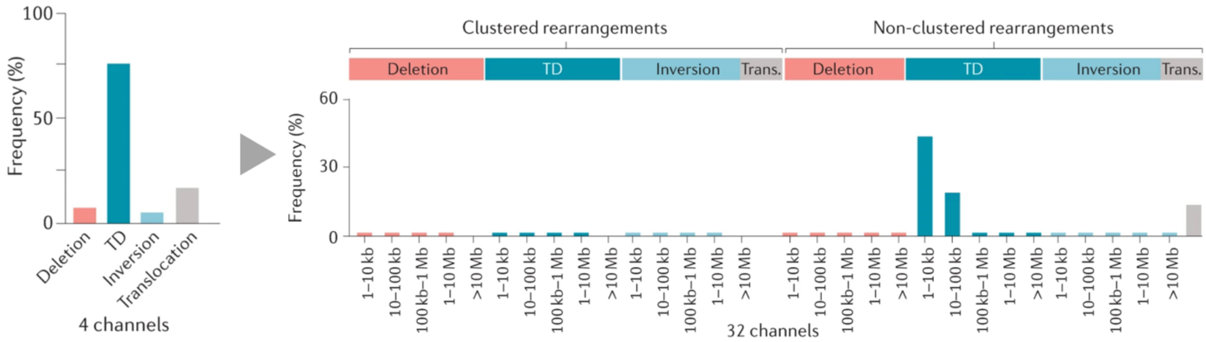


Figure 18. Mutation type classification of Rearrangement Signatures in 4 main channels and extended channels, by fragment size. Figure taken from (Koh *et al*, 2021b) *Reproduced with permission from Springer Nature* (see Appendix*)*

Computational Approaches for Analyzing Mutational Signatures

The original definition of mutational signatures is based on the de-novo extraction of signatures from cancer samples. De-novo extraction involves decomposing the entire mutational profile of a cohort into individual signatures with non-negative matrix factorization (NMF). Once a catalog of stable signatures was defined, other approaches such as signature refitting were developed, which are also suitable for cohorts with lower sample numbers. In signature refitting, mutational profiles of samples are modelled as a linear combination of multiple known spectra, resulting in an estimate of signature activities within each sample (Omichessan *et al*, 2019; Baez-Ortega & Gori, 2019). Below the main aspects and differences of these two approaches are discussed.

NMF is useful in decomposing large and sparse datasets, such as mutations found in entire genome sequences. The main goal is to decompose the given matrix into two lower-rank matrices, such that the product of these two matrices approximates the original data matrix. This calculation requires that all elements of the input matrix are non-negative. Mathematically, this process can be described as follows. Any sequenced genome contains a set of mutations $m_{g,}$ which in sum represent the combination of all present mutational processes (p) with a given activity called exposure (e) and distribution of mutation types (k) (Nik-Zainal *et al*, 2012a).The mutation type k is variably defined depending on which channels are chosen for analysis. For Single nucleotide substitution signatures, k=96, for indel signatures k=83, and for double base substitution signatures k=78. The input matrix with all present mutations can thus be expressed as the sum of the product between processes and exposures:

$$m_g^k \approx \sum_{n=1}^{N} p_n^k e_g^n$$

By representing exposures to mutational processes (e) and mutational catalogs (m_g) as matrices, this term can be applied to all mutation types (k) and genomes (g).

$$\begin{pmatrix} m_1^1 & m_2^1 & \cdots & m_{G-1}^1 & m_G^1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ m_1^K & m_2^K & \ldots & m_{G-1}^K & m_G^K \end{pmatrix} \approx \begin{pmatrix} p_1^1 & p_2^1 & \cdots & p_{N-1}^1 & p_N^1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_1^K & p_2^K & \ldots & p_{N-1}^K & p_N^K \end{pmatrix} \times \begin{pmatrix} e_1^1 & e_2^1 & \cdots & e_{G-1}^1 & e_G^1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ e_1^N & e_2^N & \ldots & e_{G-1}^N & e_G^N \end{pmatrix}$$

With this notation it becomes obvious that the mutational catalog matrix M is decomposed into the product of the mutational process matrix P and the exposure matrix E, or simply:
M ≈ P x E. The closer the product between P and E resembles the original data M, the lower the reconstruction error, and the more stable the solution when the decomposition is repeated in multiple iterations (Baez-Ortega & Gori, 2019). While the mathematical approach to this problem is not new and was originally applied to biological data mora than 20 years ago (Lee & Seung, 1999, 2001), the application to genomic data only followed in 2012 (Nik-Zainal *et al*, 2012a).

The NMF framework of course has some limitations. The first and obvious limitation is that the type and number of unique processes recoverable from a given input depends on the type and number of channels defined. Consequently, mechanistic insight into signature generation might be limited, depending on whether biologically meaningful mutational channels were defined (Koh *et al*, 2020a). This is especially critical for mutational signature classes which do not cover the entire possible combinatorial space of mutations and sequence contexts for a given mutation type, *i.e.* indel signatures, copy number signatures, and rearrangement signatures (Koh *et al*, 2021). If too few channels are defined, signatures of differing origin might not be separable. If too many channels are redundant, this can lead to calling of spurious and biologically meaningless signatures. Hence it is paramount to define a reasonable number of informative channels to derive biological insight.

The second limitation deals with the accuracy of the extracted signatures, which depends on several properties of the input data, including the number of genomes available (sample size), the number of mutations present in the genomes, and the number of signatures to be extracted. Generally, the more signatures are extracted, the more the data is overfit and the more the stability of the solution decreases. Hence, the extraction of more signatures requires an increase in sample size. Fortunately, more cancer genome sequences have become available with time, enabling the discovery (Alexandrov *et al*, 2020; Nik-Zainal *et al*, 2016) and documentation of standard signatures in a dedicated database (COSMIC – Consensus of Somatic Mutations in Cancer) (Tate *et al*, 2019).

The selection of the optimal number of signatures to extract remains a critical parameter influencing the analysis. An alternative approach to manually determining the number of active mutational processes in a mutational catalogue is expectation maximization. This approach incorporates an underlying probabilistic model which uses the Bayesian Information Criterion (BIC) to assess the number of active signatures present (Fischer *et al*, 2013). A major assumption of this model is that all input samples are independent from each other, which might not hold true in all situations. Newer implementations of NMF algorithms incorporate automated rank selection, for instance NMFk, to objectively approach the optimal number of signatures for extraction (Islam *et al*, 2022). By comparing the distance between the original and recreated profiles, the NMFk approach aims to maximize the tradeoff between both, the stability of the solution and the correctness of the reconstructed data (Nebgen *et al*, 2020).

Especially in smaller cohorts, NMF can lead to high false positive rates, calling unstable signatures. However, the availability of a curated set of mutational signatures opens another possibility of analysis that allows to analyze even single samples. Instead of extracting signatures de-novo, the mutational catalogue of a sample or cohort can be modeled as a linear combination of known signatures. This approach is commonly called signature refitting (Baez-Ortega & Gori, 2019; Omichessan *et al*, 2019). While this approach offers the advantage that the sample size is not a limiting factor, is has several other limitations. First, matching the samples' mutational profile to a finite set of known signatures prohibits the discovery of new signatures. Second, many mutational signatures share features across the defined mutational types k. In the case of SBS signatures there are 60 known signatures and 19 signatures of possible sequence artifacts, which commonly share mutational features such as C>A, C>G, and C>T mutations. Modelling a mutational catalogue as a linear combination of all possible mutational signatures thus poses the risk of misattributing mutations of shared features between signatures. This occurs because standard refitting algorithms will include signatures to improve the fit to the data, even when the signatures contribute very little to the overall mutational profile or make no biological sense. To avoid this problem, it makes sense to include prior knowledge and restrict the set of known signatures used for refitting to signatures which are expected to be active. Using prior knowledge reduces the risk of overfitting, however, it also introduces bias and is difficult to apply when little or no prior knowledge is available (Baez-Ortega & Gori, 2019).

To overcome bias introduced with prior knowledge, the backwards removal approach can be applied, where the signature which contributes least to the fit is iteratively removed until a minimal set of signatures that fits the data well is found. This works by calculating the cosine similarity between the original and the modelled profile prior to and after removing the least contributing signature. If the difference in cosine similarity distance between the two iterations is higher than a given cutoff, the removed signature should be retained and the best minimal set of fitted signatures is found. Since the stopping threshold is arbitrarily set, this process should be repeated with multiple thresholds to empirically find the best subset of known mutational signatures explaining a given samples' profile. Furthermore, some bias may be introduced with the order of removal of signatures. An improved implementation of this approach works by choosing a reasonable set of mutational signatures and refitting every possible combination subset of signatures to the samples. This process is repeated n-1 times, where n is the total number of known signatures, for each subset n-2, n-3, n-4 etc., to retain the combination of signatures which best models the data. This best subset approach allows to incorporate prior knowledge and is more accurate than the backwards removal approach but becomes computationally infeasible when too many known signatures are included (Blokzijl *et al*, 2018).

Finally, bootstrapped refitting circumvents bias introduced with the use of prior knowledge while still working on cohorts with lower sample numbers. This approach is especially useful in verifying the stability of a refit in cohorts with low sample numbers. First, the mutational matrix is resampled with replacement, using the original distribution of mutations as weights. A standard signature refit is performed for each sampling iteration, allowing to estimate statistics on the accuracy of the refit by recording the contribution of each signature to each iteration of the resampled mutation matrix. Signatures which are found active in most bootstrapped iterations can thus be interpreted as a stable refit and are likely truly active in the given sample (Blokzijl *et al*, 2018).

Mutational Signatures of DNA Repair Deficiencies

Mutational signatures reflect the action or dysfunction of an entire pathway and thus capture a level of biological activity which is not accessible through studying individual mutations only. One of the best examples is signature SBS3, which describes the so called BRCAness phenotype, i.e., a defect in homologous recombination. Several coding mutations within the BRCA1 and BRCA2 genes are known to cause a defect in the pathway due to impacting the function of the BRCA proteins. These variants are included in clinical tests for cancer predisposition screening and risk assessment. Furthermore, patients with mutations in BRCA genes benefit from PARP1 inhibitor treatment, as a synthetic lethality exists between the processes these genes facilitate (Helleday, 2011; Lord & Ashworth, 2017). Importantly, the interaction which leads to PARP1 dependence and thus gives the therapeutic effect, happens on the pathway level. Proteins enact their function not just through their sequence and structure but also through their specific post-translational modifications, subcellular localization, and time dependent interaction with other proteins and complexes. Hence, individual mutations do not reflect all possible disturbances to pathway functionality. Mutational signatures, on the other hand, capture overactivity or dysfunction of pathways and records these in distinct mutational patterns in the genome. Mutational signatures thus contain more information than individual mutations. Ultimately, signatures contain information about the history and progression of pathway interaction throughout the development of a tumor and help to guide patient stratification for treatment decisions. The HR-deficiency signature SBS3 is a prominent example illustrating all the points raised above. The signature can arise from known or unknown mutations in causative genes, i.e., BRCA1, BRCA2, or PALB2 (Nguyen *et al*, 2020). However, taken together, these cases comprise only about half of all cases which show a defect in HR and subsequent SBS3 activity (Figure 19) (Turner, 2017). The other half may be caused by epigenetic or other regulatory mechanisms. A recent study, for instance, identified that accumulation of succinate, fumarate, and 2-hydroxyglutarate precipitate HR defects though interfering with the function of lysine demethylases KDM4A/B (Sulkowski *et al*, 2018, 2020). The demethylases are needed clear methylation marks at DNA breaks sites to facilitate recruitment of repair factors, including ATM and TIP60 (Sulkowski *et al*, 2020). Hence, there are driver processes which are independent of mutations in causative genes, but phenocopy the BRCAness, and with it the therapeutic vulnerability to PARP1 inhibitors. Patient stratification for optimal treatment decisions is thus far more informative when based on mutational signatures.

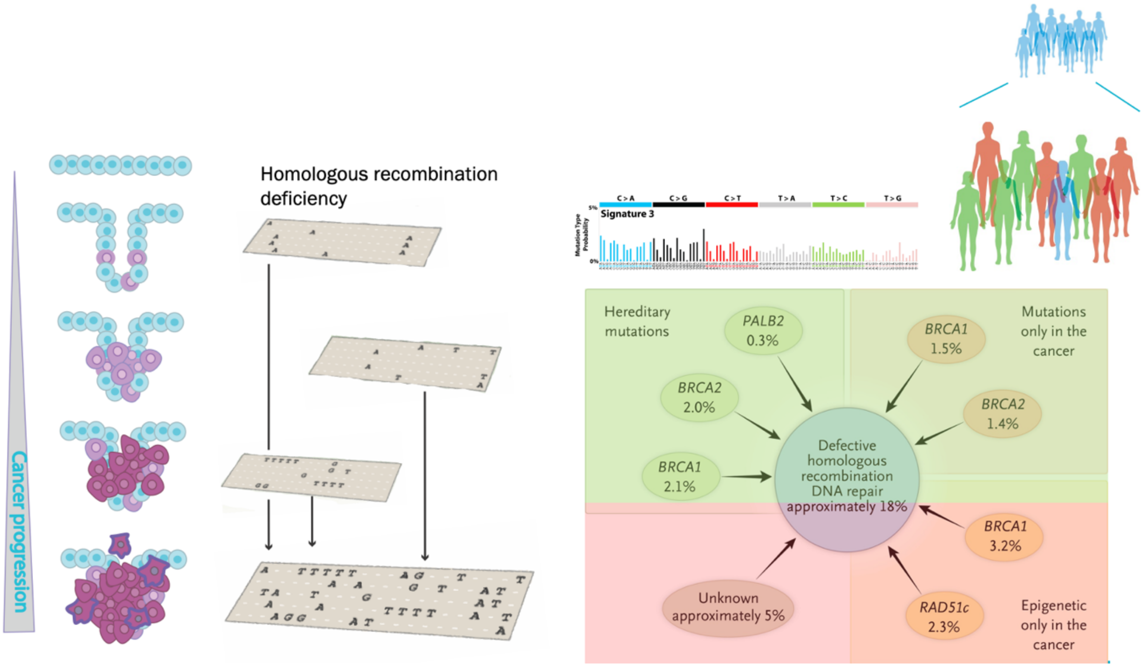


Figure 19. Schematic overview of the decomposition of the overall mutational profile into individual mutational signatures (left) and their biological and clinical interpretation using SBS3 as an example (right). Figure adapted from (Helleday *et al*, 2014; Turner, 2017).

There are many other signatures which arise due a defect in DNA repair (Alexandrov *et al*, 2020). Notably, HR signatures are not restricted to SBS signatures. Beside SBS3, ID6, and CN17 could also be attributed to defects in homologous recombination. Similarly, mismatch repair deficiency is also a prominent cause of multiple signatures across classes. Namely, SBS6, SBS15, SBS21, SBS26, SBS44, DBS7, DBS19, and ID7, were attributed to faulty activity of the mismatch repair pathway (Forbes *et al*, 2017; Tate *et al*, 2019). The variety of different signatures and sequence contexts associated with mismatch repair deficiency might reflect the broad spectrum of lesions generated and repair attempts conducted upon encountering a mismatch in various contexts.

The base excision repair deficiency related signatures SBS30 and SBS36 are characterized by two specific mutations in BER factors, NTHL1 and MUTHY, respectively. Mutations which disrupt the function of these glycosylases cause the accumulation of the distinctive C>T mutations for SBS30 and C>A for SBS36 (Tate *et al*, 2019; Drost *et al*, 2017b). Since MUTYH was previously implicated in the removal of 8-oxo-G lesions, it is not surprising that both experimental and computational studies on the MUTYH associated signature found great similarity to SBS18, which is proposed to be caused by oxidative damage (Viel *et al*, 2017; Pilati *et al*, 2017; Tate *et al*, 2019). Both, MUTYH and NTHL1 are implicated in rare hereditary tumor syndromes, predisposing to the development of colorectal cancer, and to a lesser extend to endometrial, cervical, and bladder cancer (Magrin *et al*, 2022; Das *et al*, 2020; Robinson *et al*, 2022). The distinctive signatures caused by the respective gene defect is testament to the non-redundant functions these genes fulfill in BER. Furthermore, it illustrates that mutational signature analysis can distinguish different mechanisms of cancer development within the same tumor types, in this case even with single gene resolution in the same pathway.

Beside deficiency in a DDR pathway, there are also signatures which are caused by the overactivity of certain pathways or enzymes. One example is ID8, where the overactivity of NHEJ, possibly exclusively or in conjunction with mutations in topoisomerase2 (TOP2A) causes the accumulation of >5 bp deletions at small repeat units, as well as >5 bp deletions at sites of microhomologies of varying length (1-5 bp MH sequences) (Alexandrov *et al*, 2020; Tate *et al*, 2019). Another example of increased activity causing mutagenesis are the APOBEC-family and AID enzymes. APOBEC enzymes are cytidine deaminases, which depending on their activity and accessibility to DNA cause C>T transitions defining SBS2 and C>G transversions characterizing SBS13 (Chan *et al*, 2015). Aid enzymes are activation induced cytidine deaminases, which cause a less defined mutational spectrum in SBS84 and SBS85 than APOBEC enzymes (Kasar *et al*, 2015).

Beyond classical DNA repair factors, there are also multiple polymerases which play an important role across pathways. Hence, mutations altering the function of these polymerases can account for various signatures. SBS9 is thought to be caused by faulty activity of POLH, which is involved in TLS, NER, and BER, as well as class switch recombination. Mutations in the exonuclease domain of POLE are the attributed etiology for SBS10a and SBS10b, which are characterized by distinct clusters of C>A and C>T mutations respectively (Alexandrov *et al*, 2020; Li *et al*, 2018). The distinct patterns arising from mutations in the same subdomain of the same gene might be explained by the diverse roles POLE plays in SSBR, NER, BER, as well as in replication (Ma *et al*, 2018). Therefore, the mutational imprint the activity of mutated POLE leaves behind, may be context dependent. Finally, mutations in subunit 1 of POLD impairs the proofreading function, which introduces C>A type mutations during processes such as NER, TLS, and replication. POLD1 mutations are the attributed cause for SBS10c and SBS10d (Robinson *et al*, 2021).

DDR Signatures with Emergent Etiologies

Interestingly, not all signature etiologies can be traced to a single gene or pathway. These signatures have emergent properties as they arise due to the interaction of two or more genes or processes. SBS14 and SBS20 are examples of this principle. SBS14 is the result of concurrent mutations in polymerase epsilon (POLE) and MMR-deficiency and SBS20 arises in a context where defective MMR co-occurs with POLD1 mutations. Both signatures are mainly characterized by C>A mutations but show very distinctive patterns within this mutation class (Hodel *et al*, 2020; Meier *et al*, 2018).

Apart from SBS signatures, emergent properties of signatures also become apparent in other signature classes, such as in rearrangement signatures (Nik-Zainal *et al*, 2016). Two DNA repair related emergent signatures are the POLQ and FANC related rearrangement signatures which both occur in a BRCA1 deficient genetic context. Specifically, in these rearrangement signatures, characterized by tandem duplications (TD), the genetic background of HR-deficiency interacts with the repair pathway choice in a time-dependent manner during replication. The POLQ specific mutational outcome of microhomology flanked indels at breakpoints, as well as templated insertion, occurred within early replicating domains of the genome (Kamp *et al*, 2020; Mateos-Gomez *et al*, 2015; Ceccaldi *et al*, 2015). Oppositely, a rearrangement signature with a different pattern of TDs, was proposed to be linked to FANC gene mutations, found to be enriched in late replicating domains (Li *et al*, 2020). Specifically, the POLQ dependent repair in HR-deficient backgrounds, points to an exploitable therapeutic vulnerability, highlighting the clinical impact of the biological insight that can be derived from mutational signatures.

Other examples of emergent signatures include signatures of chromosomal instability (Drews *et al*, 2022). The signatures of chromosomal instability are defined based on features of copy number alterations that were previously observed to occur on a chromosomal level, including breakpoint count per 10 megabases, breakpoint count per chromosome arm, length of copy number segments, and the copy number change between two neighboring segments (Macintyre *et al*, 2018). Based on these features, Drews *et al* extract CX signatures they could relate back to specific observed patterns of changes. CX3, for instance, is characterized by the occurrence of long segments with single copy changes and is proposed to be caused by defective HR in the presence of replication stress and dysfunctional damage sensing. The biological interpretation of CX3 is supported by the differentially higher activity of CX3 in tumors with somatic BRCA1 mutations and RAD51C methylation, as well as correlation with known HR-deficiency signatures of other classes (SBS3 and ID6) (Drews *et al*, 2022). Thus, CX3 seems to capture a biologically complex process of mutational generation that is active in a specific biological context where multiple factors interact simultaneously to create the outcome.

Although the examples discussed above clearly showcase the utility of analyzing complex and emergent signatures, neither rearrangement signatures (RS) nor signatures of chromosomal instability (CX), have been included in the harmonized catalogue of mutational signatures (COSMIC database) (Table 1). A major difficulty of analyzing these types of signatures lies in the definition of mutational channels before matrix decomposition. Which features and sequence contexts are considered important varies between studies and thus hinders cross-study comparisons (Koh *et al*, 2021b). Defining which mutational signature channels to include into the definition of a signature type must balance including biologically relevant channels with excluding channels carrying redundant or noisy information. This problem is not trivial and will define the future of the field.

The diversity in classes of signatures caused by DNA repair defects (SBS, DBS, ID, CN, RS, CX), is testament to the complexity of the DDR, where various lesions and intermediate repair products can be substrate for DNA repair processes. The variety in substrates and possible repair reactions, successful and unsuccessful, is mirrored in the diversity of mutational outcomes recorded across all classes of signatures. In summary, signatures with emergent properties are dependent on the genetic background and other cellular processes, such as the replication timing. Considering emergent characteristics of signatures illustrates the complex context dependent choice of DNA repair and the subsequent mutational outcome and holds great promise for furthering our understanding of these processes. Likely, a full understanding of cancer etiology will require more than mapping the space of all existing signatures. Further studies will be required to elucidate interactions between signatures and between genetic backgrounds and signature generation, which also highlights the need to develop new and innovative approaches to analyze this data.

Table 1: Consensus Mutational Signatures Listed in the COSMIC Database (Tate *et al*, 2019)

Signature class	Signature	Proposed etiology
Single Base Substitutions		
	SBS1	Spontaneous deamination of 5-methylcytosine (clock-like signature)
	SBS2	Overactivity of APOBEC enzymes (cytidine deaminases)
	SBS3	HR-deficiency
	SBS4	Tobacco smoking
	SBS5	Unknown (clock-like signature)
	SBS6	MMR-deficiency
	SBS7a	Ultraviolet light exposure

	SBS7b	Ultraviolet light exposure
	SBS7c	Ultraviolet light exposure
	SBS7d	Ultraviolet light exposure
	SBS8	Unknown
	SBS9	POLH (Polymerase eta) somatic hypermutation
	SBS10a	POLE (Polymerase epsilon) exonuclease domain mutations
	SBS10b	POLE (Polymerase epsilon) exonuclease domain mutations
	SBS10c	Defective POLD1 proofreading
	SBS10d	Defective POLD1 proofreading
	SBS11	Temozolomide treatment
	SBS12	Unknown
	SBS13	Overactivity of APOBEC enzymes (cytidine deaminases)
	SBS14	Concurrent MMR-deficiency and POLE mutation (Polymerase epsilon)
	SBS15	MMR-deficiency
	SBS16	Unknown
	SBS17a	Unknown
	SBS17b	Unknown
	SBS18	Damage by ROS
	SBS19	Unknown
	SBS20	Concurrent MMR-deficiency and POLD1 mutation
	SBS21	MMR-deficiency
	SBS22	Aristolochic acid exposure
	SBS23	Unknown
	SBS24	Aflatoxin exposure
	SBS25	Chemotherapy
	SBS26	MMR-deficiency
	SBS28	Unknown
	SBS29	Tobacco chewing
	SBS30	BER-deficiency (NTHL1 mutations)
	SBS31	Chemotherapy with platinum agents
	SBS32	Treatment with Azathioprine
	SBS33	Unknown
	SBS34	Unknown
	SBS35	Treatment with platinum agents
	SBS36	BER-deficiency (MUTYH mutations)
	SBS37	Unknown
	SBS38	Indirect effect of UV-light exposure
	SBS39	Unknown
	SBS40	Unknown
	SBS41	Unknown
	SBS42	Haloalkane exposure
	SBS44	MMR-deficiency
	SBS84	AID-activity (activation induced cytidine deaminases)
	SBS85	Indirect effect of AID-activity (activation induced cytidine deaminases)
	SBS86	Unknown chemotherapy treatment
	SBS87	Chemotherapy with Thiopurine
	SBS88	Colibactin exposure
	SBS89	Unknown
	SBS90	Duocarmycin exposure
	SBS91	Unknown
	SBS92	Tobacco Smoking
	SBS93	Unknown
	SBS94	Unknown

SBS possible sequencing artefacts	SBS27, SBS43, SBS45-SBS60, SBS95	Unclear origin, possible sequence artefacts, unstable signatures
Double Base Substitution Signatures		
	DBS1	Ultraviolet light exposure
	DBS2	Tobacco smoking, exposure to acetal aldehyde
	DBS3	POLE exonuclease domain mutations (Polymerase epsilon)
	DBS4	Unknown
	DBS5	Platinum chemotherapy agents
	DBS6	Unknown
	DBS7	MMR-deficiency
	DBS8	Unknown
	DBS9	Unknown
	DBS10	MMR-deficiency
	DBS11	Unknown
Insertion and Deletion Signatures		
	ID1	Slippage of the replicated strand during DNA replication
	ID2	Slippage of the replicated strand during DNA replication
	ID3	Tobacco smoking
	ID4	Unknown
	ID5	Unknown
	ID6	HR-deficiency
	ID7	MMR-deficiency
	ID8	NHEJ overactivity or mutations in TOP2A
	ID9	Unknown
	ID10	Unknown
	ID11	Unknown
	ID12	Unknown
	ID13	Ultraviolet light exposure
	ID14	Unknown
	ID15	Unknown
	ID16	Unknown
	ID17	Mutations in topoisomerase TOP2A
	ID18	Colibactin exposure
Copy Number Variation Signatures		
	CN1	Diploidy
	CN2	Tetraploidy
	CN3	Octoploidy
	CN4	Chromothripsis
	CN5	Chromothripsis
	CN6	Chromothripsis before whole genome duplication
	CN7	Chromothripsis associated amplification
	CN8	Chromothripsis associated amplification
	CN9	Focal loss of heterozygosity (LOH), diploid and chromosomal instability
	CN10	Focal loss of heterozygosity (LOH), 1x whole genome duplication
	CN11	Focal loss of heterozygosity (LOH), 2x whole genome duplication

	CN12	Focal loss of heterozygosity (LOH), 1x whole genome duplication, chromosomal instability
	CN13	Chromosomal loss of heterozygosity (LOH)
	CN14	Chromosomal loss of heterozygosity (LOH), 1x whole genome duplication
	CN15	Chromosomal loss of heterozygosity (LOH), 2x whole genome duplication
	CN16	Chromosomal loss of heterozygosity (LOH), 1x whole genome duplication, chromosomal instability
	CN17	HR-deficiency and tandem duplication
	CN18	Unknown
	CN19	Unknown
	CN20	Unknown
	CN21	Unknown
Possible Sequencing artefacts (CN)	CN22-24	Unclear origin, possible sequence artefacts, unstable signatures

Experimental Elucidation of Signature Etiologies

Generally, there are computationally driven top-down approaches and experimentally driven bottom-up approaches to infer signature etiology. For computational approaches a large sample size is required to even detect rare signatures. These top-down approaches have resulted in the discovery of new substitution and indel signatures (SBS, DBS, ID) recently (Degasperi *et al*, 2022; Alexandrov *et al*, 2020). Importantly, the increasing availability of sequencing data allows not just to discover new signatures, but aids interpretability since the data can be used to explore correlations between signature activity and clinical parameters or genetic background.

While computational extraction of signatures has yielded a great variety of different signatures (Table 1), the advances of in-silico signature detection have outpaced the experimental validation. Assigning an etiology is difficult due to the challenge of creating a clean experimental system which allows to study the generation of mutational signatures in the absence of confounding factors. Bottom-up approaches are useful for elucidating a mechanistic connection between a pathway and a signature. To experimentally validate mutational signatures a clean experimental system with a carefully matched control is needed to estimate both, background mutagenesis and de-novo mutagenesis due to the intervention of interest.

The first studies which showed that bottom-up signature validation is possible, focused on isogenic cell models and organoid systems (intestinal organoids), respectively. Both studies chose to focus on CRISPR-Cas9 mediated knockouts of DDR genes to observe mutagenesis over time, compared between parental and sub-clonal knock-outs (Drost *et al*, 2017a; Zou *et al*, 2018a). A follow up study could elucidate more specific mutational processes associated with individual DDR genes (Zou *et al*, 2021). And finally, a broad effort focused on chemical mutagenesis explored mutational signature generation in pluripotent stem cells in response to chemical exposure (Kucab *et al*, 2019).

However, despite rapid advances in decoding signature etiologies in bottom-up and top-down approaches, many signature etiologies remain unknown (Table 1). Most bottom-up experimental approaches have focused on defining mutational signature generation of clearly defined single exposures. However, studies of multiple perturbations and studies in relevant cell type specific or *in-vivo* models are needed to define which signatures are primary or secondary in a specific perturbation and tissue context. Understanding the complex map of DNA damage and repair process interactions during mutagenesis and cancer development would allow to leverage the available sequencing data for patient benefit. Expanding the space of mutation classes to analyze or finding new signatures in large datasets is not enough. Accurate biological interpretation of signature etiologies and mechanisms is needed to find therapeutically actionable vulnerabilities and define clinically relevant sub-groups within patient populations. Only then, we can hope to best utilize the treatments already available and move toward more personalized and more effective cancer treatment and prevention.

CHAPTER 2: RESULTS

Aims of the Thesis

The mathematical framework of mutational signatures, coupled with the availability of many cancer genomes, has enabled rapid elucidation of mutational signatures of distinct mutational processes. The exploration of the mutational landscape has yielded a catalogue of well-established signatures (COSMIC database) and uncovered cancer vulnerabilities which are clinically actionable. Still, more than half of the known signatures have an unknown etiology. Owing to the wide availability of tumor genome sequences, the development of genomic stability within tumors is more well studied than the emergence of genomic instability prior to cancer development. Especially chemical, environmental, or metabolic exposures are critical factors in determining risk of cancer development and remain understudied in how they impact genomic stability and the mutational landscape.

Understanding the molecular cause of mutational processes remains the focus of bottom-up *in-vitro* or *in-vivo* studies, yet few studies use tissue specific models to study genomic stability in response to controlled perturbations. This project aimed to utilize organoid technology to study the evolution of genomic stability in intestinal stem cells *in-vivo* in response to long term exposure to a high fat diet. Using a model of diet induced obesity, we aimed to understand how the systematic dietary and metabolic changes in obesity impact genomic stability in a tissue specific manner.

The specific aims included:

1. Establishing experimental protocols and organoid culturing techniques to clonally enrich intestinal stem cells for whole genome sequencing
2. Process and analyze the whole genome sequencing data
3. Explore the mutational landscape with mutational signature analysis
4. Training in bioinformatics methods comprising genome analysis and mutational signature analysis

Prologue

In the following I present the results of my PhD project, which is under review with Scientific Reports, titled “Mutational Landscape of Intestinal Stem Cells After Long-term In Vivo Exposure to High Fat Diet”. In this manuscript we comparatively explore the mutational landscape of intestinal stem cells derived from mice fed either a standard diet (SD) or a high fat diet (HFD). We recover single base substitution signatures SBS1, SBS5, and SBS18 and indel signatures ID1 and ID2 in equal proportions in both diet groups. All recovered signatures are attributable to normal mutational processes associated with aging, cellular replication, and oxidative or metabolic stress experienced *in-vivo* or during tissue culturing. Thus, we demonstrate that diet induced obesity alone, in the absence of other perturbations or driver gene mutations, is not sufficient to induce differential or enhanced mutational profiles that would indicate an increase in genomic instability.

This research article is under review with Scientific Reports.
Meyenberg and Hakobyan *et al*, **Mutational Landscape of Intestinal Stem Cells After Long-term In Vivo Exposure to High Fat Diet**, Scientific Reports (2022)

Research Article

Mutational Landscape of Intestinal Stem Cells After Long-term In Vivo Exposure to High Fat Diet

Mathilde Meyenberg[1,2 +], Anna Hakobyan[1,3, +], Nikolina Papac-Milicevic[4], Laura Göderle[4], Mateo Markovic[4], Ji-Hyun Lee[5,6], Bon-Kyoung Koo[5,6], Israel Tojal da Silva[7], Christoph J. Binder[4], Jörg Menche[1,3,8]*, and Joanna I. Loizou[1,2*]

[1]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, 1090, Austria
[2]Center for Cancer Research, Comprehensive Cancer Centre, Medical University of Vienna, 1090 Vienna, Austria [3]Department of Structural and Computational Biology, Max Perutz Labs, University of Vienna, 1030 Vienna, Austria
[4]Department of Laboratory Medicine, Medical University of Vienna, Vienna, 1090, Austria
[5]Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna BioCenter (VBC), Dr. Bohr-Gasse 3, 1030 Vienna, Austria.
[6]Center for Genome Engineering, Institute for Basic Science, 55, Expo-ro, Yuseong-gu, Daejeon, 34126, Republic of Korea
[7]Laboratory of Computational Biology and Bioinformatics, A.C. Camargo Cancer Center, São Paulo, 01508-010, Brazil
[8]Faculty of Mathematics, University of Vienna, 1090 Vienna, Austria
*joerg.menche@univie.ac.at
*joanna_loizou@hotmail.com
+these authors contributed equally to this work

ABSTRACT

Obesity is a modifiable risk factor in cancer development, especially for gastrointestinal cancer. While the etiology of colorectal cancer is well characterized by the adenoma-carcinoma sequence, it remains unclear how obesity influences colorectal cancer development. Dietary components of a high fat diet along with obesity have been shown to modulate the cancer risk by perturbing the homeostasis of intestinal stem cells, yet how adiposity impacts the development of genomic instability has not been studied. Mutational signatures are a powerful way to understand how a complex biological response impacts genomic stability. We utilized a mouse model of diet-induced obesity to study the mutational landscape of intestinal stem cells after a 48-week exposure to an experimental high fat diet *in vivo.* By clonally enriching single stem cells in organoid culture and obtaining whole genome sequences, we analyzed and compared the mutational landscape of intestinal stem cells from normal diet and high fat diet mice. Single nucleotide substitution signatures and indel signatures present in our cohort are found equally active in both diet groups and reflect biological processes of normal aging, cellular replication, and oxidative stress induced during organoid culturing. Thus, we demonstrate that in the absence of activating mutations or chemical exposure, high fat diet alone is not sufficient to increase genomic instability.

Keywords: mutational signatures, obesity, intestinal stem cells, intestinal organoids, high fat diet

Introduction

Global obesity rates have been steadily increasing for the past forty years[1]. Obesity is accompanied by many comorbidities such as increased likelihood of type II diabetes, hypertension, and nonalcoholic fatty liver disease[1,2]. Among the biggest health impacts is the increase in cancer risk which accompanies body fat accumulation[3–6]. The International Agency of Research on Cancer (IARC) has recognized the overwhelming epidemiological evidence which links the chronic obese condition with increased cancer risk, in particular for organs along the gastro-intestinal axis[7]. Especially the risk of developing colorectal cancer (CRC) is highly influenced by dietary risk factors and high body mass index (BMI)[8]. With the clear association between high BMI and CRC risk, gaining understanding of the underpinning disease etiology could inform preventative as well as therapeutic programs.

Colorectal cancer development is defined by a well described progression of mutations, known as the adenoma-carcinoma sequence[9]. Deactivating mutations in adenomatous polyposis coli (*APC*) are initiating mutations, leading to constitutive Wnt/β-catenin signaling. Colorectal cancer develops through three different molecular pathways, the chromosomal instability pathway (CIN), the microsatellite instability pathway (MSI), and the CpG island methylation pathway (CIMP)[10]. Although the development of CRC is heterogeneous and sometimes involves overlapping pathways, all three pathways are defined by genomic instability which enables the acquisition of further mutations in a set of tumor suppressor and oncogenes, including *KRAS* and *BRAF* (often mutually exclusive), *TP53, PIK3CA*, and *SMAD4*[10,11]. Interestingly, it was shown that concomitant loss of APC and p53 is sufficient to induce high levels of chromosomal instability, characteristic for the CIN pathway[12].

Despite well-defined molecular genetics in CRC development, it remains unclear how a high fat diet (HFD) impacts this series of events.

With the advent of advanced tissue culturing techniques, it has become possible to study the most relevant cell populations *in vitro*[13]. In the case of CRC, the cell population of origin are the rapidly cycling LGR5 positive (leucine rich repeat containing G protein coupled receptor 5) intestinal stem cells (ISCs), residing at the bottom of the crypt[14]. These cells have been demonstrated to be sensitive to dietary and metabolic perturbation, modulating the risk of cancer initiation[15–18]. A prolonged exposure to HFD constituents has been shown to confer stemness features on non-stem cell progenitors, thus increasing the pool of actively replicating cells[16,19]. The HFD component palmitic acid was found to initiate this effect via the activation of *PPAR-∂* (peroxisome proliferator-activated receptor delta) signaling, which induces canonical Wnt-signaling[16,19]. Another prominent metabolite commonly associated with diet induced obesity is cholesterol. Extended exposure to high cholesterol levels were found to also drive proliferation of ISCs and increase the rate of tumorigenesis in an *APC* deficient background[17].

Although it has been demonstrated that a HFD directly modulates signaling in the stem cell niche, the effect on genomic stability has not been studied yet. Beyond describing mutations in individual genes, mutational signatures offer a framework to systematically study how genomic instability arises in cancer development. Mutational signatures are a mathematical framework that allows to define patterns of mutations within their sequence context. The specific mutational imprint of a signature on the genome is the reflection of the dysregulation or dysfunction of DNA damage and repair pathways and other biological processes[20]. Since the conception of mutational signatures in 2013[21,22], it has become possible to investigate cancer genomes at a global level and capture patterns which describe complex underlying biological mechanisms. Bottom-up *in vitro* studies, measuring the mutagenic effect of an exposure or gene knockout, have proven to be especially useful in defining signature etiologies[23–26].

Here, we investigated whether exposure to prolonged high fat diet generates distinct mutational processes in intestinal stem cells. Because mutational signatures effectively capture biologically complex processes, they serve as a good readout for studying effects on genomic stability. We sequenced and analyzed clonal intestinal organoids derived from mice which were fed an experimental HFD for 48 weeks. After data processing and variant calling, we obtained sufficient numbers of single base substitution (SBS) and indel (ID) mutations to investigate SBS and ID signatures, as well as coding mutations. For both diet groups, we recover expected signatures related to aging, tissue culture processing, and cellular replication. We demonstrate that differential mutagenesis is not initiated by HFD alone in the absence of other disturbance events, such as chemical exposure or mutations in CRC driver genes.

Results

Mouse Model of Dietary Induced Obesity

To study the long-term effect of obesity on genomic stability in intestinal stem cells, we set up a cohort of age matched male C57/BL6J mice (Fig 1A). After random assignment to cages with either standard chow (SD) or HFD, the mice were started on the respective diet course at the age of 5 weeks for 48 continuous weeks. At set time intervals of 6, 12, 28, and 48 weeks, a random subsample of HFD and SD mice was drawn and sacrificed to harvest ISCs for culturing. Organoids were picked and cultured to clonality before obtaining whole genome sequences (30x) for 5 obese and 5 lean mice from the last timepoint (48 weeks). For each mouse, 4 independent organoid clones and the matched tail were sequenced to distinguish acquired variants from germline variants.

Our model of diet induced obesity relies on the choice of supplied diet. In the high fat diet condition, mice derive 60% of all calories from fat, while the majority of calories in the normal diet (SD) derive from carbohydrates (55.5%) (Fig. 1B). The exact diet composition is described in supplementary table 1 and 2. Despite lower overall food consumption in the HFD group (Fig. 1C), mean weekly caloric intake was higher in the HFD group (92.7 kcal/week) compared to the SD group (80.2 kcal/week). C57BL/6J mice have been well characterized as model organisms for diet induced obesity, capturing essential aspects of metabolic dysregulation and weight gain[27,28]. Our cohort also exhibited the marked increase in total cholesterol upon exposure to HFD (Fig. 1D) and a significant increase in body weight after 3 weeks on the HFD (Fig 1E), recapitulating the metabolic dysregulations and resulting phenotype associated with obesity.

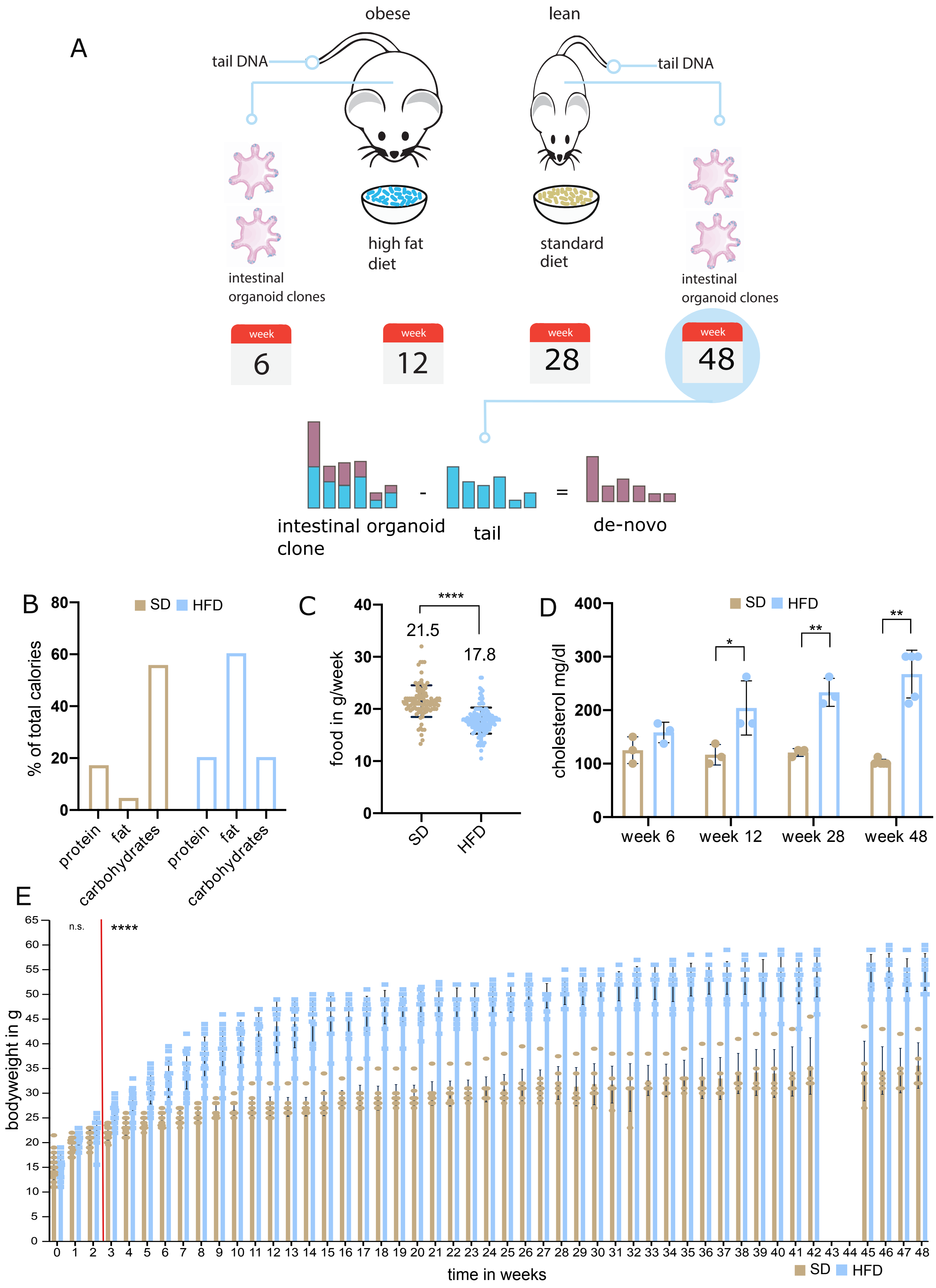


Figure 1. (A) Schematic display of experimental workflow. **(B)** Macronutrients of experimental diets shown by percent contribution to total calories. HFD is shown in light blue and SD is shown in light brown. (**C**) Food consumption per diet group,

measured per cage and divided by the number of mice per cage. The group average and statistical significance is indicated above (unpaired t-test, two-tailed) (**D)** Plasma cholesterol content in mg/dl shown per diet group at each point of the time course. N=3 for each group at timepoints week 6, 12, 28 and N=5 at week 48 (pairwise t-test, two-tailed). **(E)** Weekly weight measurements for diet groups. Dots indicate measurements for individual mice. Statistically significant weight gain was observed after 3 weeks on the HFD (indicated by red line) Statistical significance was tested using multiple unpaired t-tests with alpha = 0.001 (Holm-Sidak correction method for multiple testing, not assuming consistent standard deviation between groups).

Qualitative Analysis of Mutational Profiles in SD and HFD fed Mice

Since the genome records past and ongoing mutational processes, we reasoned that longer exposure to HFD would result in a stronger signal. Hence, we focused our sequencing efforts on the last time point (48 weeks). The obtained raw reads were processed according to GATK (The Genome Analysis Toolkit) best practices (Fig 2A). To obtain a high confidence set of mutations, we utilized two mutation callers, Mutect2[29,30] and Strelka2[31]. Mutations which were found by both Mutect2 and Strelka2 and passed the respective quality filter settings were included for further analysis. This yielded a total of 48,742 single nucleotide variants (SNV), 165 double nucleotide substitutions (DSB), and 6662 indels (insertions and deletions). Due to low numbers of mutations, DSBs were excluded from further analysis. As an additional quality control step, we checked the variant allele frequency (VAF) distribution of all organoid clones and included only clonal samples, where the VAF distribution is centered around 0.5 (Supplementary Figure 1).

We first explored the overall mutational landscape for SNVs per diet group. Surprisingly, we found a slightly higher number of total mutations in the SD group than in the HFD group. We observed a significantly higher number of mutations in the SD group for C>G, C>T outside of CpG regions, T>C, and T>G (Fig 2B). The profile of relative contributions, across the 7 mutation channels, however, is similar between two diet groups. Next, we examined the mutational profiles in 96 channels. The mean mutational profile per diet group exhibits few characteristic peaks, with the exception in the C>A and C>T components. The aggregated profile of the HFD group has a cosine similarity of 0.9929 to the SD group (Fig 2C). We furthermore observe highly similar profiles between mice of either diet group (Supplementary Figure 2A, B). To quantify how similar the mutational profiles of samples across diet groups are, we computed the pairwise cosine similarity between all samples, which ranges from 0.9020 to 0.9776 (mean = 0.9558) (Supplementary Figure 2C).

The high cosine similarity between all samples implies the absence of strong differential mutational processes. To test this, we used a bootstrap resampling method of the 96-channel mutation matrix, adapted from Zou *et al.*, for SD and HFD samples[24]. This allows us to detect potential qualitative differences in mutational profiles which remain uncovered due to low sample size and high signal to noise ratio. The global bootstrapped mutational profile of the SD mice has a cosine similarity of 0.9933 when compared to the profile of the HFD mice (Supplementary Figure 2D).

In summary, this suggests that no strong qualitative differences exist for mutagenic processes in either diet group.

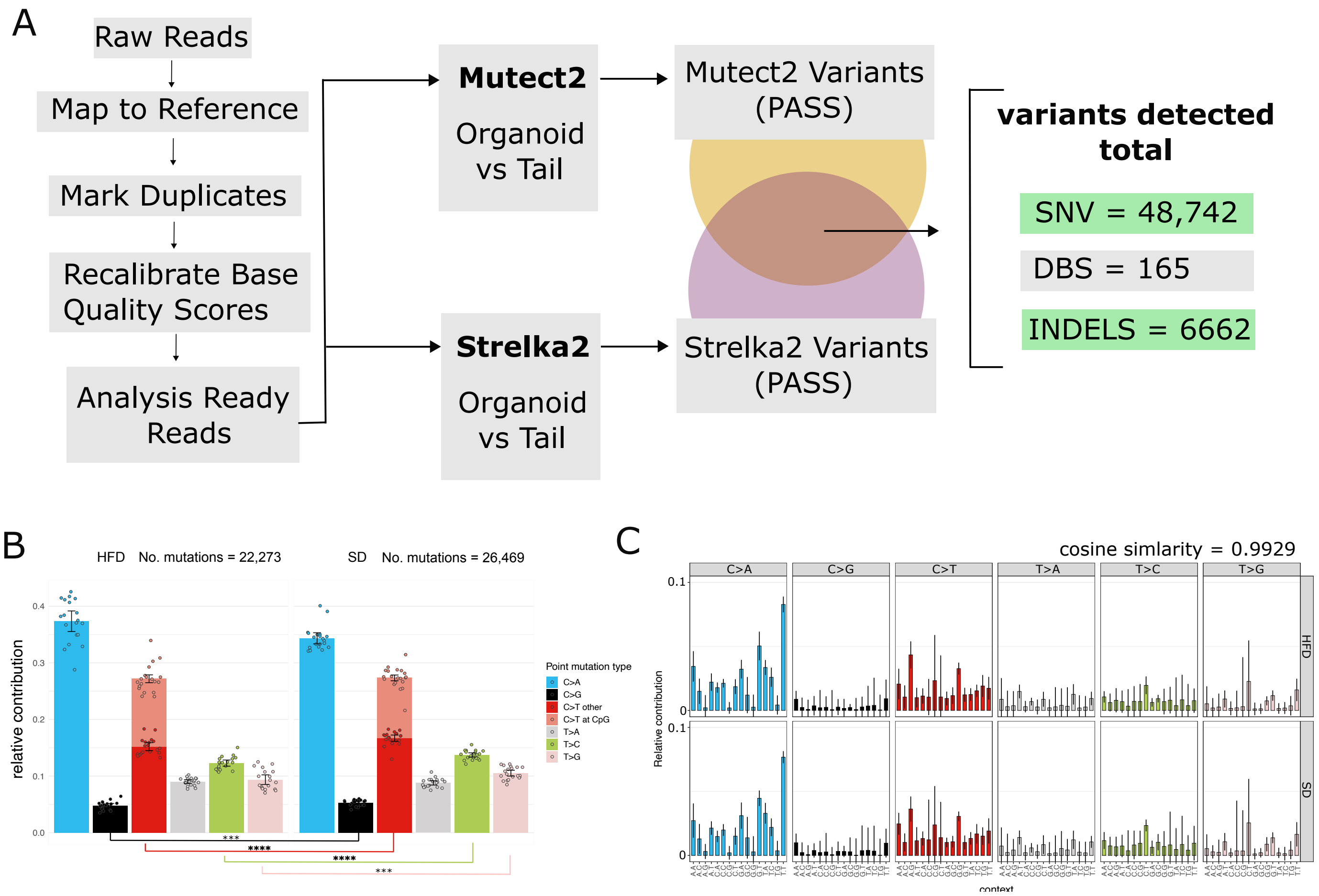


Figure 2. (**A**) Raw reads from paired end 150 bp Illumina sequencing were processed according to GATK best practices, including marking and removal of duplicates and recalibration of base quality scores. Analysis ready reads were processed by two mutation callers, Mutect2 and Strelka2. Variants called by both tools were included in the analysis. In total, 48,742 single nucleotide variants, 165 double base substitution variants, and 6,662 insertions and deletions could be detected. **(B)** Relative contribution of SNVs in six mutation classes for HFD samples (left panel) and SD samples (right panel). C>T mutations within CpG sites are shown as a separate category. Individual dots indicate organoid samples, error bars show ±1 sd from the mean, asterisks indicate results from pairwise t-test (two-sided) comparing mutation numbers for each mutation category, alpha = 0.05 **(C)** Average mutational profile of SNVs in 96 channels shown for HFD (upper panel) and SD (lower panel). Error bars indicate ±1 sd.

Mutational Signature Analysis of Single Nucleotide Variant Profiles

De-Novo Signature Extraction of SNV Signatures

Despite the lack of qualitative differences in the mutational profiles of the two diet groups, we next sought to explore which mutational signatures are active in the diet groups to determine whether quantitative differences exist. We first employed non-negative matrix factorization (NMF) with automated rank selection based on the NMFk method to determine the optimal number of *de-novo* signatures to extract [32,33]. Classically, NMF algorithms use heuristics to determine the optimal rank based on either stability of the solution, or on automatic relevance determination (ARD), which is a measure of precision of the chosen model to explain the data[33]. In contrast, the NMFk method for automatic rank determination seeks to optimize the tradeoff between both, the stability of the solution and the accuracy of the reconstructed data, measured as the distance between original and reconstructed profiles (mean sample cosine distance). This method allows to robustly extract a meaningful number of signatures from noisy data while minimizing the number of false positive signatures[32].

Applied to our data, both stability and mean sample cosine distance decline the more signatures are extracted. Thus, the most optimal solution consists of extracting a single signature (Fig 3A). The presence of a single consensus profile in the cohort would indicate that there are no distinguishing signatures between diet groups. The *de-novo* extracted signature can furthermore be decomposed into known signatures from the catalog of somatic mutations in cancer (COSMIC) database[34]. According to this decomposition, the *de-novo* signature consists of 48.1% SBS5, 42.56% SBS18, and 9.34% SBS1 (Fig 3A).

Comparing the per sample contribution of the decomposed signatures reveals an equal distribution of signature activities across samples, regardless of diet used (Fig 3B).

Signature Refitting of SNV Signatures

In cohorts with lower sample numbers such as ours, an alternative approach to *de-novo* extraction is signature refitting, where the mutational catalog of the samples is fitted to the catalog of known signatures (COSMIC) to find a subset which best explains the observed mutational catalog. This approach takes a defined set of known signatures and performs a refit in an iterative manner. After each iteration the reconstructed and original profile are compared and the lowest-contributing signature is eliminated from the set. Signatures will stop being removed when the cosine similarity between the reconstructed and original profile between two iterations has changed more than a given threshold. Thus, only signatures which are necessary to model the observed data are retained in the set. By repeating this process n-1 times for all sets of n-1, n-2, n-3 *etc*., where n is the total number of known signatures, we find that SBS1, SBS5, SBS18, and SBS40 explain 99.6% of observed mutations in both diet groups (Fig 3C). Only four samples showed minimal activity of SBS15 (defective mismatch repair), a signature directly attributable to an increase in genomic instability. The equally minimal number of mutations attributed to SBS15 in both diet groups, however, suggest no differential potential in mismatch repair among SD and HFD. In summary, the distribution of the fitted signatures is highly similar across samples and is not diet specific.

To quantify the activity of the most active signatures, we applied the signature presence test from the mSigAct package[35,36] to SBS1, SBS5, SBS18, and SBS40. This statistical test builds two refit models, one including and one excluding the signature of interest, while minimizing the reconstruction error. Following this, the likelihood ratio test between the two models is computed. For ratios greater than 1, the likelihood of the signature being active is significantly higher than the alternative hypothesis. The signature presence test confirmed the results obtained from signature refitting. Of the 4 tested signatures, SBS5 is the least active, as already observed before (Fig 3D). Although some variation in signature activity can be observed between individual samples, no signature shows a systematic difference between diets.

All signatures we found are equally active in both diet groups and are likely attributable to normal aging processes and the culturing process prior to sequencing. SBS1 is a clock-like signature which is attributed to aging due to spontaneous deamination of 5-methylcytosines, which leads to a C>T transition[21]. The activity of SBS1 observed in both groups thus likely reflects the normal aging process. Additionally, both groups showed high numbers of C>A mutations, which were largely attributed to SBS18. This signature has been proposed to be caused by damage due to reactive oxygen species[25,37] and might thus have arisen during the routine experimental handling of the samples or due to exposure to metabolic byproducts in the intestine. The remaining signatures SBS5 and SBS40 share similarly flat profiles. Although only SBS5 has been clearly identified as a clock-like signature, SBS40 was also found to correlate with age[22,38]. Thus, the activity of both signatures may be explained by normal aging processes. Taken together, the results from *de-novo* extraction and signature refitting, confirm that the experimental HFD did not induce or impact different mutational processes for single nucleotide substitutions compared to the standard diet.

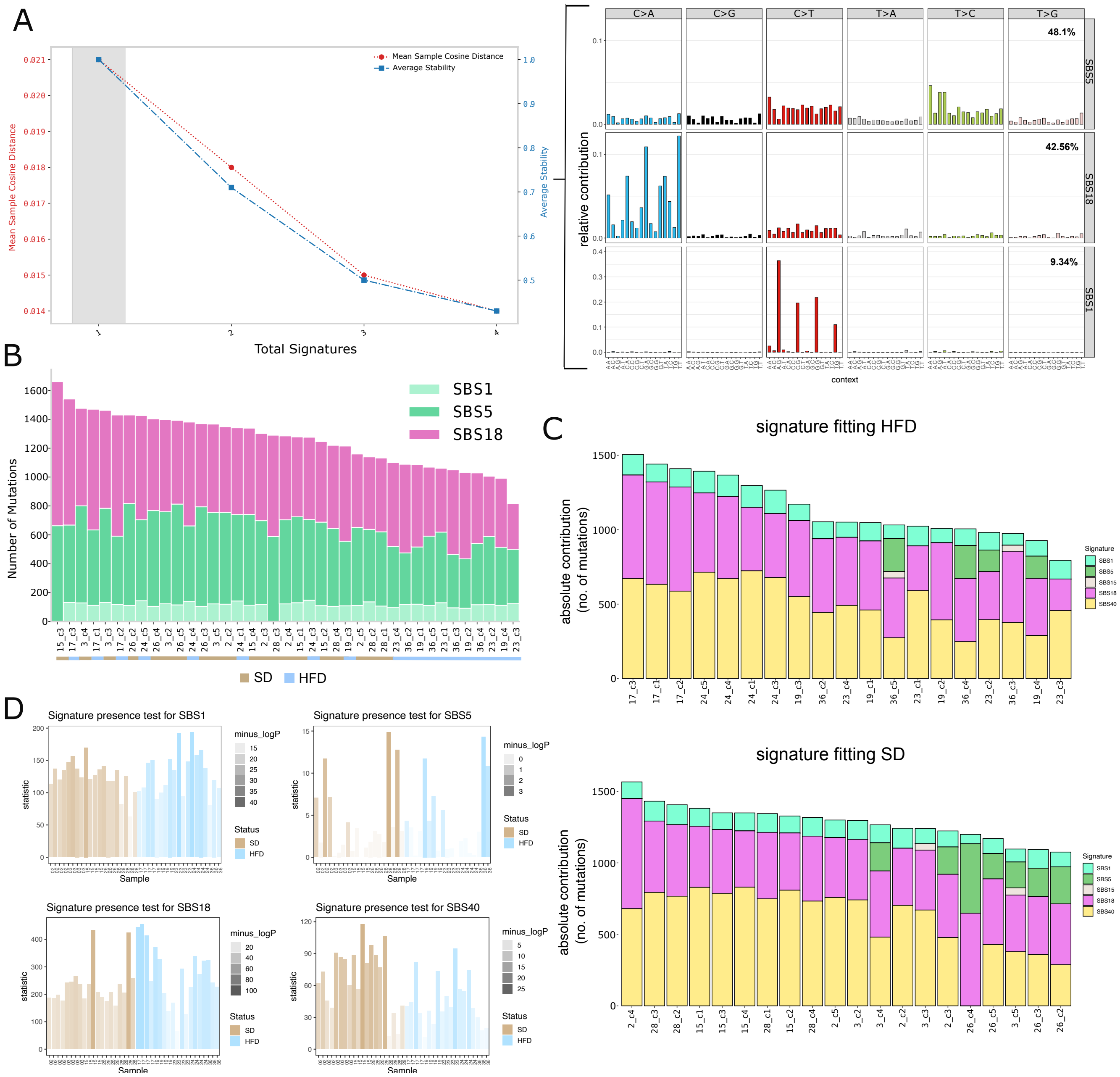


Figure 3. (**A**) NMF for signature extraction ranging from 1-4 signatures. Red line indicates mean sample cosine distance (MSCD), blue line indicates average stability (AS), gray bar indicates preferred solution, maximizing the tradeoff between MSCD and AS. The decomposition of the extracted signature into known COSMIC signatures and their calculated percent contribution is shown to the right. **(B)** NMF results from A shown as per sample absolute signature contributions (number of mutations), diet status of the samples is indicated at the bottom. **(C)** Best subset signature refitting using signatures commonly active in colorectal cancer. Per sample absolute signature contributions (number of mutations) are shown for HFD samples (upper panel) and SD samples (lower panel). **(D)** Signature Presence test for 4 most active signatures. The y-axis indicates the likelihood ratio between the signature fitting with and without the tested signature. The translucence of the bars, shown for individual organoid clones, is indicative of the level of significance (-log p).

Mutational Signature Analysis of Indel Profiles

Comparison of Indel Profiles between Diet Groups

Aside from SNVs, numerous mutational processes also generate insertions and deletions. This class of mutations generates signatures different to SNV signatures. We therefore analyzed the 6662 indel mutations in our cohort to compare whether differences in indel generating mutational processes exist between the diet groups. We only considered clonal samples with a VAF distribution centered around 0.5 (Supplementary Figure 3). Indel mutations can be analyzed in 16 or in 83 curated channels, representing the main and extended sequence context respectively[39]. The curated indel types range from a single base pair deletion or insertion, up to indels longer than 5bp. Additionally, 1-5 bp deletions flanked by microhomologies are considered, since such mutations are indicative of defective double strand break repair processes[40]. Indel profiles in both sequence contexts were highly similar between diet groups, with a cosine similarity of 0.9925 for main indel contexts (Fig 4A), and 0.9941 for extended indel contexts (Fig 4B). Only 5+bp deletions flanked by microhomologies were significantly increased in the SD compared to the HFD cohort (Fig 4A). However, since the total number of mutations in that category is less than 10, this likely represents a random variation and carries no specific biological meaning. Indeed, all mice, regardless of diet group, exhibited highly similar indel profiles, both for the main and the extended sequence context (Supplementary figure 4A-D). Furthermore, all samples showed a pairwise cosine similarity greater than 0.84 (Supplementary Figure 4E). Conclusively, the high cosine similarity between indel profiles of the diet groups as well as among individual samples suggest that no indel generating processes are unique to either diet.

De-Novo Signature Extraction of Indel Signatures

We next applied the same analysis workflow we established for SNV signatures to all insertions and deletions. NMF with automated rank selection, found one indel signature as the optimal solution because extraction of more than one signature led to sharp decrease in average stability (Fig 4C left panel). The decomposition of the single *de-novo* signature estimated three known COSMIC signatures to be active, ID1 (22.46%), ID2 (40.88%), and ID12 (36.66%) (Fig 4C right panel). The distribution of the signature contribution to the individual samples does not differ between diet groups (Fig 4D).

Signature Refitting for Indel Signatures

Exploring indel signatures further with refitting analysis allowed us to confirm the results obtained from *de-novo* extraction. Using best subset refitting with all 18 known indel signatures, we find ID1, ID2, and ID12 most active and similarly distributed across samples (Fig 4E). Minor activity observed for ID7 (MMR deficiency[38]) and ID9 (etiology unknown[38]), may be due to signature misattribution for the common C and T deletions found in our cohort. Since the low number of mutations may be limiting in this analysis, we also pooled the mutational matrix of each diet group and performed a best subset refit to all COSMIC indel signatures. The results show an equal distribution of ID1, ID2, and ID12 activity across diet groups (Supplementary figure 4F). To confirm the stability of the refitting we performed bootstrapped refitting. The mutational matrix is resampled 1000 times with replacement, using the original mutational profile as weight. For each bootstrap iteration a refit is calculated, recording estimated signature activity. The higher the consensus of refits across bootstrap iterations, the more stable the refit. The results confirm that ID1, ID2, and ID12 are the most active signatures in our cohort, regardless of diet consumed. ID7 was found active only in the SD group and was attributed less than 10% of all mutations in that group (Fig 4F). Finally, we quantified the signature activity of ID1, ID2, ID7, and ID12 for all samples, using the signature presence test (Fig 4G). The results confirm that ID2, and ID12 (etiology unknown[38]) are the most active signatures in both diet groups, since the majority of mutations is attributed to these signatures across all samples. ID1 is the third most active indel signature, followed by ID7, which is active only in some samples and completely absent in 23% of all samples.

None of the identified indel signatures are differentially active between tested diets. Signatures ID1 and ID2 are both proposed to arise due to slippage of the replicated (ID1), and template strand (ID2) during replication, producing the characteristic 5+bp T-insertions and 6+bp T-deletions. These signatures have been observed to be active in all samples and are only increased in backgrounds with mismatch repair deficiency (MMR)[38]. In our cohort, we have not observed a strong activity of either SNV or indel signatures associated with defective mismatch repair. The low activity of MMR deficiency signature ID7 in some samples may partially explain the high activity observed for ID1 and ID2. However, this process is equally active in both diet groups (Fig 4E, G). Notably, ID1 and ID2 activity were found increased in conditions of chronic inflammation of the intestinal tract in patients[41]. Even though obesity is associated with changes in metabolic and hormonal signaling associated with forming an inflammatory environment[6], we do not observe an increase in ID1 and ID2 activity that would indicate strong changes in inflammatory signaling. Thus, the activity of ID1 and ID2 we find in both diet groups suggest mutational processes ongoing during normal cellular replication. In summary, the results indicate that the experimental HFD did not invoke or influence mutational processes of indel generation.

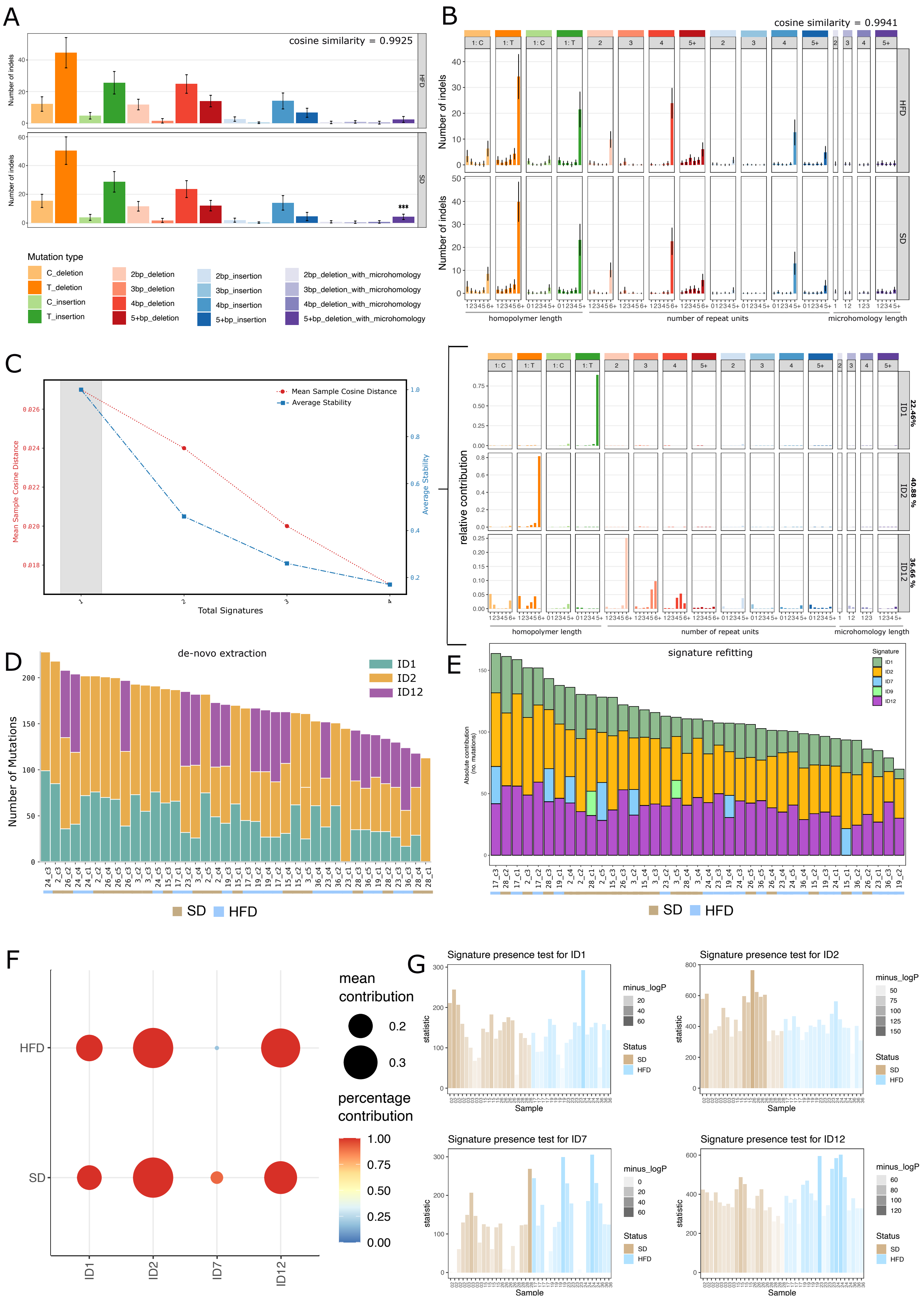


Figure 4. **(A)** Main indel contexts (16-channels) profiles aggregated by diet group (mean), error bars indicate ±1 sd. Statistical significance was assessed using

multiple pairwise t-tests, not assuming consistent standard deviation (Holm-Sidak correction method for multiple testing, alpha = 0.01) **(B)** Mean extended indel profile by diet group (83 channels), error bars indicate ±1 sd from the mean. **(C)** NMF diagnostic plot for signature extraction ranging from 1-4 signatures. Red line indicates mean sample cosine distance (MSCD), blue line indicates average stability (AS), gray bar indicates preferred solution, maximizing the tradeoff between MSCD and AS. The decomposition of the extracted signature into known COSMIC indel signatures and their calculated percent contribution is shown to the right. **(D)** NMF results from A shown as per sample absolute signature contributions (number of mutations), the diet status of the samples is indicated on the x-axis. **(E)** Best subset signature refitting using all 18 known indel signatures. Per sample absolute signature contributions (number of mutations) are shown. Diet status is indicated on the x-axis. **(F)** Bootstrapped refitting of indel signatures (best subset approach using all known indel signatures, 100 iterations). Size of dots indicates the mean contribution of the signature for all bootstrap iterations where this signature was found. The color scale represents the percentage of bootstrap iterations where the signature was found active. **(G)** Signature Presence test for 4 most active signatures found in refitting and bootstrapped refitting. The x-axis indicates the sample, the y-axis indicates the likelihood ratio between the signature fitting with and without the tested signature. The translucence of the bars, shown for individual organoid clones, is indicative of the level of significance (-log p).

Coding Mutations

Finally, we wondered whether the absence of specific mutational processes also precluded the accumulation of specific deleterious mutations which might initiate the adenoma-carcinoma sequence and thus predispose to tumor development. To test this, we explored the coding mutations which accumulated in either diet group. Due to low numbers of coding mutations in our cohort, we included all mutations which passed the filtering criteria from the Mutect2 variant caller. Filtering for the most mutated genes revealed a remarkable overlap, 9 out of the top 10 most mutated genes are shared between the SD and HFD group (Fig 5 A-B). The largest fraction of alterations are missense mutations. None of the mutated genes have a known role in intestinal cancer development. Taken together, we found no specific mutations which might explain how obesity increases risk of cancer development in the intestinal tract.

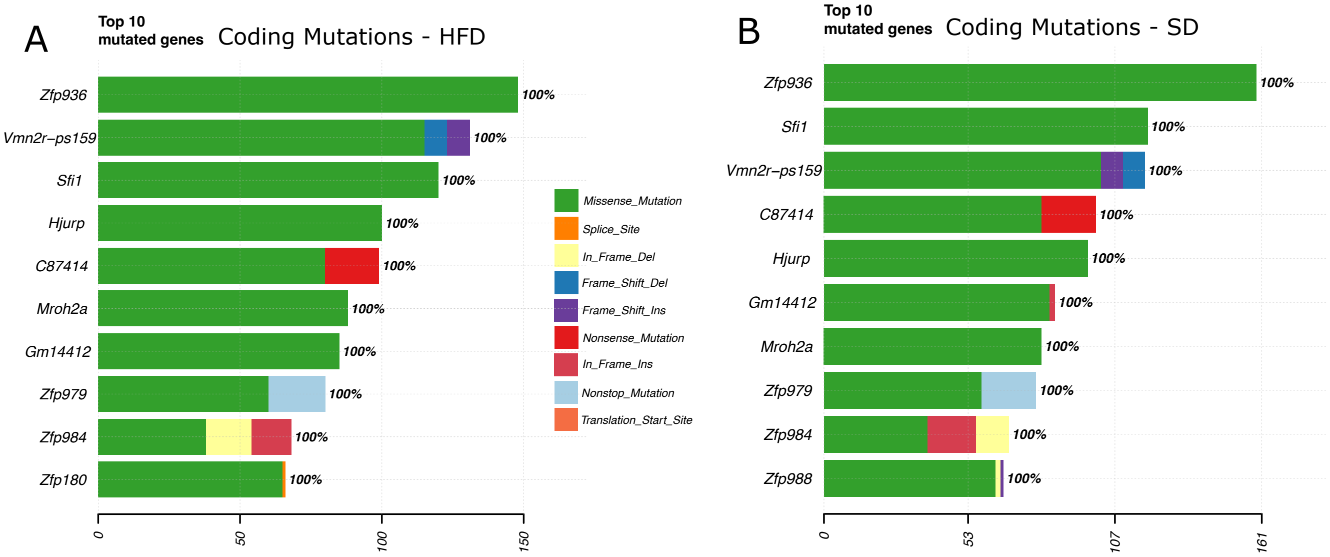


Figure 5. (A) Top 10 coding mutations in HFD (B) and SD mice. The type of substitution is indicated by color. The percentage of samples with a mutation in the gene is indicated to the right.

Discussion

Obesity is a chronic disease which epidemiologically has been shown to increase the risk of developing cancer in the intestinal tract[3–7]. High BMI and dietary factors such as consumption of a western style diet high in fats have been demonstrated to have a positive association with CRC risk through modulation of signaling in the intestinal stem cell niche[15–19]. However, the effect of obesity on the genomic stability of intestinal stem cells has not been investigated yet. We hypothesized that chronic exposure to a HFD impacts on DNA damage and repair signaling or associated processes and thus shapes the landscape of genomic stability in intestinal stem cells. To investigate the mutational landscape in response to diet induced obesity we used whole genome sequencing on clonal organoid populations derived from intestinal stem cells of mice exposed to experimental high fat or control diets, respectively.

Our results show that HFD alone, on an isogenic background, and in the absence of other predisposing mutations, does not induce differential mutational signatures compared to a standard control diet. All mutational signatures we recovered are equally active in both diet groups and represent normal ongoing mutational processes associated with aging (SBS1, SBS5), cellular replication (ID1, ID2), or oxidative stress experienced either *in-vivo* or during the culturing process during sample preparation (SBS18). Overall, signatures we recover are in agreement with previous findings, reporting the activity of SBS1, 5, and 18, as well as ID1 and ID2 as normal aging and metabolism associated processes in colonic crypts[41]. Other signatures recovered were SBS40 and ID12, which are both signatures with unknown etiology and were active in both diet groups. We furthermore found no coding mutations in common CRC driver genes or other genes associated with developing genomic instability. Thus, in the absence of any other predisposing mutations, diet-induced obesity associated alterations to any molecular pathways in the stem cell niche are not enough to generate an excess of mutations, specific mutational patterns or coding mutations that would predispose to cancer. The lack of mutagenesis in the HFD condition, both in terms of numbers of mutations and mutational signatures, would suggest that the DNA repair machinery is working efficiently in the diet-induced obesity condition, ensuring genomic stability. How a HFD impacts on the mutagenesis in a cancer predisposed background (e.g. ApcMin) or in the presence of DNA damaging agents remains to be investigated.

In the face of convincing epidemiological evidence, it remains important to understand how a western style diet modulates cancer risk in the gastrointestinal tract. By investigating mutagenesis in intestinal stem cells upon long-term exposure to a high fat diet *in vivo*, we show that HFD alone, in the absence of other perturbation events, such as mutations or chemical exposure, is not sufficient to initiate specific mutational patterns. Our results might lead to future studies to investigate combinatorial effects of HFD with other perturbations, to continue elucidating the etiology of obesity induced cancers.

Methods

Mouse work

All experimental protocols were approved by the institutional animal experimentation committee of the Medical University of Vienna and the Austrian Ministry of Science under ethical permit number 66.009/0179-V/3b/2019. All methods were performed in accordance with relevant guidelines and regulations. The study was reported in accordance with the ARRIVE guidelines. Experimental mice were age-matched males on a C57BL/6J background. The total number of mice included the study was 28. Control and treatment groups (diet groups) were randomly assigned cage numbers before the addition of the experimental research diets. Researchers were not blinded to the assigned treatment.

Shortly, from 5 weeks of age, after a 1-week acclimatization period on SD, mice were fed SD or HFD for 6, 12, 28, and 48 weeks (Research diets, D12492i, rodent diet with 60 kcal% fat, for diet composition see supplementary table 1,2). Mice were housed at the Department of Biomedical Research or the Department of Laboratory Animal Science and Genetics of the Medical University of Vienna, Austria with a 12-hour dark-/light-cycle with ad libitum access to water and food. Weight gain and food consumption of experimental animals were monitored on a weekly basis. At experimental exitus mice were sacrificed after 3 hours of fasting. Blood, plasma, intestinal tissues, and intestinal crypts from the jejunum for organoid culture were isolated. Blood was collected from the vena cava with a syringe, stored in collection tubes containing EDTA, and spun down to 15 minutes at 2000g. The supernatant plasma was retrieved and snap frozen in liquid nitrogen.

Organoid Culture

Isolated tissue from jejunum was gently rinsed with ice cold PBS (20 mL, without Mg++ and Ca++) using a syringe. The intestinal tube was cut lengthwise and covered with fresh PBS (1-2 mL, without Mg++ and Ca++)). The villi were gently scraped off using a microscope coverslip. Following this, the tissue was cut into ca. 0.5 cm long pieces and added to a tube containing ice cold PBS (50 mL, without Mg++ and Ca++). The tissue pieces were washed by gently inverting the tube before collecting the tissue pieces and repeating this washing process 2 more times with fresh PBS. After washing, tissue pieces were collected and incubated in enzyme-free dissociation buffer (StemCell Catalog #100-0485) for 10 min at room temperature on a tube roller. After incubation, the tube was vigorously shaken to loosen the crypts from the remaining tissue. The resulting solution was filtered through a 70 µm cell strainer and centrifuged for 3 minutes at 1200 rpm. Supernatant was discarded and the pellet was resuspended in fresh PBS (1 mL). Multiple aliquots of 50 µL, 100 µL, and 200 µL were transferred to 1.5 mL tubes and centrifuged for 5 minutes at 500 rcf. The supernatant was removed carefully and Matrigel (20 µL) was added and mixed with the pellet. The mixture was plated into 48-well tissue culture plates (20 µL per well), the plate inverted and incubated for 5 min at 37ºC to allow the Matrigel to polymerize. Finally, the droplets were covered with 250 µL- 300 µL of WENR culturing medium (Advanced DMEM/F12, 1%Glutamax(200mM), 1% HEPES (1M), 1% Penicillin/Streptomycin, 2% B27(50x, Thermo Fischer Catalog #17504044), 0.25% n-acetyl-L-cysteine (500mM), 0.05% Recombinant Murine EGF (500 µg/mL), 0.1% Recombinant Murine Noggin (100 µg/mL, Peprotech Catalog #250-38), 0.2%

Primocin (50 mg/mL, InVivoGen Catalog #ant-pm-05), 0.01% Y-27632 (100 mM, Adooq Bioscience Catalog #A11001-50), 1% Nicotinamide (1M), 50% Wnt3A conditioned medium as described previously, 10% R-spondin conditioned medium prepared as described previously).

After 5-7 days in culture, organoids were recovered from Matrigel and dissociated into single cells using 0.05% Trypsin-EDTA (incubation at 37ºC for 5-12 minutes) and mechanical disruption via vigorous pipetting. Single cells were plated in increasingly diluted aliquots and checked under the microscope for complete dispersion. Resulting clonal organoids were picked with a pipette after 7-10 days in culture (medium change every 2-3 days), disrupted with 0.05% Trypsin-EDTA, and cultured until enough material was available for DNA extraction.

Whole Genome Sequencing and Variant Calling

Organoids were extracted from the Matrigel by adding protease K (800 U, ~20µg), centrifuging the solution at 500 rcf for 5 minutes, and discarding the supernatant. Total DNA (~ 1 µg/sample) was extracted using a QIAamp DNA Micro Kit (Qiagen Catalog #56304). Library preparation (350 bp inserts) and sequencing (150 bp PE) on a NovaSeq6000 platform (Illumina) was carried out with Novogene, Cambridge, UK. Raw reads were processed according to GATK4 best practices recommendation for data pre-processing for variant discovery. Reads were mapped to the mm10/GRCm38 mouse reference genome. All bam files were downsampled to match the file with the lowest coverage using the DownsampleSam command from Picard tools with accuracy=0.001. Variants in organoid clones were called with Mutect2 and Strelka2, using the tail DNA as a reference. Variants with filter status PASS which were called by both tools were included in the analysis. For each sample, the variant allele frequency (VAF) distribution was plotted. All samples which did not have a distribution centered around 0.5 were excluded from further analysis.

Mutational Signature Analysis

De-novo mutational signature extraction using NMF was performed using SigprofilerExtractor[32]. Signature refitting and plotting was performed using the MutationalPatterns package in R[42]. The bootstrapping analysis of the SNV signatures was conducted as described previously[24]. Briefly, bootstrapped resampling was applied to generate 10,000 replicates of the mutational matrix for SD samples and HFD samples respectively, using the underlying distribution of signatures across the 96 channels as weight. The results were aggregated by diet group to generate an average bootstrapped mutational profile, which was then compared between groups using cosine similarity.

References

1. Jaacks, L. M. *et al.* The Obesity Transition: Stages of the global epidemic. *lancet. Diabetes Endocrinol.* **7**, 231 (2019).
2. Blüher, M. Obesity: global epidemiology and pathogenesis. *Nat. Rev. Endocrinol. 2019 155* **15**, 288–298 (2019).
3. Hopkins, B. D., Goncalves, M. D. & Cantley, L. C. Obesity and cancer mechanisms: Cancer metabolism. *Journal of Clinical Oncology* **34**, 4277–4283 (2016).
4. Calle, E. E., Rodriguez, C., Walker-Thurmond, K. & Thun, M. J. Overweight, Obesity, and Mortality from Cancer in a Prospectively Studied Cohort of U.S. Adults. *http://dx.doi.org/10.1056/NEJMoa021423* **348**, 1625–1638 (2009).
5. Friedenreich, C. M., Ryder-Burbidge, C. & McNeil, J. Physical activity, obesity and sedentary behavior in cancer etiology: epidemiologic evidence and biologic mechanisms. *Mol. Oncol.* **15**, 790–800 (2021).
6. Avgerinos, K. I., Spyrou, N., Mantzoros, C. S. & Dalamaga, M. Obesity and cancer risk: Emerging biological mechanisms and perspectives. *Metabolism: Clinical and Experimental* **92**, 121–135 (2019).
7. Lauby-Secretan, B. *et al.* Body Fatness and Cancer — Viewpoint of the IARC Working Group. *N. Engl. J. Med.* **375**, 794–798 (2016).
8. Tran, K. B. *et al.* The global burden of cancer attributable to risk factors, 2010–19: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **400**, 563–591 (2022).
9. Fearon, E. F. & Vogelstein, B. A Genetic Model for Colorectal Tumorigenesis. *Cell* **61**, 759–767 (1990).
10. Bogaert, J. & Prenen, H. Molecular genetics of colorectal cancer. *Ann. Gastroenterol.* **27**, 9 (2014).
11. Midthun, L. *et al.* Concomitant KRAS and BRAF mutations in colorectal cancer. *J. Gastrointest. Oncol.* **10**, 577 (2019).
12. Drost, J. *et al.* Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43–47 (2015).
13. Miyoshi, H. & Stappenbeck, T. S. In vitro expansion and genetic modification of gastrointestinal stem cells in spheroid culture. *Nat. Protoc.* **8**, 2471–2482 (2013).
14. Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608–611 (2009).
15. Kim, E. *et al.* Rapidly cycling Lgr5 + stem cells are exquisitely sensitive to extrinsic dietary factors that modulate colon cancer risk Official journal of the Cell Death Differentiation Association. *Cit. Cell Death Dis.* **7**, (2016).
16. Beyaz, S. *et al.* High-fat diet enhances stemness and tumorigenicity of intestinal progenitors. *Nature* **531**, 53–58 (2016).
17. Wang, B. *et al.* Phospholipid Remodeling and Cholesterol Availability Regulate Intestinal Stemness and Tumorigenesis. *Cell Stem Cell* **22**, 206-220.e4 (2018).
18. La Vecchia, S. & Sebastián, C. Metabolic pathways regulating colorectal cancer initiation and progression. *Seminars in Cell and Developmental Biology* **98**, 63–70 (2020).
19. Beyaz, S. & Yilmaz, Ö. H. Molecular Pathways: Dietary Regulation of Stemness and Tumor Initiation by the PPAR-δ Pathway. *Clin. Cancer Res.* **22**, 5636–5641 (2016).
20. Koh, G., Degasperi, A., Zou, X., Momen, S. & Nik-Zainal, S. Mutational signatures: emerging concepts, caveats and clinical applications. *Nat. Rev. Cancer* **21**, 619–637 (2021).
21. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
22. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–21 (2013).

23. Drost, J. *et al.* Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science (80-.).* **358**, 234–238 (2017).
24. Zou, X. *et al.* Validating the concept of mutational signatures with isogenic cell models. *Nat. Commun.* **9**, 1744 (2018).
25. Kucab, J. E. *et al.* A Compendium of Mutational Signatures of Environmental Agents. *Cell* (2019). doi:10.1016/j.cell.2019.03.001
26. Zou, X. *et al.* A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nat. cancer* **2**, 643 (2021).
27. Speakman, J., Hambly, C., Mitchell, S. & Król, E. Animal models of obesity. *Obes. Rev.* **8**, 55–61 (2007).
28. Collins, S., Martin, T. L., Surwit, R. S. & Robidoux, J. Genetic vulnerability to diet-induced obesity in the C57BL/6J mouse: physiological and molecular characteristics. *Physiol. Behav.* **81**, 243–248 (2004).
29. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213 (2013).
30. Broad Institute. Mutect2 – GATK. *Mutect2* (2019). Available at: https://gatk.broadinstitute.org/hc/en-us/articles/360041416112-Mutect2. (Accessed: 12th January 2023)
31. Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods 2018 158* **15**, 591–594 (2018).
32. Islam, S. M. A. *et al.* Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *Cell Genomics* **2**, 100179 (2022).
33. Nebgen, B., Vangara, R., Hombrados-Herrera, M. A., Kuksova, S. & Alexandrov, B. A neural network for determination of latent dimensionality in Nonnegative Matrix Factorization. *Mach. Learn. Sci. Technol.* 1–17 (2020). doi:10.1088/2632-2153/aba372
34. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941 (2019).
35. Ng, A. W. T. *et al.* Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia. *Sci. Transl. Med.* **9**, (2017).
36. mSigAct/mSigAct_2.2.0.pdf at v2.2.0-branch · steverozen/mSigAct. Available at: https://github.com/steverozen/mSigAct/blob/v2.2.0-branch/data-raw/mSigAct_2.2.0.pdf. (Accessed: 20th December 2022)
37. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nat. 2013 5007463* **500**, 415–421 (2013).
38. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nat. 2020 5787793* **578**, 94–101 (2020).
39. Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* **10**, (2018).
40. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet. 2014 159* **15**, 585–598 (2014).
41. Olafsson, S. *et al.* Somatic Evolution in Non-neoplastic IBD-Affected Colon. *Cell* **182**, 672-684.e11 (2020).
42. Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: Comprehensive genome-wide analysis of mutational processes. *Genome Med.* **10**, 1–11 (2018).

Acknowledgements

The authors thank the members of the Koo lab for sharing their know-how of organoid technology, and the members of Menche lab and Loizou lab for their critical revision of the manuscript. We would like to thank Dr. Núria López-Bigas and Dr. Abel Gonzalez-Perez (Institute for Research in Biomedicine, Barcelona, Spain) for helpful discussions and feedback.

Funding

MM was supported by a DOC fellowship of the Austrian Academy of Sciences (25757, awarded to MM). The Loizou lab was funded by an ERC Synergy Grant (DDREAMM grant agreement ID:855741, awarded to J.I.L.) and the Austrian Science Fund (F79 Spezialforschungsbereiche, awarded to J.I.L). CeMM is funded by the Austrian Academy of Sciences.

Author contributions statement

M.M., A.H., J.M., B.K.K., C.B. and J.I.L. conceived the experiment(s), M.M., N.P.M., L.G., and Mateo M. conducted the experiment(s), M.M., A.H., I.T.S and J.M. analyzed the results. All authors reviewed the manuscript.

Data availability statement

All code used to analyze the data and produce the figures is available on https://github.com/menchelab/hfd-mutagenesis.

Additional information

The authors declare no conflicts of interest. JIL is currently an employee of AstraZeneca

Supplementary Material

Supplementary Table 1. Composition of the experimental control diet (SD).

High Fat Diet- Total energy density: 5.21 kcal/g

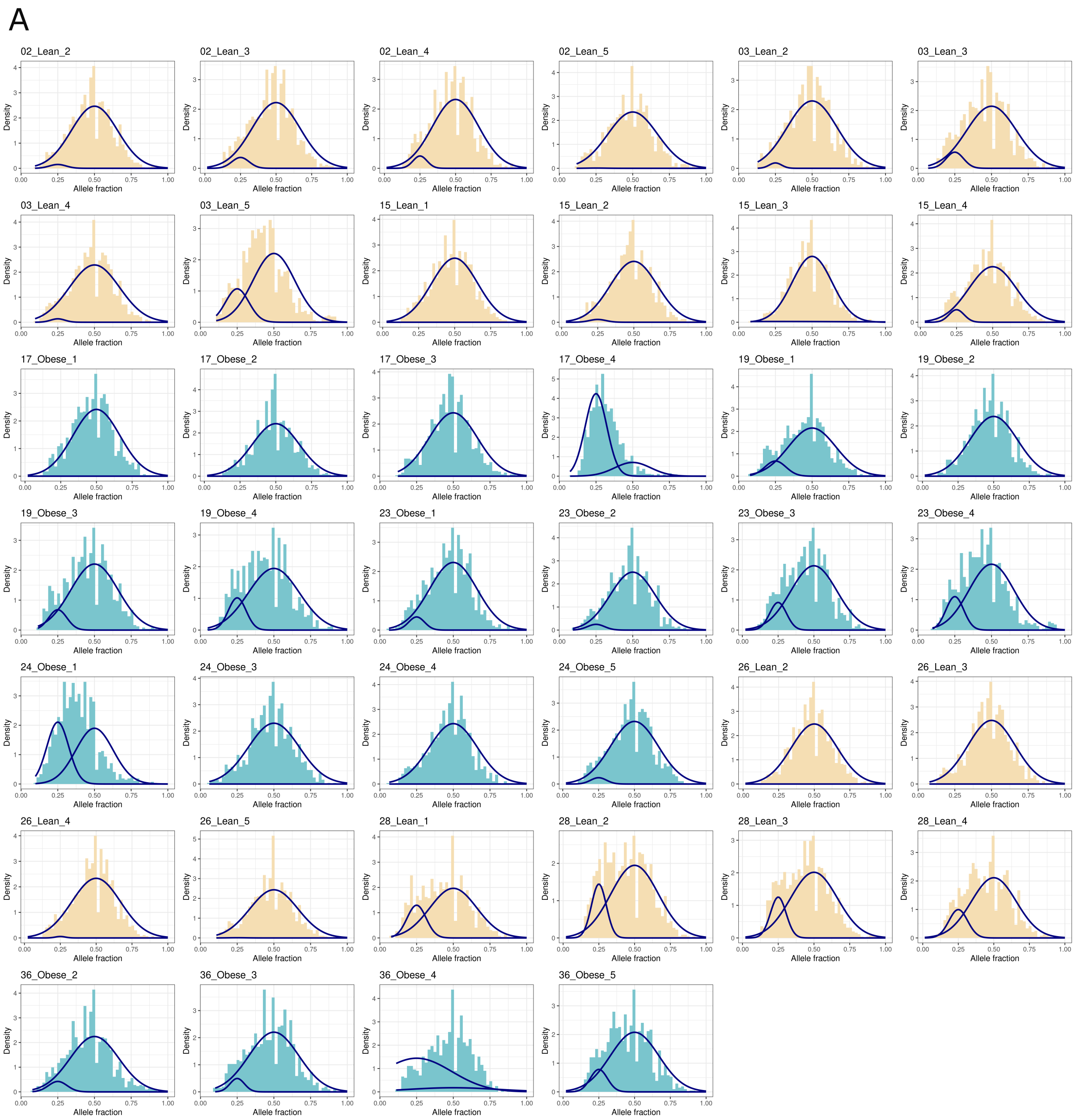
Carbo-hydrates	g/kg	Fatty Acids	g/kg	Protein	g/kg	Vitamins	g/kg	Minerals	g/kg	Other	g/kg
Lodex 10	161.53	Soybeann Oil (USP)	32.31	Casein (Lactic), 30 Mesh	258.45	Choline Bitartrate	2.58	Potassium Citrate (Monohydrate)	213.22	Solka Floc, FCC200(Fiber)	64.61
Sucrose	94.0	Lard	316.6	L-Cysteine	3.88	Vitamin E Acetate (50%)	6.46	Calcium Phosphate (Dibasic)	167.99	Blue FD&C #1, Aluminium Lake 35-42% (Dye)	0.065
						Niacin (B3)	1.94	Calcium Carbonate (light, USP)	71.07		
						Biotin (1%)	1.3	Sodium Chloride	33.47		
						Pantothenic Acid (B5)	1.03	Magnesium Sulfate (Heptahydrate)	33.29		
						Vitamin D3 (100,00 IU/gm)	0.65	Magnesium Oxide (Heavy, DC USP)	5.41		
						Vitamin B12 (0.1% Mannitol)	0.65	Ferric Citrate	2.71		
						Vitamin A Acetate (500,000 IU/gm)	0.52	Manganese Carbonate Hydrate	1.58		
						Pyridoxine HCl (B6)	0.45	Zinc Carbonate	0.72		
						Riboflavin (B2)	0.39	Chromium Potassium Sulfate	0.25		
						Thiamine HCl (B1)	0.39	Copper Carbonate	0.16		
						Folic Acid	0.13	Ammonium Molybdate Tetrahydrate	0.039		
						Menadione Sodium Bisulfite	0.054	Sodium Fluoride	0.029		
								Sodium Selenite	0.0065		
								Potassium Iodate	0.0065		

Supplier: https://researchdiets.com/formulas/d12492
Data sheet: https://researchdiets.com/formulas/d12492

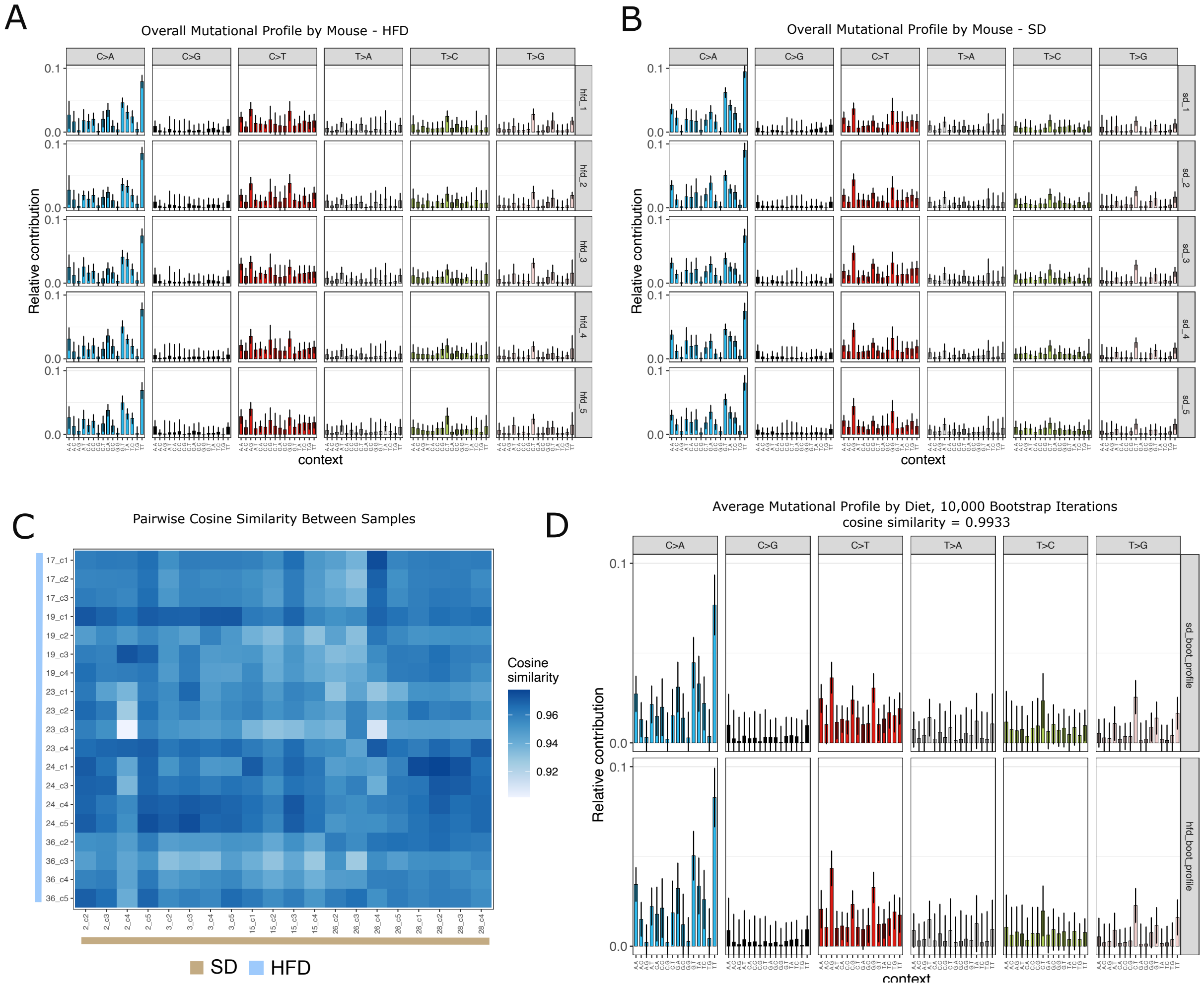
Supplementary Table 2. Composition of the experimental high fat diet (HFD).
Control Diet (Normal Diet) - Total energy density: 3.73 kcal/g

Carbo-hydrates	g/kg	Fatty Acids	g/kg	Protein	g/kg	Vitamins	Unit per kg	Minerals	g/kg	Other	g/kg
N-free extract	550	C16:0	5.0	Arginine	8.0	Vitamin A	15,000 IU	Calcium	0.01	Ash	70
		C18:0	2.0	Cysteine	3.5	Vitamin D3	1,200 IU	Phosphorus	0.0065	Fiber	43
		C20:0	0.1	Histidine	4.0	Vitamin E	0.09	Sodium	0.003		
		C18:1	9	Isoleucine	6.5	Vitamin K	0.005	Magnesium	0.0025		
		C18:2	19	Leucine	17.0	Thiamine (B1)	0.015	Iron	0.2		
		C18:3	7.5	Lysine	8.0	Riboflavin (B2)	0.01	Iodine	0.004		
				Methionine	4.0	Pyridoxine (B6)	0.01	Copper	0.015		
				Phenylalanine	8.5	Cobalamine (B12)	0.05	Cobalt	0.0015		
				Threonine	6.0	Biotin	0.2	Manganese	0.12		
				Tryptophan	2.0	Choline	1.0	Selenium	0.0002		
				Tyrosine	6.0	Folate	0.002	Zinc	0.075		
						Niacin	0.04				
						Pantothenic Acid (B5)	0.02				

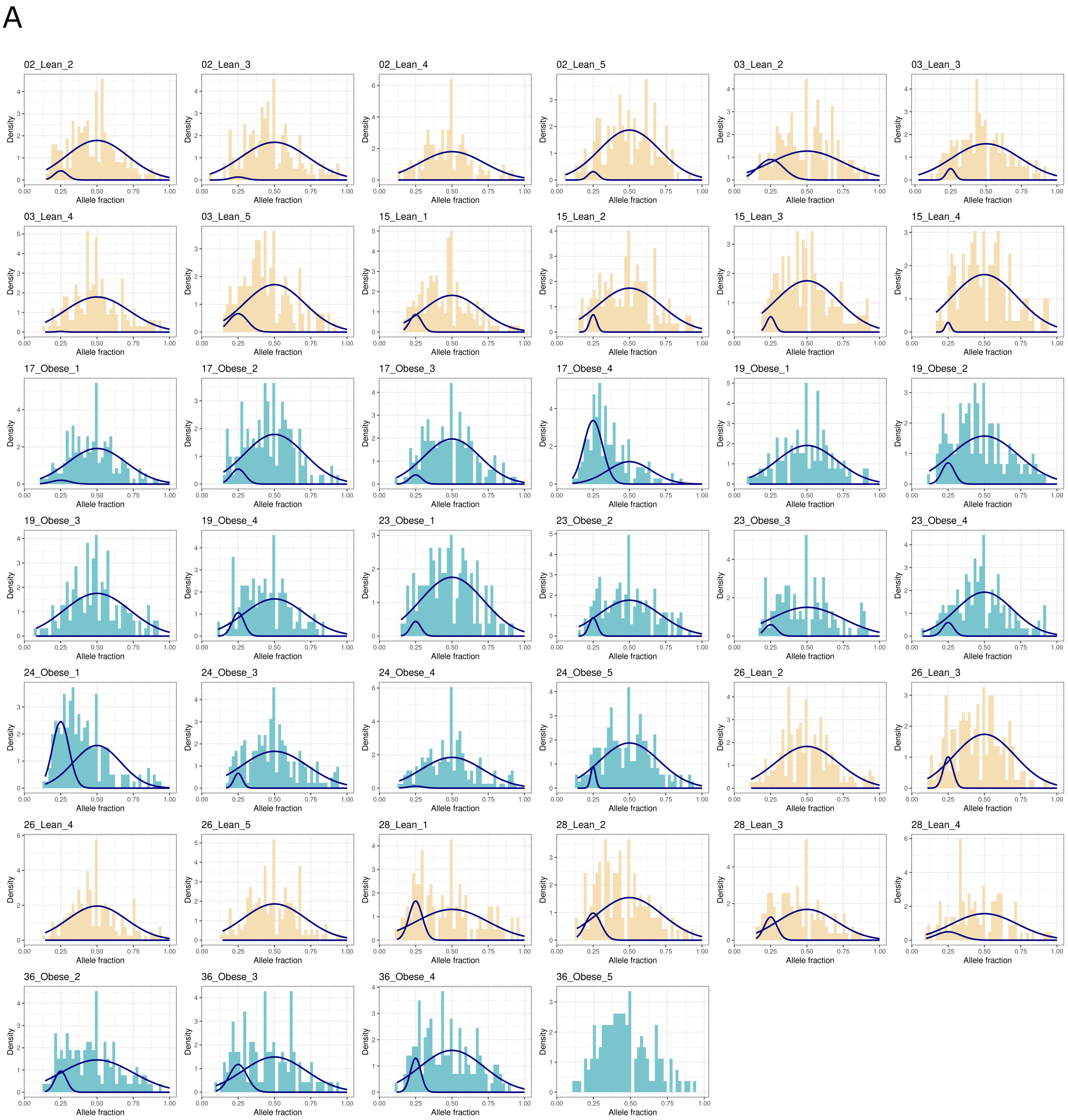
Supplier: http://www.lasvendi.com/en/lasqcdiets-eng/mice-rats/rod16-r-eng.html
Data sheet: https://www.lasvendi.com/files/PDF-EN/lasqcdiet_rod16_rad_data_eng.pdf



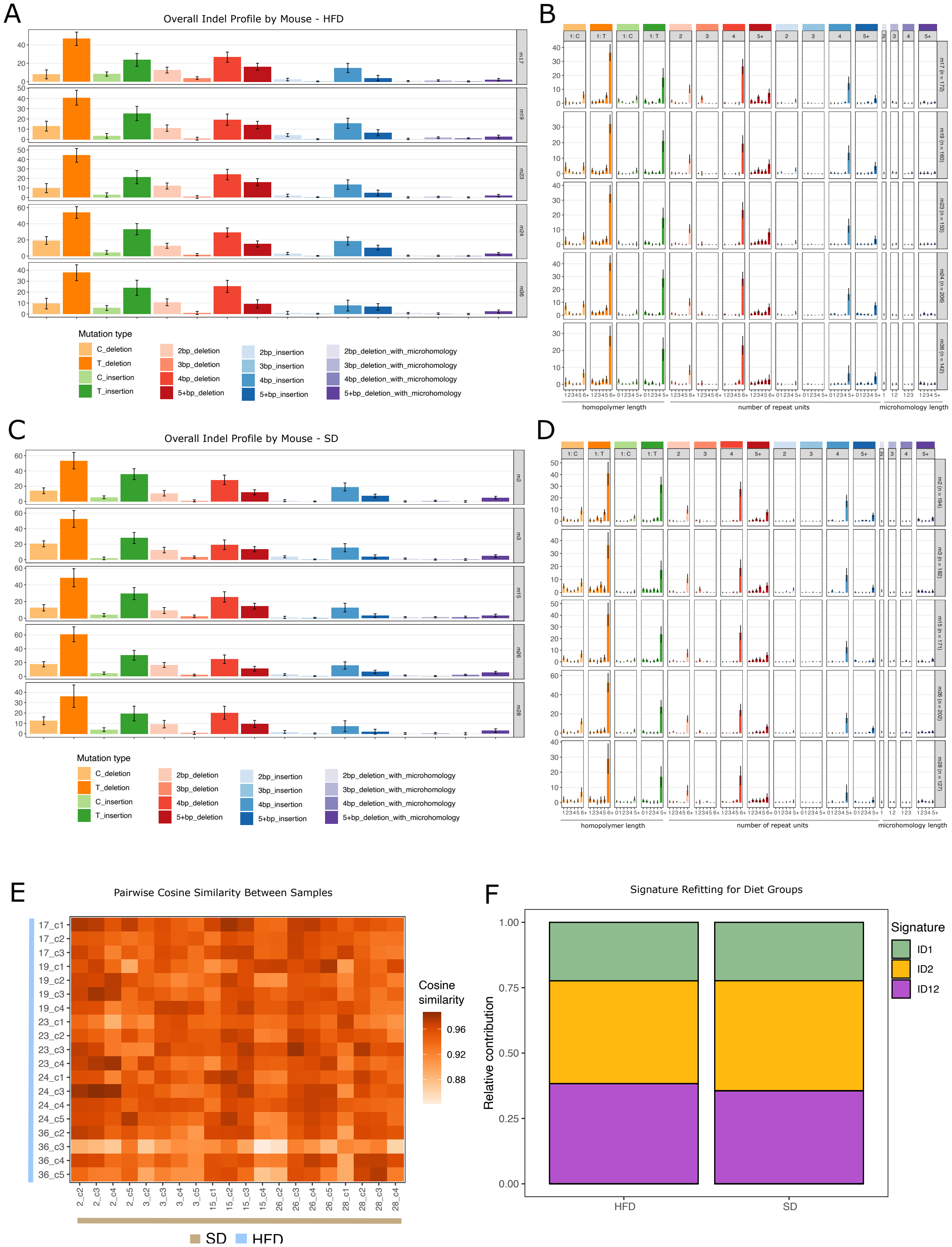
Supplementary Figure 1. (A) Variant allele frequency distribution (VAF) of single nucleotide variants (SNVs) for each organoid clone modeled with a Gaussian distribution, after deduction of germline variants found in the mouse tail sequences. Gaussian mixture model was fit with fixed means at 0.25 and 0.5 to identify the proportions of clonal and sub-clonal cell populations. The resulting distributions are shown in blue.



Supplementary Figure 2. (A) Average 96-channel SNV mutational profile per mouse in the HFD group. Error bars indicate ±1 standard deviation from the mean. **(B)** Average 96-channel SNV mutational profile per mouse in the SD group. Error bars indicate ±1 standard deviation from the mean. **(C)** Pairwise cosine similarity matrix between mutational profiles of all clones from SD and HFD mice respectively. The color scale has been adjusted to reflect the range of represented values from 0.9 to 1.0. **(D)** Aggregated mutational profiles by mean for each diet group after 10,000 bootstrap iterations of the mutational matrix of SD and HFD clones respectively. Error bars indicate ±1 standard deviation from the mean. The reported cosine similarity between the two averaged bootstrapped profiles is 0.9933.



Supplementary Figure 3. (A) Variant allele frequency distribution (VAF) of insertions and deletions (indels) for each organoid clone modeled with a Gaussian distribution, after deduction of germline variants found in the mouse tail sequences. Gaussian mixture model was fit with fixed means at 0.25 and 0.5 to identify the proportions of clonal and sub-clonal cell populations. The resulting distributions are shown in blue.



Supplementary Figure 4. (A) Average 16-channel indel mutational profile (main indel contexts) per mouse in the HFD group. Error bars indicate ±1 standard deviation from the mean. **(B)** Average 83-channel indel mutational profile (extended indel contexts) per mouse in the HFD group. Error bars indicate ±1 standard deviation from the mean. **(C)** Average 16-channel indel mutational profile (main indel contexts) per mouse in the SD group. Error bars indicate ±1 standard deviation from the mean. **(D)** Average 83-channel indel mutational profile (extended indel contexts) per mouse in the SD group. Error bars indicate ±1 standard deviation from the mean. **(E)** Pairwise cosine similarity matrix between indel mutational profiles of all organoid clones from SD and HFD mice. The color scale has been adjusted to reflect the range of represented values from 0.84 to 1.0. **(F)** Best subset refitting of diet groups to known indel signatures. The relative contribution per diet group is shown.

CHAPTER 3: DISCUSSION

Extended Interpretation of the Results

Strengths and Limitations of the Experimental System

Out study had multiple requirements of an experimental model. First, our *in-vivo* model needed to reflect diet induced obesity with common hallmarks of obesity. Since the C57/BL6J mice have been well-established as a model of diet induced obesity (Speakman *et al*, 2007; Collins *et al*, 2004), this requirement for our study was fulfilled. Furthermore, we were able to observe a significant increase in plasma cholesterol, and accumulation of white adipose tissue accompanied with weight gain in our HFD group. Directly harvesting intestinal stem cells and culturing these to clonality, allowed us to conduct a cell type specific investigation into mutagenesis in ISCs in response to high fat diet and the accompanying diet induced obesity. We deliberately chose to study inbred, age-matched mice with a wild-type genetic background to minimize confounding factors.

To justify our choice, we need to examine the most common model of tumorigenesis of the intestinal tract, the APCmin mouse. In this mouse model, a dominant truncating mutation in exon 1 of the APC gene leads to accumulation of polyps in the entire intestinal tract (Fodde *et al*, 1994). This mimics the phenotype of the human inherited cancer predisposition familial adenomatous polyposis (FAP). Loss of the APC gene has been found to be a strong driver for intestinal polyp formation. The polyps represent an early disease stage in the adenoma carcinoma sequence and have been found to have a polyclonal structure, even within small single adenomas (Merritt *et al*, 1997). However, to detect mutations and thus the activity of mutational signatures with confidence in a cohort with low sample sizes, a clonal structure is needed. Furthermore, loss of APC leads to different regionality of tumor formation along the length of the intestinal tract, reflecting differing mechanisms of tumor initiation, depending on the induction of the APC mutation, which would further confound the identification of mutational signatures (Haigis *et al*, 2004). Finally, APC-induced polyp formation is subject to many genetic modifiers, which is a further confounding factor introducing variability into the study (Rakoff-Nahoum & Medzhitov, 2007; Novelli *et al*, 1999). Indeed, the genetic modifier PLA2G2A (Phospholipase A2 Group IIA) was observed to not act in systemic manner but instead found to exert its effect heterogeneously and restricted to single ISCs or crypts (Novelli *et al*, 1999). Taken together, the use of the APCmin model would introduce too many confounding factors which would impede signature identification and interpretation.

Another factor to consider is the anatomical site to harvest ISCs from. Tumor formation in the APCmin mouse model is predominant in the small intestine, reflecting a major difference in the anatomical site between mouse model and humans. Although mouse cancer models of the large intestine exist, they often rely on additional chemical treatment (Bürtin *et al*, 2020), which would induce polyclonal lesions and likely produce a strong mutational imprint that would make identification of other, endogenous mutational processes, difficult. Thus, we excluded their use on the same basis we excluded the APCmin model. Nevertheless, the difference in anatomical site of cancer formation is important to consider in our chosen model as well. We chose to harvest intestinal stem cells from the middle part of the small intestine, the jejunum, for two reasons. First, even though murine tumor formation is not completely analogous to human tumor formation, sporadic lesions would most likely be expected to develop in the small intestine and therefore present the anatomic region of relevance for our model. Secondly, gene expression changes in response to a high fat diet were found most pronounced in the jejunum. Specifically, genes associated with cell cycle, inflammation and lipid metabolism were found altered in the high fat diet condition (de Wit *et al*, 2008). We therefore reasoned that intestinal stem cells in our model of diet induced obesity would be most affected in the jejunum.

For any study, the choice of model is critical for biological interpretation. We chose the anatomical site and sub-region based on previous findings that demonstrate the small intestine to be the most relevant site for the study of sporadic tumor formation in mice. It is known that the tumor formation in the murine small intestine, even with strong genetic drivers such as APC, does not capture all aspects of human carcinogenesis (Boivin *et al*, 2003; Washington & Zemper, 2019). Thus, studying the models' most relevant affected cell population for changes in the mutational landscape in response to a HFD is a logical choice for our question. We acknowledge that the difference in anatomical site is a limitation of our study, since this has implications for interpreting the etiology of tumor formation because regional differences in immune signaling and microbiome composition are likely to play a role in human CRC development as well (Tilg *et al*, 2018; Peng *et al*, 2020).

Functional Interpretation of Single Base Substitution Signatures

Mutational signatures are by nature compositional and not linearly separable. As discussed previously, compositional data describes parts of a whole, where the sum of all parts equals one. Applied to mutational signatures, this means that the mutation counts attributed to each channel of each signature found active, scaled to a relative contribution, add up to one. A natural consequence of these attributes is that mutations attributed to one channel of one signature, cannot be also attributed to the same channel of another signature, even when it is equally likely that a given mutation may have contributed to either signature.

The properties of the data thus pose a natural problem in the interpretation of signature etiologies. Since multiple signatures share mutational features across multiple channels, no definitive quantitative statement about absolute activity of mutational signatures can be made. This may or may not be critical depending on how many features overlap between signatures and how similar the proposed etiologies are. For example, SBS6 and SBS15 share highly similar but specific features in the C>T component and are thus likely to be confused both in NMF and by signature refitting algorithms. For a functional interpretation, however, the relative activity of multiple similar signatures with similar proposed etiologies only increases the confidence that the shared underlying process – defective mismatch repair in the case of SBS5 and SBS15 – is truly active. In contrast to this, if signatures share many features but not their proposed etiologies, functional interpretation quickly becomes difficult. Signatures with flat profiles such as SBS3 (homologous recombination deficiency), SBS5 (unknown), and SBS40 (unknown) are good examples of this. All three signatures have mutations attributed to all of the 96 available channels are thus have a high similarity among themselves. However, attributing mutations to SBS3 results in a different functional interpretation than attributing mutations to SBS5 or SBS40 respectively. As discussed in the introduction, SBS3 activity has direct and usable clinical implications, whereas detection of the activity of other signatures (SBS5/SBS40) with unknown etiologies does not.

Consequently, when we aim for an accurate functional interpretation, we need to consider signature activity in conjunction with the activity of other signatures detected, because if their signals overlap, their activity might not be accurately defined. Mutational signatures are a tool with specific strengths and weaknesses. The major strengths lie in the ability to capture the outcome of complex and overlapping biological processes while retaining single nucleotide resolution. The main weakness of this approach lies in the challenge to balance the analysis between mathematical possibility and biological sensibility. As already discussed in the introduction, the ability to understand the unique mechanisms that generate a specific type of mutation depends on the channels defined prior to NMF. Mathematically, it is possible to define endless possibilities of channels. However, if many uninformative channels are defined and included in the analysis, extracted signatures are likely biologically meaningless. Reversely, defining too few channels prohibits a clear distinction between signatures (Koh *et al*, 2020, 2021). Optimal interpretation thus demands an intermediate number of channels across multiple types of mutation classes. This allows to distinguish between unique processes on the one hand, while still capturing somewhat redundant activity of multiple similar signatures on the other hand, increasing confidence in the activity of a given process. In the following, I discuss the activity of signatures we identified in our cohort while considering these concepts.

Single Base Substitution Signature 1

We found SBS1 active in equal proportions in all samples from both diet groups. SBS1 is a clock-like signature which is found active in all samples. The signature is characterized by C>T transitions, specifically enriched in ACG, CCG, GCG, and TCG sequence contexts. SBS1 is proposed to be caused by spontaneous or enzymatic deamination of 5-methylcytosines (Nik-Zainal *et al*, 2012b; Alexandrov *et al*, 2013). This chemically converts cytosine to thymine, leaving behind G:T mismatches. If these lesions are not removed prior to the next round of DNA replication, the mutation becomes fixed as a T, resulting in a switch from a G:C to an A:T base pair. This mutational outcome was observed to correlate with age in most cancers and normal cell types. While the acquisition rate of SBS1 mutations varies substantially between cell types, these changes correlate with theoretical rates of stem cell division in various tissues, suggesting that SBS1 may function as a mitotic clock, timing the rate of cell division (Alexandrov *et al*, 2015; Moore *et al*, 2021; Lee-Six *et al*, 2019).

For SBS1, the absolute number of mutations attributed to this signature is very similar between NMF and the best subset refitting approach we used. Since SBS1 is defined by a few clear features in four C>T channels, spuriously assigned activity of this signature due to misattribution of mutations is highly unlikely. Coupled with the expectation that SBS1 is active in all samples, the assigned activity of SBS1 in our cohort makes biological sense. Because we observe equal proportions of SBS1 activity in all samples, we conclude that there is no difference in the rate of spontaneous deamination of 5-methylcytosines between the diet groups. In conjunction with the proposed etiology of SBS1, this could also indicate that HFD did not induce an increase in cell cycle progression, as faster cell division would also increase the rate of 5-methylcytosines deamination and thus would increase SBS1 activity.

Single Base Substitution Signature 5

SBS5 was one of two ubiquitous signatures found in our cohort in de-novo extraction, accounting for almost half of all mutations in a given sample. In contrast, with best subset refitting, activity of SBS5 was found only in 11 out of 39 samples, accounting for less than 20-25% of all mutations present in a sample.

SBS5 is a ubiquitous signature which is found in almost all samples and exhibits a clock like activity. While SBS5 is strongly correlated with age, it was also found associated to smoking and oxidative damage (Alexandrov *et al*, 2016; Kim *et al*, 2020). The rate of generation is different among tissues, indicating that the rate of generation of this signature is independently regulated (Alexandrov *et al*, 2015). The flat profile of the signature could indicate a mutational process which is common and unspecific to the sequence context. However, a deeper analysis of the main components, C>T and T>C revealed further insights. False incorporation of uracil and hypoxanthine, which are read as thymine and guanine respectively would lead to an A>G/T>C transitions, explaining the observed C>T component (Alexandrov *et al*, 2015). Additionally, T>C mutations in SBS5 show transcriptional strand bias, which could indicate that some mutations resulted from DNA adducts which are substrates for TC-NER (Alexandrov *et al*, 2015). Indeed, SBS5 was found increased in bladder cancers with concurrent ERCC2 mutations, which offers a potential hypothesis for the etiology of part of this signature (Kim *et al*, 2016). In summary, SBS5 does not have one clearly defined etiology but rather seems to arise from the simultaneous action of misincorporation of nucleotide analogs and oxidative/ alkylating damage and subsequent transcription coupled NER.

For the results of our study, the equal activity of SBS in both diets groups can be interpreted in the following ways. The amount of oxidative damage, lesions with adducts, and false incorporation of nucleotide analogs may be unchanged between diet groups or is adequately compensated for by functioning transcription coupled NER. Furthermore, it is unlikely that HFD induced an increase in proliferation and cell cycle progression, as this would increase the dependence on NER and over time lead to an increase of SBS5 associated mutations. This interpretation is in accordance with our interpretation of the activity of SBS1.

Single Base Substitution Signature 40

SBS40 has a similarly flat profile to SBS5 with an unknown etiology. Like SBS5, SBS40 is ubiquitous across many sample types and positively correlated with aging (Alexandrov *et al*, 2020). High cosine similarity between signature pairs not only means impaired algorithmic decoupling, but also that the number of mutations attributed to each signature will depend on mutation counts attributed to all the other signatures with high cosine similarity. This reasoning suggests that SBS5 and SBS40 might be somewhat mutually exclusive. The high similarity between these two signatures might suggest similar etiologies. Like SBS5, SBS40 might thus be the by-product of a combination of simultaneously acting mutagenic processes. Furthermore, the wide array of tissues in which either SBS5 or SBS40 are active (Moore *et al*, 2021; Lee-Six *et al*, 2019) would suggest that their etiology is linked to fundamental cellular processes that are common throughout tissues and cell types.

Ultimately, SBS5 and SBS40 have uncertain etiologies and a similar profile with a high cosine similarity (0.83), resulting in low deconstruction fidelity, both in NMF and in refitting approaches. This might explain the discrepancy of mutational activity attributed to SBS5 versus SBS40 in our results obtained from NMF compared to the results from best subset refitting. In NMF, only SBS5 was identified, while in refitting, most of the mutations were attributed to SBS40 instead of SBS5.

Single Base Substitution Signature 18

Next to SBSS5, SBS18 was the most active signature we identified in our cohort. SBS18 is defined by peaks in the C>A and C>T channels, although C>A peaks are more pronounced, especially at ACA, CCA, GCA, and TCT sequence contexts. SBS18 is highly similar in profile to SBS36, which has been proposed to be caused by defects in base excision repair due to mutations in the DNA glycosylase MUTYH (Viel *et al*, 2017). SBS18 activity is commonly observed in cancer samples and normal tissues, as well as in samples from experimental bottom-up studies of mutational signatures (Lee-Six *et al*, 2019; Moore *et al*, 2021; Rouhani *et al*, 2016). Whether the oxidative damage stems from metabolic activity *in-vivo* or oxidative damage incurred during culturing *ex-vivo*, the equivalent activity of SBS18 in both diet groups, suggests no increased damage generation or adequate compensation through functional base excision repair.

Single Base Substitution Signature 15

In the signature refitting approach, we identified minute activity of SBS15 in 2 SD samples and 2 HFD samples. SBS15 is characterized by major C>T peaks in GCN sequence contexts. Minor contributing peaks are C>T mutations in other sequence contexts, C>A mutations, and T>C mutations, which each contribute less than 10% of all mutations for this signature. SBS15 is one of six single base substitution signatures associated with a defect in mismatch repair. Zou and Owusu *et al* have demonstrated that specifically MSH6 mutations can recapitulate the profile of SBS15, drawing a causal link between MMR pathway mutations and this specific signature profile (Zou *et al*, 2018).

To exclude the possibility of an artefact during refitting, we supplemented the analysis with bootstrapped refitting, pooled by diet group. The results show that no residual activity of SBS15 can be detected in 1000 bootstrap iterations (Figure 20). Thus, we conclude that neither diet group exhibited specific mutational activity that would indicate a dysfunction in mismatch repair.

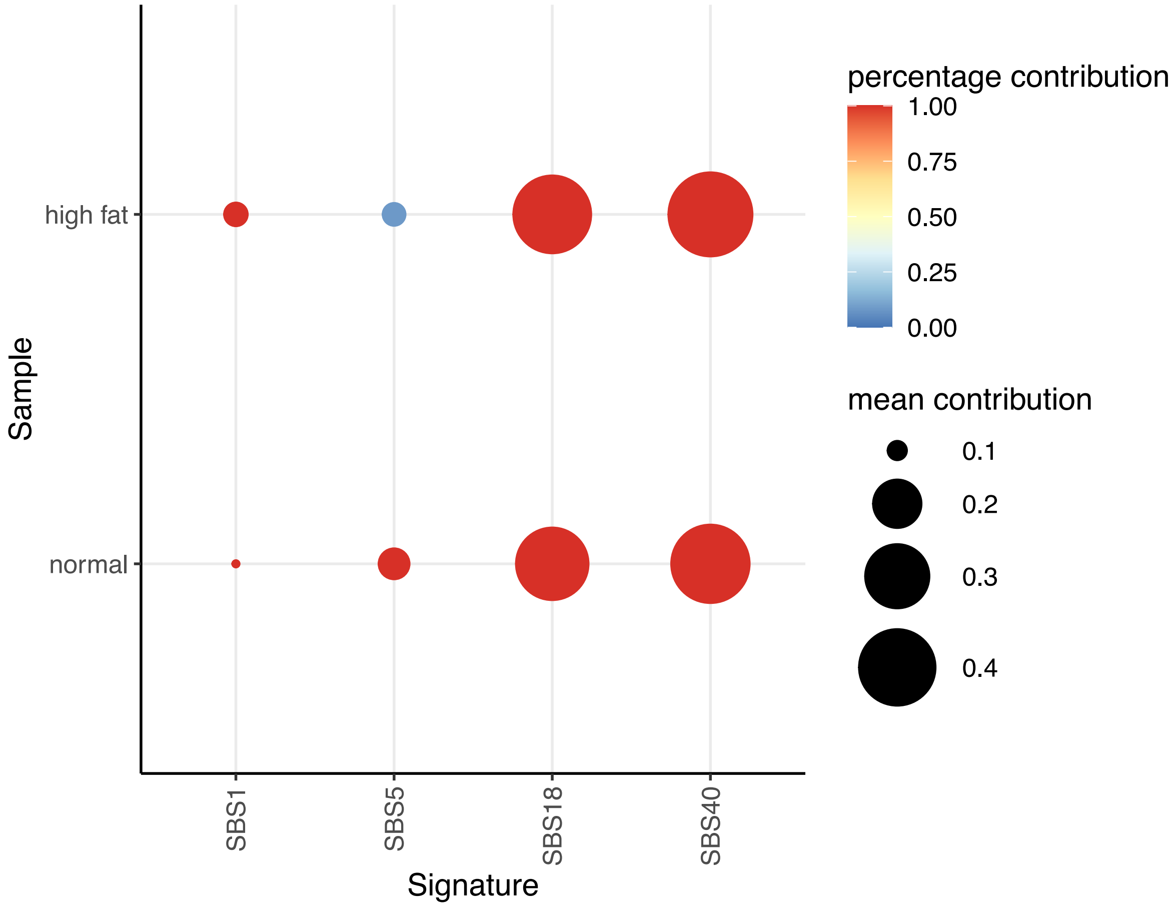


Figure 20. Bootstrapped refitting of SBS signatures (1000 iterations) The size of the dots indicates the mean contribution of the signature on the x-axis for all bootstrap iterations where this signature was found active. The color scale indicates the percentage of bootstrap iterations where the signature was detected. Bootstrapping was performed by pooling the mutational matrix per diet group and including signatures SBS1, SBS5, SBS15, SBS18, and SBS40.

Functional Interpretation of Indel Signatures

Indel signatures are a generalized set of sequence contexts based on size, type of repeat unit (homopolymer or polynucleotide repeat tract), and whether microhomologies exist at indel junctions. This definition of indel channels was chosen because indels cannot be exhaustively described like SBS signatures and thus, these broad categories were chosen as a first attempt to reveal biological underpinnings of processes generating indels. However, several of the proposed 83 channels were not informative when applied to a large cancer cohort (Alexandrov *et al*, 2020; Koh *et al*, 2021a). As already mentioned, the classification of mutations into biologically meaningful channels before non-negative matrix decomposition is key to uncovering new biological insight. Therefore, the currently defined channels might not be sufficient descriptors for capturing meaningful indel generating processes. Other classifications have not been tried yet but might provide additional insights into the biology of this mutation class.

Indel Signatures 1 and 2

Of all indel signatures we detected, ID1 and ID2 were the most active in all samples. ID1 is a characteristic signature defined by a large peak marking 1-base pair T-insertions at homopolymers of length five or more. This one insertion type makes up over 85% of all indel types which are part of the ID1 signature. ID2, similarly to ID1, is described by a single prominent peak. Differently from ID1, the defining peak is a 1 base pair T deletion at homopolymers with length 6 or more. ID1 and ID2 are common signatures which are found in almost all samples and positively correlate with age (Alexandrov *et al*, 2020). This clock-like behavior, where the total mutation burden attributed to ID1 and ID2 increases with age, indicates that the underlying mutational mechanisms may be associated with normal cellular processes. Indeed, the proposed etiology of ID1 is the slippage of the replicated strand during DNA replication and the proposed etiology of ID2 involves slippage of the template DNA strand during replication. Furthermore, ID1 and ID2 activity have been observed to increase in genetic backgrounds with DNA mismatch repair deficiency (Alexandrov *et al*, 2020). It might be plausible that even without dMMR, increased proliferation speeds might cause replication fork slippage. Whether fast cell cycle progression specifically increases mutation counts associated to ID1 and ID2 remains a hypothesis to be tested.

We have not observed differential activity of ID1 or ID2 between diet groups or accelerated accumulation of mutations for other clock-like signatures (SBS1, SBS5, SBS50). Thus, the conclusion is, that the ID1 and ID2 activity observed in our cohort likely reflect normal rates of replication associated indel generation.

Indel Signature 7

Although we found limited evidence of ID7 activity in our cohort, I briefly discuss the implications of the proposed dMMR etiology of this signature. ID7 is mainly characterized by 1 base pair C and T deletions in homopolymer sequence contexts with a length of 5+ and 6+ base pairs. Additionally, 2 base pair deletions at repeat units of length 4,5 or 6 bp are also represented. ID7 was observed to occur in cancer samples which harbored many mutations attributed to ID1 and ID2, while simultaneously showing activity for SBS signatures commonly associated with mismatch repair deficiency (SBS6, SBS15, SBS21, SBS16, SBS44) (Alexandrov *et al*, 2020). Thus, the proposed etiology of ID7 is thought to be due to defective mismatch repair. If the underlying process generating ID7 is truly a defect in mismatch repair, it could make sense that ID7 exacerbates ID1 and ID2. Mechanistically, in a genetic background with mismatch repair deficiency, more mismatches are expected to occur in homopolymer stretches, such as the sequence contexts in which ID1 and ID2 characteristic T- insertions and deletions accumulate.

However, the association of ID7 with dMMR rests upon the correlation with SBS signature activities which are known to be due to defective mismatch repair. Since signature etiologies are complex, association must not be confused with causation. While it is reasonable to assume that mismatch repair deficiency increases ID1 and ID2 mutations, this association is not conclusive evidence that ID7 is a standalone signature which represents an independent readout for detecting mismatch repair deficiency. Since MMR is a repair process needed in many cellular contexts – during enhanced replication, during normal replication, for repairing mistakes during other repair processes – it is reasonable to assume, that multiple different mutations within different sequence contexts are overlapping, even if the underlying causative process is identical. Thus, the defined readouts of 1 base pair C and T deletions characteristic for ID7 might simply be an unobserved byproduct of already defined SBS signatures due to dMMR.

Ultimately it remains unknown whether ID7 represents a unique and independent indel generation process that is biologically distinct from other processes which cause the already known SBS signatures of defective mismatch repair. Thus, in the context of our results, I would interpret the presence of ID7 only in conjunction with the activity of SBS15. Since we only observed minor activity of SBS15 and ID7, defective mismatch repair is likely not a strong driver of mutagenesis in our cohort. Consequently, this means, that the activity of ID1 and ID2 is likely not influenced by a dysfunction in MMR, but rather explainable by normal cellular replication.

Indel Signatures 9 and 12

We detected two other indel signatures in our cohort, ID9 and ID12 (etiology unknown). ID9 was only detected in the refitting approach for two SD samples and could not be validated with bootstrapped refitting. ID12, on the other hand, was a ubiquitous signature, which was recovered in NMF, best subset refitting, as well as bootstrapped refitting. In all three approaches, ID12 was found equally active in both diet groups.

ID12 is defined by one prominent peak of 2bp deletions in 6+ bp repeat regions. Aside from this major characteristic, ID12 shows mutation counts in 3-4 bp deletions at larger repeat regions and 1 bp C and T deletions in homopolymers of varying lengths. ID9 consists of 1bp C and T deletions in homopolymers of different lengths. Since stable ID9 activity could not be verified with bootstrapping, it is plausible that the presence of ID9 in signature refitting was caused by misattribution of the 1bp C and T deletion channels, which are shared between ID9 and ID12. The other 3-4 bp deletions which define ID12 are not shared with other known indel signatures, possibly indicating that these channels carry a specific meaning. However, functional interpretation of this signature bears two caveats. First, ID12 has only been identified in small proportions in only 2 cancers (liposarcoma, prostate adenocarcinoma) in a single study, which introduced the current framework for defining indel signatures (Alexandrov *et al*, 2020). Second, whether the proposed channels are optimal for capturing relevant indel mutations is debated, because many channels stayed uninformative so far and no other frameworks have been tried (Koh *et al*, 2021). Thus, it remains unknown if ID12 presents a unique and meaningful mutational process or is the byproduct of other processes.

It is curious to see that ID12 is ubiquitously active in both diet groups, identified by all computational approaches used. In our context, under the assumption that the defined mutational channels carry specific meaning, this could potentially indicate a housekeeping process, akin to the replication, aging, and oxidative stress signatures we identified.

Interpretation of Mutagenesis in a Tissue Specific Context

In summary, we recover the activity of mutational processes related to aging (SBS1) due to mechanisms involved in spontaneous degradation of the DNA molecule, cellular replication (SBS1, SBS5/SBS40, ID1, ID2), related to mechanisms of DNA slippage and other clock-like mutagenesis, and oxidative stress (SBS18) explainable by metabolic stress and ROS encountered *in-vivo* or *ex-vivo* during culturing. The equal activity of mutational signatures across both diet groups and their associated etiologies of normal processes indicates functional DNA repair in both groups, which suggests that HFD and associated systemic changes are not strongly affecting DNA damage and repair processes to drive mutagenesis beyond baseline processes. In the introduction, DNA repair was discussed as a tissue specific process, and thus the results should be interpreted in a tissue specific context as well.

The crypt-villus architecture, coupled with the rapid renewal of the intestinal epithelial lining poses tissue specific constraints on mutagenesis. First, the renewal of the entire intestinal epithelium within 3-5 days rapidly eliminates mutated cells which are not long-lived intestinal stem cells. Within the crypt, however, long lived ISCs underlie mechanisms governed by Wnt and Notch signaling, regulating the balance between stem cell division and differentiation (Funk *et al*, 2020). As discussed in the review articles presented in chapter 1, multiple cell populations which can compensate for the impairment of loss of ISCs exist. Thus, mutated cells, which have not yet gained strong driver mutations might undergo negative selection. Additionally, the crypt architecture poses another barrier, even for replicating ISCs which already acquired mutations, as these cells would stay confined within the crypts, unless such a cell can overcome the crypt compartment and expand clonally through crypt fission. Estimates of the driver mutation rate across different tissues provide evidence for this argument. In normal human colorectal epithelium, only 1% of cells contained driver mutations (Lee-Six *et al*, 2019), whereas esophageal epithelium contained 50% (Martincorena *et al*, 2018) and skin 30% cells with driver mutations (Martincorena *et al*, 2015). Hence, the barrier for oncogenic transformation is especially high in the intestinal epithelium because the high rate of overturn, presence of compensatory cell populations, and physical confinement to the crypt play together to negatively select mutated cells which haven not gained a significant fitness advantage yet. Considering these mechanisms of control, the results of Beyaz *et al* can be interpreted from a tissue specific perspective. Their study found enhanced tumorigenesis in HFD fed APCmin mice via a mechanism increasing the pool of available stem cells, thus increasing the chance to produce a cell which gains enough fitness advantage to overcome these barriers, at least in a cancer pre-disposed background.

Our results indicate that HFD alone is not sufficient to induce increased or differential mutagenesis in the absence of other perturbances such as driver mutations or chemical agents. The signatures we recover indicate that, a wild-type background, DNA repair seems to be functioning regardless of diet and if genomic stability is not increased, the chances for producing a cell which can overcome the tissue specific barriers for oncogenic transformation are comparatively low.

Future Outlook

Throughout this thesis, the utility for understanding cancer etiology and clinical applicability of mutational signatures was explored in depth and contrasted with the caveats present in experimental design and computational analysis. The current state of the art for assigning signature etiologies relies on statistical association between signature activity and other omics data and on signature activity mapping resulting from genetic or chemical perturbation experiments. However, neither approach is sufficient to uncover all etiologies due to several inherent problems. The computational approach of associating signature etiologies to certain genetic backgrounds based on other omics measurements reflects association, not causation, which is especially critical in trying to discern primary causes from secondary effects. The experimental bottom-up approaches, while suitable for detecting the effects of a single perturbation, still do not result in a one-to-one mapping of perturbations to signature etiologies because a single perturbation may spark an adaptive response that leads to the generation of multiple signatures (Koh *et al*, 2020). Finally, the overlap between different signatures and the emergent nature of some signatures, as discussed in chapter 1, make it difficult to resolve similar or functionally connected signatures.

Just like experimental and computational approaches synergized to detect, expand, and annotate the known mutational signatures, further elucidation of signature etiologies will likely benefit from equally complementary approaches. On the one hand, a more fine-grained mechanistic understanding of DNA damage and repair processes are needed to define informative channels for biologically relevant mutation types. On the other hand, a more connected birds-eye perspective of the overlap and connection between different signatures would allow to ask deeper questions about their co-occurrence or mutual exclusivity in different biological contexts.

The first objective is enabled by new DNA damage and repair centric sequencing techniques (Lensing *et al*, 2016; Zatopek *et al*, 2019; Hussmann *et al*, 2021; Mingard *et al*, 2020), and the development of high-fidelity sequencing methods which increase the signal to noise ratio (Abascal *et al*, 2021). The ability to detect mutations, even in a polyclonal sample, additionally circumvents the need for single cell selection and culturing, thus further reducing artefacts otherwise introduced in the experimental workflow. The second objective may be reached by developing new computational frameworks which allow to probe the interaction between signatures. In such a framework, signature co-occurrence may indicate a shared underlying etiology or a mechanistic connection where the activity of one signature precedes another. Alternatively, one can image negative interactions between signatures where exclusivity might indicate synthetic lethality of the underlying pathways.

Applied to uncovering the etiologies of obesity associated cancer, this might mean studying HFD exposure in combination with different perturbation events, including modelling the evolution of mutational signatures throughout the adenoma carcinoma sequence. By analyzing the exact lesions generated and how they are repaired, coupled with a systematic overview of signature interactions in a time dependent manner, it may become possible to connect distinct DNA damage and repair processes to specific activity increases in mutational signatures. Ultimately, this would allow to move beyond association and instead draw a causal link between processes and signatures.

REFERENCES

Abascal F, Harvey LMR, Mitchell E, Lawson ARJ, Lensing S V., Ellis P, Russell AJC, Alcantara RE, Baez-Ortega A, Wang Y, *et al* (2021) Somatic mutation landscapes at single-molecule resolution. *Nat 2021 5937859* 593: 405–410

Adhikari S, Choudhury S, Mitra P, Dubash J, Sajankila S & Roy R (2012) Targeting Base Excision Repair for Chemosensitization. *Anticancer Agents Med Chem* 8: 351–357

Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S & Stratton MR (2015) Clock-like mutational processes in human somatic cells. *Nat Genet* 47: 1402–1407

Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, Totoki Y, Fujimoto A, Nakagawa H, Shibata T, *et al* (2016) Mutational signatures associated with tobacco smoking in human cancer. *Science* 354: 618

Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, *et al* (2020) The repertoire of mutational signatures in human cancer. *Nat 2020 5787793* 578: 94–101

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin A V., Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, *et al* (2013) Signatures of mutational processes in human cancer. *Nature* 500: 415–21

Aravind L, Walker DR & Koonin E V (1999) Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res* 27

Avgerinos KI, Spyrou N, Mantzoros CS & Dalamaga M (2019) Obesity and cancer risk: Emerging biological mechanisms and perspectives. *Metabolism* 92: 121–135 doi:10.1016/j.metabol.2018.11.001 [PREPRINT]

Baez-Ortega A & Gori K (2019) Computational approaches for discovery of mutational signatures in cancer. *Brief Bioinform* 20: 77

Bartkova J, Hořejší Z, Koed K, Krämer A, Tort F, Zleger K, Guldberg P, Sehested M, Nesland JM, Lukas C, *et al* (2005) DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* 434: 864–870

Bartkova J, Rezaei N, Liontos M, Karakaidos P, Kletsas D, Issaeva N, Vassiliou LVF, Kolettas E, Niforou K, Zoumpourlis VC, *et al* (2006) Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature* 444: 633–637

Beyaz S, Mana MD, Roper J, Kedrin D, Saadatpour A, Hong SJ, Bauer-Rowe KE, Xifaras ME, Akkad A, Arias E, *et al* (2016) High-fat diet enhances stemness and tumorigenicity of intestinal progenitors. *Nature* 531: 53–58

Bianchi JJ, Zhao X, Mays JC & Davoli T (2020) Not all cancers are created equal: Tissue specificity in cancer genes and pathways. *Curr Opin Cell Biol* 63: 135–143 doi:10.1016/j.ceb.2020.01.005 [PREPRINT]

Blasiak J (2021) Single-Strand Annealing in Cancer. *Int J Mol Sci* 22: 2167

Blokzijl F, Janssen R, van Boxtel R & Cuppen E (2018) MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med* 10

Blüher M (2019) Obesity: global epidemiology and pathogenesis. *Nat Rev Endocrinol 2019 155* 15: 288–298

Bogaert J & Prenen H (2014) Molecular genetics of colorectal cancer. *Ann Gastroenterol* 27: 9

Boivin GP, Washington K, Yang K, Ward JM, Pretlow TP, Russell R, Besselsen DG, Godfrey VL, Doetschman T, Dove WF, *et al* (2003) Pathology of mouse models of intestinal cancer: Consensus report and recommendations. *Gastroenterology* 124: 762–777

Bürtin F, Mullins CS & Linnebacher M (2020) Mouse models of colorectal cancer: Past, present and future perspectives. *World J Gastroenterol* 26: 1394

Caldecott KW (2008) Single-strand break repair and genetic disease. *Nat Rev Genet 2008 98* 9: 619–631

Calle EE, Rodriguez C, Walker-Thurmond K & Thun MJ (2009) Overweight, Obesity, and

Mortality from Cancer in a Prospectively Studied Cohort of U.S. Adults. *http://dx.doi.org/101056/NEJMoa021423* 348: 1625–1638

Ceccaldi R, Liu JC, Amunugama R, Hajdu I, Primack B, Petalcorin MIR, O'Connor KW, Konstantinopoulos PA, Elledge SJ, Boulton SJ, *et al* (2015) Homologous recombination-deficient tumors are hyper-dependent on POLQ-mediated repair. *Nature* 518: 258

Ceccaldi R, Sarangi P & D'Andrea AD (2016) The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol 2016 176* 17: 337–349

Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, Kim J, Kwiatkowski DJ, Fargo DC, Mieczkowski PA, *et al* (2015) An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat Genet* 47: 1067–72

Chang HHY, Pannunzio NR, Adachi N & Lieber MR (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* 18: 495–506

Christofori G & Semb H (1999) The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. *Trends Biochem Sci* 24: 73–76 doi:10.1016/S0968-0004(98)01343-7 [PREPRINT]

Ciccia A & Elledge SJ (2010) The DNA Damage Response: Making It Safe to Play with Knives. *Mol Cell* 40: 179–204

Collins S, Martin TL, Surwit RS & Robidoux J (2004) Genetic vulnerability to diet-induced obesity in the C57BL/6J mouse: physiological and molecular characteristics. *Physiol Behav* 81: 243–248

COSMIC (2020) COSMIC | Mutational Signatures. (https://cancer.sanger.ac.uk/signatures/cn/cn3/) [PREPRINT]

Counter CM, Avilion AA, Lefeuvre CE, Stewart NG, Greider CW, Harley CB & Bacchetti S (1992) Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J* 11: 1921–1929

Das L, Quintana VG & Sweasy JB (2020) NTHL1 in Genomic Integrity, Aging and Cancer. *DNA Repair (Amst)* 93: 102920

DeBerardinis RJ & Chandel NS (2020) We need to talk about the Warburg effect. *Nat Metab* 2: 127–129 doi:10.1038/s42255-020-0172-2 [PREPRINT]

Degasperi A, Zou X, Amarante TD, Martinez-Martinez A, Koh GCC, Dias JML, Heskin L, Chmelova L, Rinaldi G, Wang VYW, *et al* (2022) Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. *Science* 376

Dianov GL (2011) Base excision repair targets for cancer therapy. *Am J Cancer Res* 1: 845

Drews RM, Hernando B, Tarabichi M, Haase K, Lesluyes T, Smith PS, Morrill Gavarró L, Couturier DL, Liu L, Schneider M, *et al* (2022) A pan-cancer compendium of chromosomal instability. *Nature* 606: 976–983

Drost J, van Boxtel R, Blokzijl F, Mizutani T, Sasaki N, Sasselli V, de Ligt J, Behjati S, Grolleman JE, van Wezel T, *et al* (2017a) Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science* 358: 234–238

Drost J, Van Boxtel R, Blokzijl F, Mizutani T, Sasaki N, Sasselli V, De Ligt J, Behjati S, Grolleman JE, Van Wezel T, *et al* (2017b) Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science (80-)* 358: 234–238

Drost J, Van Jaarsveld RH, Ponsioen B, Zimberlin C, Van Boxtel R, Buijs A, Sachs N, Overmeer RM, Offerhaus GJ, Begthel H, *et al* (2015) Sequential cancer mutations in cultured human intestinal stem cells. *Nature* 521: 43–47

Duan M, Ulibarri J, Liu KJ & Mao P (2020) Role of nucleotide excision repair in cisplatin resistance. *Int J Mol Sci* 21: 1–13 doi:10.3390/ijms21239248 [PREPRINT]

Duchartre Y, Kim YM & Kahn M (2016) The Wnt signaling pathway in cancer. *Crit Rev Oncol Hematol* 99: 141–149

Fearon EF & Vogelstein B (1990) A Genetic Model for Colorectal Tumorigenesis. *Cell* 61: 759–767

Fischer A, Illingworth CJR, Campbell PJ & Mustonen V (2013) EMu: probabilistic inference of mutational processes and their localization in the cancer genome. *Genome Biol* 14: R39

Fodde R, Edelmann W, Yang K, Van Leeuwen C, Carlson C, Renault B, Breukel C, Alt E, Lipkin M, Khan PM, *et al* (1994) A targeted chain-termination mutation in the mouse Apc gene results in multiple intestinal tumors. *Proc Natl Acad Sci* 91: 8969–8973

Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, *et al* (2017) COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Res* 45: D777–D783

Friedberg EC (2001) How nucleotide excision repair protects against cancer. *Nat Rev Cancer* 1: 22–33

Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA & Ellenberger T (2005) DNA Repair and Mutagenesis. *DNA Repair Mutagen*

Friedenreich CM, Ryder-Burbidge C & McNeil J (2021) Physical activity, obesity and sedentary behavior in cancer etiology: epidemiologic evidence and biologic mechanisms. *Mol Oncol* 15: 790–800

Funk MC, Zhou J & Boutros M (2020) Ageing, metabolism and the intestine. *EMBO Rep* 21

Gorgoulis VG, Vassiliou LVF, Karakaidos P, Zacharatos P, Kotsinas A, Liloglou T, Venere M, DiTullio RA, Kastrinakis NG, Levy B, *et al* (2005) Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* 434: 907–913

Grundy GJ & Parsons JL (2020) Base excision repair and its implications to cancer therapy. *Essays Biochem* 64: 831

Haigis KM, Hoff PD, White A, Shoemaker AR, Halberg RB & Dove WF (2004) Tumor regionality in the mouse intestine reflects the mechanism of loss of Apc function. *Proc Natl Acad Sci* 101: 9769–9773

Hainaut P & Pfeifer GP (2016) Somatic TP53 mutations in the era of genome sequencing. *Cold Spring Harb Perspect Med* 6

Hajdu SI (2011a) A note from history: Landmarks in history of cancer, part 1. *Cancer* 117: 1097–1102

Hajdu SI (2011b) A note from history: Landmarks in history of cancer, part 2. *Cancer* 117: 2811–2820 doi:10.1002/cncr.25825 [PREPRINT]

Halazonetis TD, Gorgoulis VG & Bartek J (2008) An oncogene-induced DNA damage model for cancer development. *Science* 319: 1352–1355

Hanahan D & Folkman J (1996) Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell* 86: 353–364 doi:10.1016/S0092-8674(00)80108-7 [PREPRINT]

Hanahan D & Weinberg RA (2000) The hallmarks of cancer. *Cell* 100: 57–70 doi:10.1016/S0092-8674(00)81683-9 [PREPRINT]

Hanahan D & Weinberg RA (2011) Hallmarks of cancer: The next generation. *Cell* 144: 646–674 doi:10.1016/j.cell.2011.02.013 [PREPRINT]

Harper JW & Elledge SJ (2007) The DNA Damage Response: Ten Years After. *Mol Cell* 28: 739–745 doi:10.1016/j.molcel.2007.11.015 [PREPRINT]

Heather JM & Chain B (2016) The sequence of sequencers: The history of sequencing DNA. *Genomics* 107: 1–8 doi:10.1016/j.ygeno.2015.11.003 [PREPRINT]

Helleday T (2011) The underlying mechanism for the PARP and BRCA synthetic lethality: Clearing up the misunderstandings. *Mol Oncol* 5: 387

Helleday T, Eshtad S & Nik-Zainal S (2014) Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* 15: 585–598 doi:10.1038/nrg3729 [PREPRINT]

Herbig U, Jobling WA, Chen BPC, Chen DJ & Sedivy JM (2004) Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21CIP1, but not p16INK4a. *Mol Cell* 14: 501–513

Hills SA & Diffley JFX (2014) DNA replication and oncogene-induced replicative stress. *Curr Biol* 24: R435–R444 doi:10.1016/j.cub.2014.04.012 [PREPRINT]

Hodel KP, Sun MJS, Ungerleider N, Park VS, Williams LG, Bauer DL, Immethun VE, Wang

J, Suo Z, Lu H, *et al* (2020) POLE Mutation Spectra Are Shaped by the Mutant Allele Identity, Its Abundance, and Mismatch Repair Status. *Mol Cell* 78: 1166-1177.e6

Hoeijmakers JHJ (2009) DNA Damage, Aging, and Cancer. *https://doi.org/101056/NEJMra0804615* 361: 1475–1485

Hopkins BD, Goncalves MD & Cantley LC (2016) Obesity and cancer mechanisms: Cancer metabolism. *J Clin Oncol* 34: 4277–4283 doi:10.1200/JCO.2016.67.9712 [PREPRINT]

Hussmann JA, Ling J, Ravisankar P, Yan J, Cirincione A, Xu A, Simpson D, Yang D, Bothmer A, Cotta-Ramusino C, *et al* (2021) Mapping the genetic landscape of DNA double-strand break repair. *Cell* 184: 5653-5669.e25

Islam SMA, Díaz-Gay M, Wu Y, Barnes M, Vangara R, Bergstrom EN, He Y, Vella M, Wang J, Teague JW, *et al* (2022) Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *Cell Genomics* 2: 100179

Iwai Y, Ishida M, Tanaka Y, Okazaki T, Honjo T & Minato N (2002) Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade. *Proc Natl Acad Sci U S A* 99: 12293–12297

Jaacks LM, Vandevijvere S, Pan A, McGowan CJ, Wallace C, Imamura F, Mozaffarian D, Swinburn B & Ezzati M (2019) The Obesity Transition: Stages of the global epidemic. *lancet Diabetes Endocrinol* 7: 231

Jackson SP & Bartek J (2009) The DNA-damage response in human biology and disease. *Nature* 461: 1071–1078

Jhunjhunwala S, Hammer C & Delamarre L (2021) Antigen presentation in cancer: insights into tumour immunogenicity and immune evasion. *Nat Rev Cancer* 21: 298–312 doi:10.1038/s41568-021-00339-z [PREPRINT]

Junttila MR & Evan GI (2009) P53 a Jack of all trades but master of none. *Nat Rev Cancer* 9: 821–829 doi:10.1038/nrc2728 [PREPRINT]

Kamp JA, van Schendel R, Dilweg IW & Tijsterman M (2020) BRCA1-associated structural variations are a consequence of polymerase theta-mediated end-joining. *Nat Commun* 11

Kasar S, Kim J, Improgo R, Tiao G, Polak P, Haradhvala N, Lawrence MS, Kiezun A, Fernandes SM, Bahl S, *et al* (2015) Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nat Commun* 6

Kebudi R, Kiykim A & Sahin MK (2019) Primary Immunodeficiency and Cancer in Children; A Review of the Literature. *Curr Pediatr Rev* 15: 245–250

Kelley MR, Logsdon D & Fishel ML (2014) Targeting DNA repair pathways for cancer treatment: What's new? *Futur Oncol* 10: 1215–1237 doi:10.2217/fon.14.60 [PREPRINT]

Kim J, Mouw KW, Polak P, Braunstein LZ, Kamburov A, Tiao G, Kwiatkowski DJ, Rosenberg JE, Van Allen EM, D'Andrea AD, *et al* (2016) Somatic ERCC2 Mutations Are Associated with a Distinct Genomic Signature in Urothelial Tumors. *Nat Genet* 48: 600

Kim YA, Wojtowicz D, Sarto Basso R, Sason I, Robinson W, Hochbaum DS, Leiserson MDM, Sharan R, Vadin F & Przytycka TM (2020) Network-based approaches elucidate differences within APOBEC and clock-like signatures in breast cancer. *Genome Med* 12

Kinzler KW & Vogelstein B (1997) Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* 386: 761–763

Koh G, Degasperi A, Zou X, Momen S & Nik-Zainal S (2021a) Mutational signatures: emerging concepts, caveats and clinical applications. *Nat Rev Cancer* 21: 619–637

Koh G, Degasperi A, Zou X, Momen S & Nik-Zainal S (2021b) Mutational signatures: emerging concepts, caveats and clinical applications. *Nat Rev Cancer* 21: 619–637

Koh G, Zou X & Nik-Zainal S (2020a) Mutational signatures: Experimental design and analytical framework. *Genome Biol* 21: 1–13

Koh G, Zou X & Nik-Zainal S (2020b) Mutational signatures: Experimental design and analytical framework. *Genome Biol* 21: 1–13

Krokan HE & Bjørås M (2013) Base excision repair. *Cold Spring Harb Perspect Biol* 5: 1–22

Kucab JE, Zou X, Morganella S, Joel M, Nanda AS, Nagy E, Gomez C, Degasperi A, Harris R, Jackson SP, *et al* (2019) A Compendium of Mutational Signatures of Environmental

Agents. *Cell* 177: 821-836.e16

Kunkel TA & Erie DA (2005) DNA mismatch repair. *Annu Rev Biochem* 74: 681–710 doi:10.1146/annurev.biochem.74.082803.133243 [PREPRINT]

Langevin F, Crossan GP, Rosado I V., Arends MJ & Patel KJ (2011) Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nat 2011 4757354* 475: 53–58

Latchman Y, Wood CR, Chernova T, Chaudhary D, Borde M, Chernova I, Iwai Y, Long AJ, Brown JA, Nunes R, *et al* (2001) PD-L2 is a second ligand for PD-1 and inhibits T cell activation. *Nat Immunol* 2: 261–268

Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F & Straif K (2016) Body Fatness and Cancer — Viewpoint of the IARC Working Group. *N Engl J Med* 375: 794–798

Lee-Six H, Olafsson S, Ellis P, Osborne RJ, Sanders MA, Moore L, Georgakopoulos N, Torrente F, Noorani A, Goddard M, *et al* (2019) The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* 574: 532–537

Lee DD & Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. *Nature* 401: 788–791

Lee DD & Seung HS (2001) Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*

Lensing S V., Marsico G, Hänsel-Hertsch R, Lam EY, Tannahill D & Balasubramanian S (2016) DSBCapture: in situ capture and direct sequencing of dsDNA breaks. *Nat Methods* 13: 855

Li GM (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res* 18: 85–98 doi:10.1038/cr.2007.115 [PREPRINT]

Li HD, Cuevas I, Zhang M, Lu C, Alam MM, Fu YX, You MJ, Akbay EA, Zhang H & Castrillon DH (2018) Polymerase-mediated ultramutagenesis in mice produces diverse cancers with high mutational load. *J Clin Invest* 128: 4179–4191

Li SKH & Martin A (2016) Mismatch Repair and Colon Cancer: Mechanisms and Therapies Explored. *Trends Mol Med* 22: 274–289 doi:10.1016/j.molmed.2016.02.003 [PREPRINT]

Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, Khurana E, Waszak S, Korbel JO, Haber JE, *et al* (2020) Patterns of somatic structural variation in human cancer genomes. *Nature* 578: 112

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

Loeb LA (1991) Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res (Chicago, Ill)* 51: 3075–3079

Lord CJ & Ashworth A (2017) PARP Inhibitors: The First Synthetic Lethal Targeted Therapy. *Science* 355: 1152

Ma J, Setton J, Lee NY, Riaz N & Powell SN (2018) The therapeutic significance of mutational signatures from DNA repair deficiency in cancer. *Nat Commun* 9: 3292 doi:10.1038/s41467-018-05228-y [PREPRINT]

Macintyre G, Goranova TE, De Silva D, Ennis D, Piskorz AM, Eldridge M, Sie D, Lewsley LA, Hanif A, Wilson C, *et al* (2018) Copy-number signatures and mutational processes in ovariancarcinoma. *Nat Genet* 50: 1262

Magrin L, Fanale D, Brando C, Corsini LR, Randazzo U, Di Piazza M, Gurrera V, Pedone E, Bazan Russo TD, Vieni S, *et al* (2022) MUTYH-associated tumor syndrome: The other face of MAP. *Oncogene 2022 4118* 41: 2531–2539

Maiorano D, Etri J El, Franchet C & Hoffmann JS (2021) Translesion synthesis or repair by specialized dna polymerases limits excessive genomic instability upon replication stress. *Int J Mol Sci* 22 doi:10.3390/ijms22083924 [PREPRINT]

Marteijn JA, Lans H, Vermeulen W & Hoeijmakers JHJ (2014) Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol* 15: 465–481 doi:10.1038/nrm3822 [PREPRINT]

Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, Cagan A, Murai K, Mahbubani K, Stratton MR, *et al* (2018) Somatic mutant clones colonize the human esophagus with age. *Science (80-)* 362: 911–917

Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, Wedge DC, Fullam A, Alexandrov LB, Tubio JM, *et al* (2015) High burden and pervasive positive selection of somatic mutations in normal human skin. *Science (80-)* 348: 880–886

Mateos-Gomez PA, Gong F, Nair N, Miller KM, Lazzerini-Denchi E & Sfeir A (2015) Mammalian Polymerase Theta Promotes Alternative-NHEJ and Suppresses Recombination. *Nature* 518: 254

Matsui WH (2016) Cancer stem cell signaling pathways.

Mehta A & Haber JE (2014) Sources of DNA Double-Strand Breaks and Models of Recombinational DNA Repair. *Cold Spring Harb Perspect Biol* 6

Mei C, Lei L, Tan LM, Xu XJ, He BM, Luo C, Yin JY, Li X, Zhang W, Zhou HH, *et al* (2020) The role of single strand break repair pathways in cellular responses to camptothecin induced DNA damage. *Biomed Pharmacother* 125: 109875

Meier B, Volkova N V., Hong Y, Schofield P, Campbell PJ, Gerstung M & Gartner A (2018) Mutational signatures of DNA mismatch repair deficiency in C. elegans and human cancers. *Genome Res* 28: 666–675

Mekonnen N, Yang H & Shin YK (2022) Homologous Recombination Deficiency in Ovarian, Breast, Colorectal, Pancreatic, Non-Small Cell Lung and Prostate Cancers, and the Mechanisms of Resistance to PARP Inhibitors. *Front Oncol* 12: 2747

Merritt AJ, Gould KA & Dove WF (1997) Polyclonal structure of intestinal adenomas in ApcMin/+ mice with concomitant loss of Apc+ from all tumor lineages. *Proc Natl Acad Sci U S A* 94: 13927–13931

Di Micco R, Fumagalli M, Cicalese A, Piccinin S, Gasparini P, Luise C, Schurra C, Garré M, Giovanni Nuciforo P, Bensimon A, *et al* (2006) Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* 444: 638–642

Midthun L, Shaheen S, Deisch J, Senthil M, Tsai J & Hsueh CT (2019) Concomitant KRAS and BRAF mutations in colorectal cancer. *J Gastrointest Oncol* 10: 577

Mingard C, Wu J, McKeague M & Sturla SJ (2020) Next-generation DNA damage sequencing. *Chem Soc Rev* 49: 7354–7377

Moore L, Cagan A, Coorens THH, Neville MDC, Sanghvi R, Sanders MA, Oliver TRW, Leongamornlert D, Ellis P, Noorani A, *et al* (2021) The mutational landscape of human somatic and germline cells. *Nature* 597: 381–386

Mukherjee S (2010) The emperor of all maladies : a biography of cancer. 571

Nebgen B, Vangara R, Hombrados-Herrera MA, Kuksova S & Alexandrov B (2020) A neural network for determination of latent dimensionality in Nonnegative Matrix Factorization. *Mach Learn Sci Technol*: 1–17

Negrini S, Gorgoulis VG & Halazonetis TD (2010) Genomic instability an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* 11: 220–228

Nguyen L, W. M. Martens J, Van Hoeck A & Cuppen E (2020) Pan-cancer landscape of homologous recombination deficiency. *Nat Commun 2020 111* 11: 1–12

Nickoloff JA, Sharma N, Taylor L, Allen SJ & Hromas R (2021) The Safe Path at the Fork: Ensuring Replication-Associated DNA Double-Strand Breaks are Repaired by Homologous Recombination. *Front Genet* 12

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

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

Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, *et al* (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534: 47–54

Nik-Zainal S & Morganella S (2017) Mutational signatures in breast cancer: The problem at the DNA level. *Clin Cancer Res* 23: 2617–2629

Novelli MR, Wasan H, Rosewell I, Bee J, Tomlinson IP, Wright NA & Bodmer WF (1999)

Tumor burden and clonality in multiple intestinal neoplasia mouse/normal mouse aggregation chimeras. *Proc Natl Acad Sci U S A* 96: 12553–12558

Nowell PC (1976) The clonal evolution of tumor cell populations. *Science* 194: 23–28

Odes EJ, Randolph-Quinney PS, Steyn M, Throckmorton Z, Smilg JS, Zipfel B, Augustine TN, Beer F De, Hoffman JW, Franklin RD, *et al* (2016) Earliest hominin cancer: 1.7-million-yearold osteosarcoma from Swartkrans cave, South Africa. *S Afr J Sci* 112

Omichessan H, Severi G & Perduca V (2019) Computational tools to detect signatures of mutational processes in DNA from tumours: A review and empirical comparison of performance. *PLoS One* 14

Peng C, Ouyang Y, Lu N & Li N (2020) The NF-κB Signaling Pathway, the Microbiota, and Gastrointestinal Tumorigenesis: Recent Advances. *Front Immunol* 11

Pereira S, Cline DL, Glavas MM, Covey SD & Kieffer TJ (2021) Tissue-Specific Effects of Leptin on Glucose and Lipid Metabolism. *Endocr Rev* 42: 1–28

Pilati C, Shinde J, Alexandrov LB, Assié G, André T, Hélias-Rodzewicz Z, Ducoudray R, Le Corre D, Zucman-Rossi J, Emile JF, *et al* (2017) Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. *J Pathol* 242: 10–15

Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin ML, Ordõez GR, Bignell GR, *et al* (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463: 191–196

Rakoff-Nahoum S & Medzhitov R (2007) Regulation of spontaneous intestinal tumorigenesis through the adaptor protein MyD88. *Science (80-)* 317: 124–127

Randolph-Quinney PS, Williams SA, Steyn M, Meyer MR, Smilg JS, Churchill SE, Odes EJ, Augustine T, Tafforeau P & Berger LR (2016) Osteogenic tumour in Australopithecus sediba: Earliest hominin evidence for neoplastic disease. *S Afr J Sci* 112: 1–7

Rim EY, Clevers H & Nusse R (2022) The Wnt Pathway: From Signaling Mechanisms to Synthetic Modulators. *Annu Rev Biochem* 91: 571–598

Robinson PS, Coorens THH, Palles C, Mitchell E, Abascal F, Olafsson S, Lee BCH, Lawson ARJ, Lee-Six H, Moore L, *et al* (2021) Increased somatic mutation burdens in normal human cells due to defective DNA polymerases. *Nat Genet* 53: 1434–1442

Robinson PS, Thomas LE, Abascal F, Jung H, Harvey LMR, West HD, Olafsson S, Lee BCH, Coorens THH, Lee-Six H, *et al* (2022) Inherited MUTYH mutations cause elevated somatic mutation rates and distinctive mutational signatures in normal human cells. *Nat Commun* 13

Rosado I V., Langevin F, Crossan GP, Takata M & Patel KJ (2011) Formaldehyde catabolism is essential in cells deficient for the Fanconi anemia DNA-repair pathway. *Nat Struct Mol Biol 2011 1812* 18: 1432–1434

Rouhani FJ, Nik-Zainal S, Wuster A, Li Y, Conte N, Koike-Yusa H, Kumasaka N, Vallier L, Yusa K & Bradley A (2016) Mutational History of a Human Cell Lineage from Somatic to Induced Pluripotent Stem Cells. *PLoS Genet* 12

San Filippo J, Sung P & Klein H (2008) Mechanism of Eukaryotic Homologous Recombination. *https://doi.org/101146/annurev.biochem77061306125255* 77: 229–257

Sanger F, Nicklen S & Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463–5467

Sato T, Stange DE, Ferrante M, Vries RGJ, Van Es JH, Van den Brink S, Van Houdt WJ, Pronk A, Van Gorp J, Siersema PD, *et al* (2011) Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* 141: 1762–72

Sato T, Vries RG, Snippert HJ, Van De Wetering M, Barker N, Stange DE, Van Es JH, Abo A, Kujala P, Peters PJ, *et al* (2009) Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* 459: 262–265

Schimmel J, van Schendel R, den Dunnen JT & Tijsterman M (2019) Templated Insertions: A Smoking Gun for Polymerase Theta-Mediated End Joining. *Trends Genet* 35: 632–644

Schrempf A, Slyskova J & Loizou JI (2021) Targeting the DNA Repair Enzyme Polymerase θ in Cancer Therapy. *Trends in Cancer* 7: 98–111 doi:10.1016/j.trecan.2020.09.007 [PREPRINT]

Senga SS & Grose RP (2021) Hallmarks of cancer—the new testament. *Open Biol* 11

Seol JH, Shim EY & Lee SE (2018) Microhomology-mediated end joining: Good, bad and ugly. *Mutat Res - Fundam Mol Mech Mutagen* 809: 81–87 doi:10.1016/j.mrfmmm.2017.07.002 [PREPRINT]

Shang Y & Meng F-L (2021) Repair of programmed DNA lesions in antibody class switch recombination: common and unique features. *Genome Instab Dis* 2: 115–125

Shrivastav M, De Haro LP & Nickoloff JA (2008) Regulation of DNA double-strand break repair pathway choice. *Cell Res* 18: 134–147

Soerjomataram I & Bray F (2021) Planning for tomorrow: global cancer incidence and the role of prevention 2020–2070. *Nat Rev Clin Oncol* 18: 663–672

Speakman J, Hambly C, Mitchell S & Król E (2007) Animal models of obesity. *Obes Rev* 8: 55–61

Speakman JR & Goran MI (2010) Tissue-Specificity and Ethnic Diversity in Obesity-Related Risk of Cancer May Be Explained by Variability in Insulin Response and Insulin Signaling Pathways. *Obesity* 18: 1071–1078

Steele CD, Abbasi A, Islam SMA, Bowes AL, Khandekar A, Haase K, Hames-Fathi S, Ajayi D, Verfaillie A, Dhami P, *et al* (2022a) Signatures of copy number alterations in human cancer. *Nature* 606: 984–991

Steele CD, Pillay N & Alexandrov LB (2022b) An overview of mutational and copy number signatures in human cancer. *J Pathol* 257: 454–465

Suhail Y, Cain MP, Vanaja K, Kurywchak PA, Levchenko A, Kalluri R & Kshitiz (2019) Systems Biology of Cancer Metastasis. *Cell Syst* 9: 109–127 doi:10.1016/j.cels.2019.07.003 [PREPRINT]

Sulkowski PL, Oeck S, Dow J, Economos NG, Mirfakhraie L, Liu Y, Noronha K, Bao X, Li J, Shuch BM, *et al* (2020) Oncometabolites suppress DNA repair by disrupting local chromatin signalling. *Nature* 582: 586–591

Sulkowski PL, Sundaram RK, Oeck S, Corso CD, Liu Y, Noorbakhsh S, Niger M, Boeke M, Ueno D, Kalathil AN, *et al* (2018) Krebs-cycle-deficient hereditary cancer syndromes are defined by defects in homologous-recombination DNA repair. *Nat Genet* 50: 1086–1092 doi:10.1038/s41588-018-0170-4 [PREPRINT]

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A & Bray F (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 71: 209–249

Supek F, Miñana B, Valcárcel J, Gabaldón T & Lehner B (2014) Synonymous mutations frequently act as driver mutations in human cancers. *Cell* 156: 1324–1335

Symington LS & Gautier J (2011) Double-strand break end resection and repair pathway choice. *Annu Rev Genet* 45: 247–271

Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, *et al* (2019) COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* 47: D941

Tilg H, Adolph TE, Gerner RR & Moschen AR (2018) The Intestinal Microbiota in Colorectal Cancer. *Cancer Cell* 33: 954–964

Tran KB, Lang JJ, Compton K, Xu R, Acheson AR, Henrikson HJ, Kocarnik JM, Penberthy L, Aali A, Abbas Q, *et al* (2022) The global burden of cancer attributable to risk factors, 2010–19: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 400: 563–591

Tsui LC & Scherer SW (2008) The Human Genome Project. In *Biotechnology: Second, Completely Revised Edition* pp 41–59.

Turner NC (2017) Signatures of DNA-Repair Deficiencies in Breast Cancer. *N Engl J Med* 377: 2490–2492

Viel A, Bruselles A, Meccia E, Fornasarig M, Quaia M, Canzonieri V, Policicchio E, Urso ED, Agostini M, Genuardi M, *et al* (2017) A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer. *EBioMedicine*

20: 39–49

Wang B, Rong X, Palladino END, Wang J, Fogelman AM, Martín MG, Alrefai WA, Ford DA & Tontonoz P (2018) Phospholipid Remodeling and Cholesterol Availability Regulate Intestinal Stemness and Tumorigenesis. *Cell Stem Cell* 22: 206-220.e4

Washington K & Zemper AED (2019) Apc-related models of intestinal neoplasia: a brief review for pathologists. *Surg Exp Pathol 2019 21* 2: 1–9

Wculek SK, Cueto FJ, Mujal AM, Melero I, Krummel MF & Sancho D (2020) Dendritic cells in cancer immunology and immunotherapy. *Nat Rev Immunol* 20: 7–24 doi:10.1038/s41577-019-0210-z [PREPRINT]

WHO (2020) Global health estimates: Leading causes of death. *World Heal Organ*: 1–2

de Wit NJ, Bosch-Vermeulen H, de Groot PJ, Hooiveld GJ, Bromhaar MMG, Jansen J, Müller M & van der Meer R (2008) The role of the small intestine in the development of dietary fat-induced obesity and insulin resistance in C57BL/6J mice. *BMC Med Genomics* 1

Wright WD, Shah SS & Heyer WD (2018) Homologous recombination and the repair of DNA double-strand breaks. *J Biol Chem* 293: 10524

Yaffe MB (2019) Why geneticists stole cancer research even though cancer is primarily a signaling disease. *Sci Signal* 12: 3483

Zatopek KM, Potapov V, Maduzia LL, Alpaslan E, Chen L, Evans TC, Ong JL, Ettwiller LM & Gardner AF (2019) RADAR-seq: A RAre DAmage and Repair sequencing method for detecting DNA damage on a genome-wide scale. *DNA Repair (Amst)* 80: 36–44

Zhao S, Wang F & Liu L (2019) Alternative lengthening of telomeres (ALT) in tumors and pluripotent stem cells. *Genes (Basel)* 10: 1030 doi:10.3390/genes10121030 [PREPRINT]

Zhou BBS & Elledge SJ (2000) The DNA damage response: Putting checkpoints in perspective. *Nature* 408: 433–439 doi:10.1038/35044005 [PREPRINT]

Zou X, Koh GCC, Nanda AS, Degasperi A, Urgo K, Roumeliotis TI, Agu CA, Badja C, Momen S, Young J, *et al* (2021) A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nat Cancer* 2: 643–657

Zou X, Owusu M, Harris R, Jackson SP, Loizou JI & Nik-Zainal S (2018a) Validating the concept of mutational signatures with isogenic cell models. *Nat Commun 2018 91* 9: 1–16

Zou X, Owusu M, Harris R, Jackson SP, Loizou JI & Nik-Zainal S (2018b) Validating the concept of mutational signatures with isogenic cell models. *Nat Commun* 9: 1744

CV

Mathilde Meyenberg

BIOINFORMATICIAN

CONTACT

date of birth: 21.04.1994
citizenship: german

Wattgasse 57/ 15
1160 , Wien

mmeyenberg@cemm.oeaw.ac.at
mathilde@menchelab.org

+ 43 (0)660 - 30 180 12

SKILLS

Python, R, Bash, SQL	●●●●●
C++	●●○○○
Machine Learning	●●●●○
Software Development	●●●○○
Automatization	●●●●○
Data Management/ Documentation	●●●●●

LANGUAGES

German (native speaker)

English (native speaker)

AWARDS

Austrian Academy of Sciences -
DOC Fellowship 2020-2022

SUMMARY

I have been working in the field of bioinformatics for more than 3 years, combining my analytical and programming skills with the experience I gained in wet-lab research.
My area of expertise includes genomics research with a focus on cancer development, high-throughput genetic screens with phenotypic readouts, and time resolved metabolomics.
I am looking forward to becoming part of a dynamic team and apply my knowledge of molecular biology and bioinformatic analyses to exciting research questions.

EDUCATION

2018-
present

Ph.D. program in Molecular Medicine
CeMM/ Medical University of Vienna
Menche Lab/ Loizou Lab

2021-
present

Master of Science in Bioinformatics
FH Campus Wien - University of Applied Sciences
expected graduation Sept. 2023

2013-
2016

Bachelor of Science in Chemistry
Bachelor of Science in Biology
(summa cum laude)
Warren Wilson College, Asheville, NC, USA

EXPERIENCE

2019-
present

Ph.D. Project Work

- analysis of whole genome sequencing data
  GATK variant calling, mutational signature analysis
- analysis of methylation, gene expression, and mutational signature profiles in publicly available cancer data (TCGA, ICGC)
- analysis of time resolved metabolomics and metabolomic networks
- image analysis of phenotypic readouts in cellular assays
  immunoflourescence foci, cell morphology

2021-
present

Data Manager and Data Manager Team Representative
CeMM - Research Center for Molecular Medicine

- development and maintenance of data migration, storage, and annotation guidelines
- training and teaching for cluster usage and data storage/ retrieval

2018-
2020

Ph.D. Project Work

- high-throughput phenotypic CRISPR screens
- intestinal organoid technology for in vitro mutational studies

2017-
2018

Graduate Research Assistant
Department of Pharmaceutical Sciences
Wayne State University, Detroit, MI, USA

2015

Research Intern Molecular Genomics
SASA- Science and Advice for Scottish Agriculture
Government Research Lab, Edinburgh, UK

PUBLICATIONS

research
paper

Mutational Landscape of Intestinal Stem Cells After Long-term
In Vivo Exposure to High Fat Diet
(under review in Scientific Reports)

review

Meyenberg M, Ferreira da Silva J & Loizou JI (2021)
Tissue Specific DNA Repair Outcomes Shape the Landscape of Genome Editing.
Frontiers in Genetics

preview

Bernardo S, **Meyenberg M** & Loizou JI (2021)
Decomposing the mutational landscape of cancer genomes with RepairSig.
Cell Systems

review

Ferreira da Silva J, **Meyenberg M** & Loizou JI (2021)
Tissue specificity of DNA repair: the CRISPR compass.
Trends in Genetics

research
paper

Moretton A, Slyskova J, Simaan ME, Arasa-Verge EA, **Meyenberg M**,
Cerrón-Infantes DA, Unterlass MM & Loizou JI (2022)
Clickable Cisplatin Derivatives as Versatile Tools to Probe the
DNA Damage Response to Chemotherapy.
Frontiers in Oncology

Additional Manuscripts in Preparation
"Pan-cancer Analysis of Interactions of Mutational Signatures"

"Signer 2.0: Exploring Mutation Signatures and Exposures to Mutational Processes"

Conferences Presentations
2019 – ***Sy-Stem Symposium*** (Stem Cell Research), Vienna (Poster Presentation)
2019 – ***YSA***, Medical University of Vienna (Poster Presentation)
2019 – ***EU-LIFE*** Scientific Meeting, Babraham Institute, Cambridge, UK (Poster Presentation)
2021 – ***Sy-Stem Symposium*** (Stem Cell Research), (Poster Presentation)
2022 – ***Contra – Computational Oncology Conference***, Andermatt, CH (Keynote Presentation)

Teaching Experience
2022 – prepared and hosted a computational workshop on mutational signature analysis for the Molecular Precision Medicine Master's Program of the Medical University of Vienna

APPENDIX

Publisher Permissions

Frontiers in Genetics

"Tissue Specific DNA Repair Outcomes Shape the Landscape of Genome Editing"

Publisher permission for Review Article doi: *10.3389/fgene.2021.728520*

"Copyright © 2021 Meyenberg, Ferreira da Silva and Loizou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms."

Copyright Statement

"Under the Frontiers Conditions for Website Use and the Frontiers General Conditions for Authors, authors of articles published in Frontiers journals retain copyright on their articles, except for any third-party images and other materials added by Frontiers, which are subject to copyright of their respective owners. Authors are therefore free to disseminate and re-publish their articles, subject to any requirements of third-party copyright owners and subject to the original publication being fully cited. Visitors may also download and forward articles subject to the citation requirements and subject to any fees Frontiers may charge for downloading licenses. The ability to copy, download, forward or otherwise distribute any materials is always subject to any copyright notices displayed. Copyright notices must be displayed prominently and may not be obliterated, deleted or hidden, totally or partially."

Taken from: https://www.frontiersin.org/journals/genetics/about#about-copyright
Accessed: 2023-01-25

Trends in Genetics

"Tissue specificity of DNA repair: the CRISPR compass"

Publisher permission for Review Article doi: 10.1016/j.tig.2021.07.010

© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http:// creativecommons.org/licenses/by-nc-nd/4.0/).

Copyright Statement

***"Open access content** published under a CC-BY user license, as well as open access content published under other types of user licenses depending on the nature of your proposed re-use (for example, commercial vs. nonprofit use), may not require written permission, subject to proper acknowledgement. Permissions vary depending on the license type, and we recommend that readers check the license details carefully before re-using the material."*

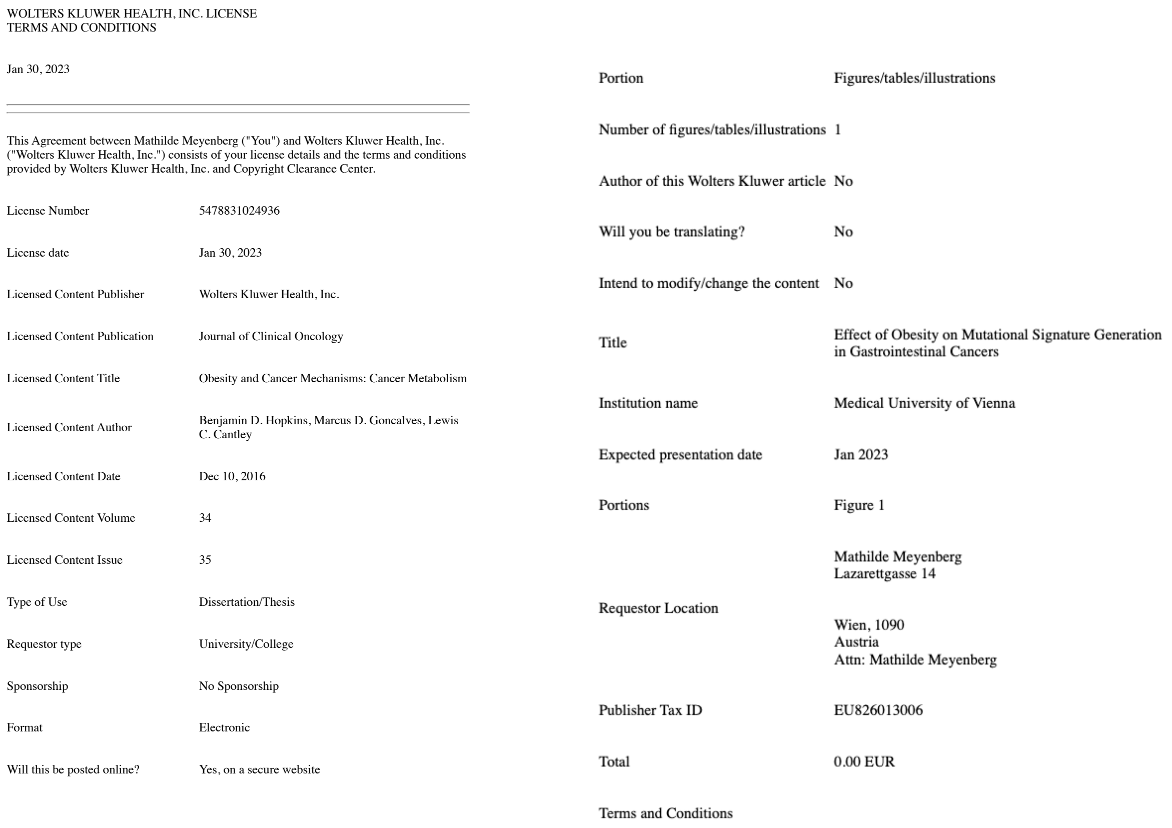
Taken from: https://www.elsevier.com/about/policies/copyright/permissions
Accessed: 2023-01-25

Licensing Information for Figures

Figure 1

WHO. Copyright. Available at: https://www.who.int/about/policies/publishing/copyright. (Accessed: 30th January 2023)

Figure 2



WOLTERS KLUWER HEALTH, INC. LICENSE
TERMS AND CONDITIONS

Jan 30, 2023

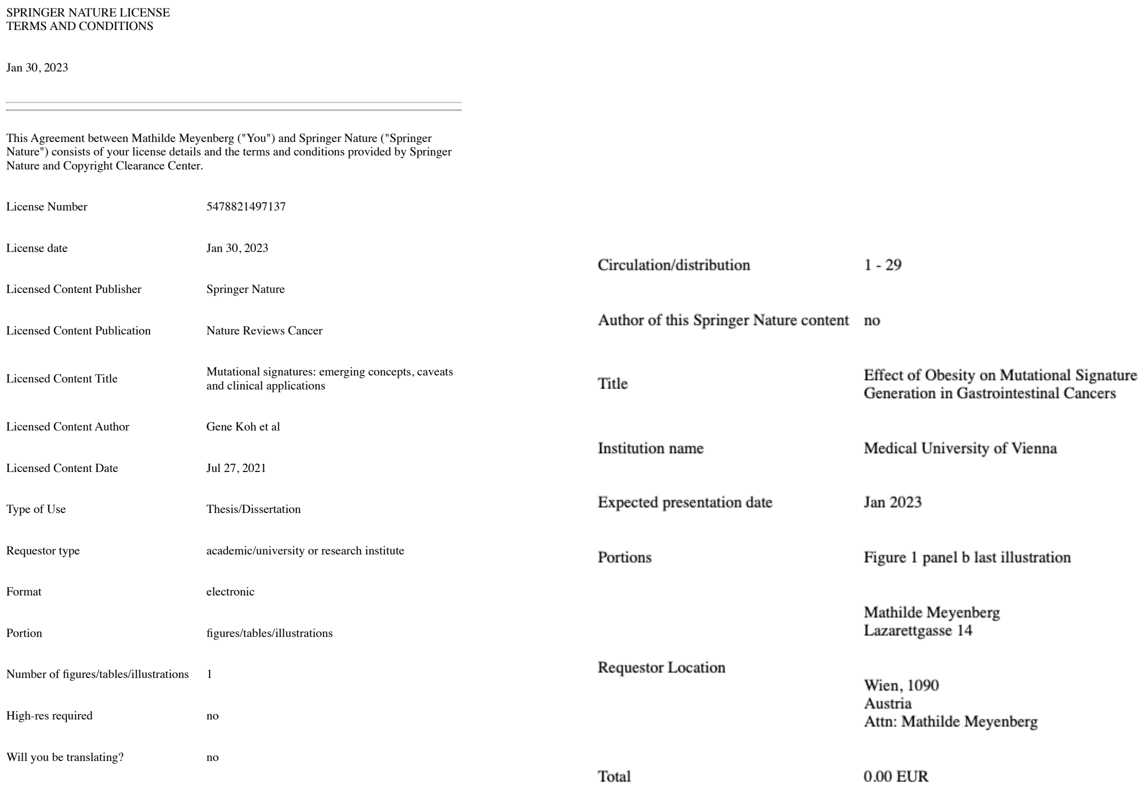
This Agreement between Mathilde Meyenberg ("You") and Wolters Kluwer Health, Inc. ("Wolters Kluwer Health, Inc.") consists of your license details and the terms and conditions provided by Wolters Kluwer Health, Inc. and Copyright Clearance Center.

License Number	5478831024936
License date	Jan 30, 2023
Licensed Content Publisher	Wolters Kluwer Health, Inc.
Licensed Content Publication	Journal of Clinical Oncology
Licensed Content Title	Obesity and Cancer Mechanisms: Cancer Metabolism
Licensed Content Author	Benjamin D. Hopkins, Marcus D. Goncalves, Lewis C. Cantley
Licensed Content Date	Dec 10, 2016
Licensed Content Volume	34
Licensed Content Issue	35
Type of Use	Dissertation/Thesis
Requestor type	University/College
Sponsorship	No Sponsorship
Format	Electronic
Will this be posted online?	Yes, on a secure website
Portion	Figures/tables/illustrations
Number of figures/tables/illustrations	1
Author of this Wolters Kluwer article	No
Will you be translating?	No
Intend to modify/change the content	No
Title	Effect of Obesity on Mutational Signature Generation in Gastrointestinal Cancers
Institution name	Medical University of Vienna
Expected presentation date	Jan 2023
Portions	Figure 1
Requestor Location	Mathilde Meyenberg Lazarettgasse 14 Wien, 1090 Austria Attn: Mathilde Meyenberg
Publisher Tax ID	EU826013006
Total	0.00 EUR
Terms and Conditions	

Figure 17

COSMIC. Licensing COSMIC Data. Available at: https://cancer.sanger.ac.uk/cosmic/license. (Accessed: 30th January 2023)

Figure 18



SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jan 30, 2023

This Agreement between Mathilde Meyenberg ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5478821497137
License date	Jan 30, 2023
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Cancer
Licensed Content Title	Mutational signatures: emerging concepts, caveats and clinical applications
Licensed Content Author	Gene Koh et al
Licensed Content Date	Jul 27, 2021
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Effect of Obesity on Mutational Signature Generation in Gastrointestinal Cancers
Institution name	Medical University of Vienna
Expected presentation date	Jan 2023
Portions	Figure 1 panel b last illustration
Requestor Location	Mathilde Meyenberg Lazarettgasse 14 Wien, 1090 Austria Attn: Mathilde Meyenberg
Total	0.00 EUR