UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA
IN SCIENZE BIOMEDICHE

Curriculum: ONCOLOGIA SPERIMENTALE E CLINICA

CICLO XXX

COORDINATORE Prof. Dello Sbarba Persio

Deaminases and beyond: Pathology, Physiology and Biotechnology.

Settore Scientifico Disciplinare MED/04

Dottorando
Dott. Munagala Uday

Tutor
Dr. Silvestro Conticello

Coordinatore
Prof. Dello Sbarba Persio

Anni 2014/2017

Per Amma, Nana e Tata...

Preface

My thesis is structured in three parts, where I report diverse investigations of deaminases. First, I will focus on investigating the role of AID/APOBECs family of deaminases as inducers for chromosomal instability. I shall display the ability of these deaminases causing chromosomal breaks using cytogenetic tools. Then I shall describe a quantitative assay developed to establish micronuclei frequency, which can be used as a tool to understand chromosome instability: induction of the AID/APOBECs in the human fibrosarcoma cell line (HT 1080) bearing a human artificial chromosome yielded elevated levels of micronuclei. Through the use of this cellular system, I provide evidence for the association of DNA repair associated to AID/APOBECs expression with chromosomal instability.

In the later part of my thesis, I will describe my work on ADAR2, the RNA editing enzyme and study its potential to act on the DNA. I will show my attempts to establish ADAR2 as base editor (new genome editing tool) to induce single dA to dG mutations to a specific sequence. Here, I will explain the strategy and construction of CRISPR/dCAS9 fusion with the deaminase domain of ADAR2(BE-ADAR2) coupled with the development of a fluorescent reporter assay to test the activity. This will be followed by development of mutants of ADAR2 that are able to target DNA. I shall describe how I developed the mutants using bacterial mutator assays. These mutants will serve as better candidates than the human ADAR2 for the use a base editor targeting A:T pairs, with potential usage in biotechnological applications such as gene therapy, antiviral treatment and cancer therapy.

The final part of my thesis consists of a manuscript I collaborated on, which has been submitted for publication.

Abbreviations

Aicda	Activation-Induced Cytidine Deaminase (gene)
AID	Activation-Induced Cytidine Deaminase (protein)
Ag	Antigen
alt-NHEJ	Alternative NHEJ
AP	Apurinic/Apyrimidinic
APC/C	Anaphase promoting complex/cyclosome Asp Abnormal spindle protein
APE	AP Endonuclease
APOBEC	apolipoprotein B mRNA editing catalytic polypeptide-like
ATM	Ataxia telangiectasia mutated Aur Aurora kinases
BCR	B-cell Receptor
BER	Base Excision Repair
BFB	Breakage-Fusion-bridge BRCA1 Breast cancer 1
bp	base pair
BRCA	Breast Cancer
Cas	CRISPR-associated Nuclease
CDK1	Cyclin B dependent kinase 1 CDKs Cyclin dependent kinases
CDR	Complementarity Determining Region
CENP	Centromere linked motor protein
CHK	Checkpoint kinase
CIN	Chromosomal instability
CLP	Common Lymphoid Precursor
c-NHEJ	Canonical NHEJ
CRISPR	Clustered regularly interspaced short palindromic repeat
CSR	Class Switch Recombination
DAPI	4’,6-Diamidino-2-phenylindole
DNA	Deoxyribonucleic Acid
DDR	DNA Damage Response

DSB	Double-stranded break
DSBR	Double-stranded break Repair
DR	Direct Repair
ER	Estrogen Receptor
FACS	Fluorescence-activated Cell Sorting GC Gene Conversion
GFP	Green fluorescent protein
GG-NER	Global Genome NER
HR	Homologous Recombination
HRR	Homologous Recombination Repair
IL-4	Interleukin 4
IR	Ionizing Radiation
IS	Insertional Sequence
kb	kilo base-pair
LPS	Lipopolysaccharide
MRN	Mre11-Rad50-Nbs1
MMEJ	Microhomology-mediated End Joining MMR Mismatch Repair
MSI	Microsatellite instability
MT	Microtubules
NAE	Nucleic Acid Editing
NBS	Nijmegen Breakage Syndrome
NER	Nucleotide Excision Repair
NHEJ	Non-homologous End Joining
NIMA	Never In Mitosis Gene A
NOS	Nitric Oxide Synthase
PARP	Poly(ADP-ribose) Polymerase
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PNK	Polynucleotide kinase
PNKP	Polynucleotide kinase/phosphatase
Rad54	DNA repair and recombination protein RAD54
RAG	Recombination Activating Genes (protein)

RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
ROS	Reactive Oxygen Species
RT	Room temperature
SHM	Somatic Hypermutation
SSB	Single-stranded break
TC-NER	Transcription-Coupled NER TCR T-cell Receptor
TDS	Translesion DNA synthesis
TE	Transposable Element
UNG	Uracil DNA glycosylase
ZFN	Zinc-finger Nuclease

Summary

Genetic abnormalities are the main cause of variability within organisms and are the basis for genetic diseases and cancers. DNA damage caused by AID, APOBEC1 and APOBEC3B was assessed by γ-H2AX immunofluorescence, where a strong activation of this marker of DNA damage was observed, thus, inciting DNA disruption and genomic damage. Furthermore, the ability of these cytidine deaminases to induce chromosomal alterations was analysed in 50 individual metaphase spreads by scoring for chromosome aberration. There was an increase in the number of aberrations. Notably, in case of AID, the number of aberration were three times higher than in the control cells, while in the case of APOBEC1 and APOBEC3B, there was a two-fold increase. This clearly indicates the role of AID, APOBEC1 and APOBEC3B causing DNA damage accounting to chromosomal instability. Micronuclei (MN) are an effective cellular indicator of CIN, and elevated frequencies of MN are observed in most solid tumors and pre-neoplastic lesions. In order to investigate the role of the AID/APOBECs as inducers of CIN we are using a quantitative assay for chromosome mis-segregation developed by exploiting the human artificial chromosome (HAC) present in human fibro- sarcoma HT1080 cells. In these cells, the HAC kinetochore can be conditionally inactivated, thus leading to formation of micronuclei. Interestingly, we observed elevated levels of formation of micronuclei after transient expression of AID, APOBEC1, APOBEC3A and APOBEC3G in presence of CytochalasinB (CytoB) an actin polymerization inhibitor. Remarkably, even in absence of co-treatment with CytoB, elevated levels of MN were observed especially in the case of AID and APOBEC1. To further verify whether the induction of C to U changes and the downstream activation of the DNA repair pathways, we inhibited Uracil-DNA glycosylase (UDG) using a bacterial UDG inhibitor (UGI). Indeed, a diminution of the levels of MN formation in cells expressing AID and APOBEC1 in the presence of UGI was readily observed, indicating that MN formation is indeed triggered downstream to the activation of the DNA repair pathways. Considering that the AID/APOBECs have been associated to the onset of cancer through their ability to mutate DNA, our finding shows another possible cancer-inducing effect of these

deaminases: their ability to induce chromosomal instability. Recent advances in targeting genomic sequences has revolutionized the genome editing field. Using customized nucleases serves as an ideal platform for targeted deletions, insertion in a broad range of organisms and cell types, but it is less ideal for obtaining changes at specific nucleotides. The fusion of AIDAPOBEC's to genome editing tool such as Cas9 and TALE has allowed the development of a base editor that aims at avoiding formation of double strand breaks and it offers a new perspective for gene editing. The limit of such base editor is that its targeting is restricted to C•G base pairs. I thus aimed at developing a A•T base editor based on the catalytic domain of ADAR2. Adenosine Deaminases Acting on RNA (ADARs), are editing enzymes that catalyze the C6 deamination of adenosine (A) to produce inosine (I) in double-stranded RNA. A-to-I editing can alter the stability of RNA structures and the coding of RNA as I is read as G instead of A by ribosomes during mRNA translation and by polymerases during RNA replication. Given the fact that human ADAR2 is able to act on DNA/RNA hybrids, I tried to use chimeras of n/dCas9 and the deaminase domain of ADAR2 to induce a single dA to dG mutation in fluorescent reporter, but our attempts failed. Thus, mutants of ADAR2 with a superior affinity to act on DNA were developed by screening mutagenized libraries through bacterial mutators assays. These mutants may be developed as an alternative base editor to target A:T pairs with potential usage in biotechnological applications such as gene therapy, antiviral treatment and cancer therapy

1. Introduction (Part I)

1.1 Cancer genomes: An organized chaos

Cells, the magnificent basic structural unit known to replicate independently, rightfully known as the "basic unit of life", they package their DNA in a systematic manner to protect it, and to regulate the cellular life through the activation of its functional regions. Cellular DNA is never bare nor unaccompanied by other proteins. Rather, it always forms a complex with various protein partners that help package it into the nucleus. This DNA-protein complex is called chromatin, wherein the mass of protein and nucleic acid is nearly equal. Within cells, chromatin usually folds into characteristic formations called chromosomes. Eukaryotes naturally possess multiple pairs of linear chromosomes, within the cellular nucleus, and these chromosomes have distinctive and variable forms. The cell division process promotes the organization of the replicated DNA into pairs of sister chromatids, which are pulled towards opposite spindle poles aiding equal distribution of the genetic material between the two daughter nuclei. This process plays a pivotal role in the maintenance of genome integrity as accurate segregation of chromosomes is of utmost importance. Mishaps during segregation (possibly due to errors in mitosis) results in aneuploidy and chromosomal instability (CIN) (Geigl *et al.*, 2008; Crasta *et al.*, 2012).

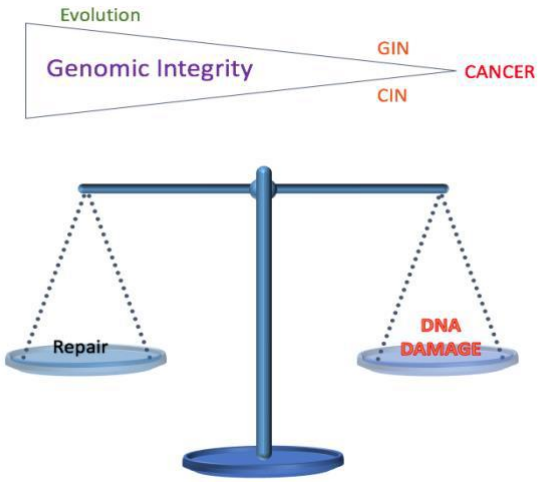


Figure 1. The balance of DNA damage and DNA repair determines the stability of a genome. As a source of evolution, mutations are beneficial for evolution, they may arise from impairment of DNA repair pathway or genotoxic stress from cellular processes (CIN, GIN). Thus, DNA repair essentially maintains the genomic integrity.

In facts, as genome are reshaped through evolution, Cancer ironically harnesses all the aspects characteristic of the evolutionary processes as cells over a period of time are mutated leading to particular traits favorable to the growth of tumor (Merlo *et al.*, 2006). Such genomic diversity bore by the cancer cell, presents itself as changes ranging from a single nucleotide to large scale alterations, primarily due to Genomic (GIN) and Chromosomal instability (CIN) (Lengauer, Kinzler and Vogelstein, 1998), fueling cellular transformation and tumor progression *(Figure 1)*. Unusual mitoses in cells from carcinomas were initially studied back in the 19th century when Von Hansemann performed early experiments to comprehend cellular transformation. These grandiose observations were a stepping stone for Theodor Boveri, and led him to propose that anomalous mitosis leads to mis-segregation of chromosomes resulting in aneuploidy, thereby promoting tumor (Boveri, 1914). Chromosome mis-segregations have been held accountable for increased mutation rate possibly due to DNA damage and repair, when this is coupled with substantial chromosomal rearrangements and aberrations it promotes aneuploidy (McGranahan *et al.*, 2012).

Genomic instability (GIN) is primarily a phenotype by which there is a high predisposition for genomic alterations. GIN, a cellular state characterized by mounting incidences of accumulated genetic alterations and a consequence of mutations affecting pathways involved in replication fidelity, cell cycle progression, checkpoint control, chromosomal segregation and repair of sporadic DNA damage. *(Figure 2*). (Gordon, Resio and Pellman, 2012).

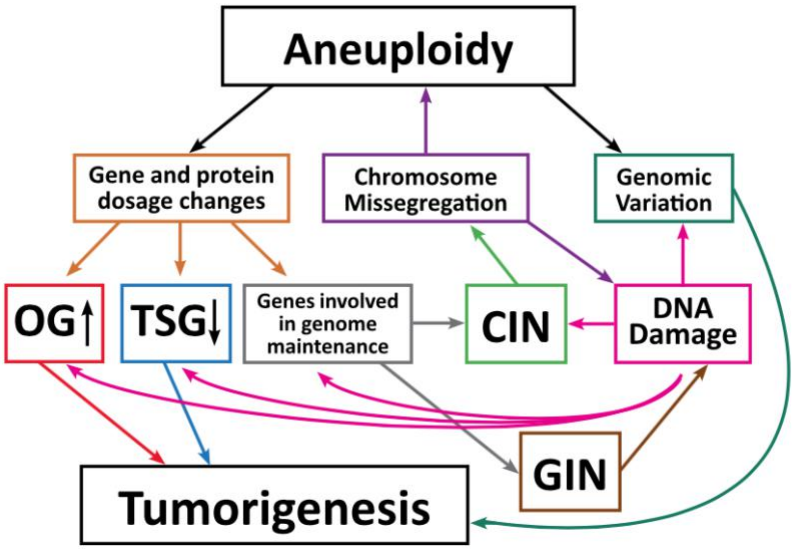


Figure 2. Tumorigenesis in an intertwined network where aneuploidy, CIN and GIN are encompassed together (figure adapted from Giam and Rancati, 2015).

1.2 Genomic instability (GIN)

Cells are primordially designed to divide and distribute an exact copy of their genetic material to the daughter cells, errors during cell division where the parental cells fail to accurately duplicate leads to genomic instability (GIN). A peculiar pattern is seen in almost all cancers associated with GIN, from minute mishaps in the DNA to errors occurring on whole chromosomes.

GIN is a rather general portrayal of the modifications/changes at the genetic level. It is an evolving hallmark in the cancer paradigm, due to it as rather evasive functionalities towards tumorigenesis. GIN is classified into:

I) Changes at ***nucleotide level*** base substitutions, deletions or insertions of nucleotides altering gene function leading to nucleotide instability (NIN). These are an outcome of faulty DNA repair pathways such as base excision repair (BER) and nucleotide excision repair (NER). E.g.: Mutations occurring in the *PIK3A* gene, resulting in breast cancer.

II) At the ***genomic*** level, alterations in the DNA repair pathway spawn microsatellite instability where short DNA repeats are scattered all over the human genome.

III) DNA methylation or gene amplification can result in epigenetic changes at the ***chromatin*** level. This has been known to promote sporadic tumors.

IV) ***Chromosomally***, inaccurate mitotic inter- or intra-chromosome recombination have been observed resulting in neomorphic (gain-of-function) or hypermorphic (overexpression) mutation leading to neoplasia. Chromosomal instabilities (CIN) accountable for aneuploidy defined with altered chromosome structure and number (Jefford and Irminger-Finger, 2006) (Lengauer, Kinzler and Vogelstein, 1998)..

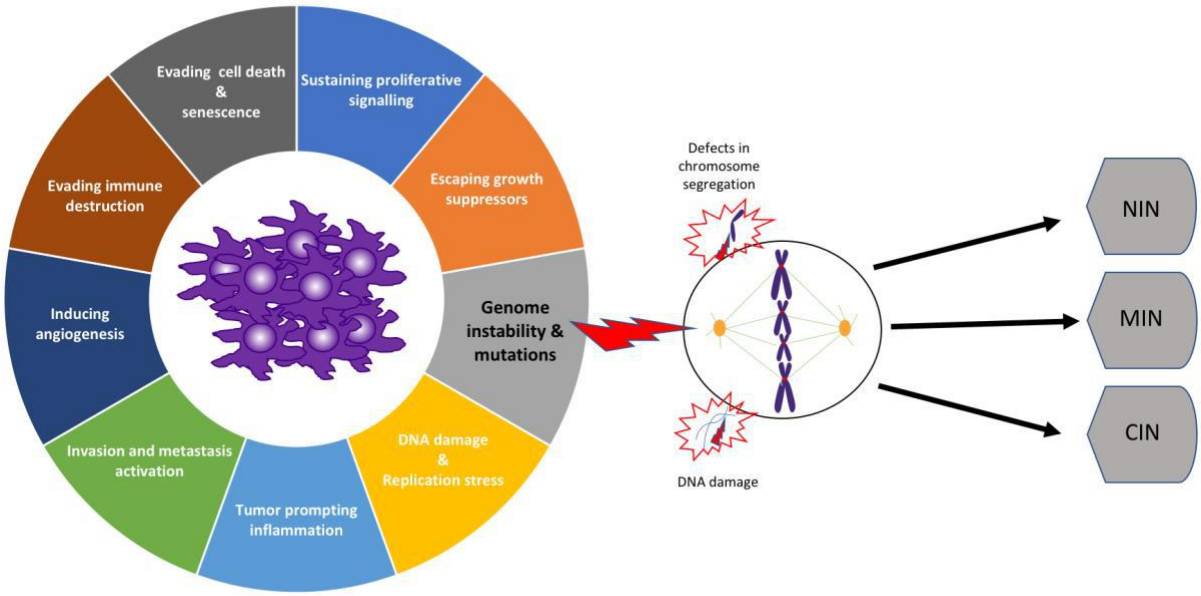


Figure 3. Genomic Instability (GIN) as an evolving hallmark of cancer: A brief schematic representation of how alterations in GIN when encounters with functions such as DNA repair, cell-cycle and mitotic checkpoints promoting fertile backdrop for tumor propagation leading to NIN, MIN and CIN.[figure has been modified from (Hanahan and Weinberg, 2011)]

1.2.1 Microsatellite instability (MIN)

Microsatellites are stretches of DNA where single nucleotides (mononucleotides), or units of two or more nucleotides (e.g., di-, tri-, tetra-, or pen tanucleotides) are repeated (*figure.4*). The stretches range from a few nucleotide constituents to hundreds of them. At least 500,000 microsatellites are present in the human genome. They are usually observed

across the genome, sometimes in intergenic regions, usually located in the introns of genes (De La Chapelle and Hampel, 2010).

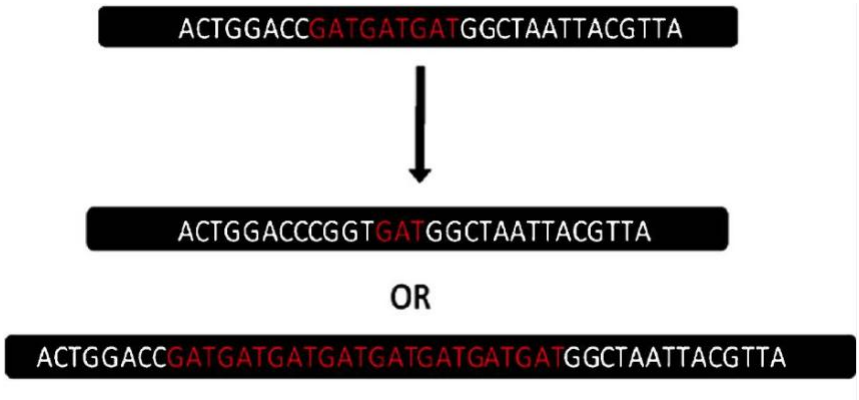


Figure. 4: MIN: Long and short stretches of microsatellites diversify throughout the genome (red). They are arisen due to defects in the MMR (Pikor *et al.*, 2013).

Microsatellite instability (MIN) is the outcome in cancer genomes, of defective mismatch repair (MMR) of the cell. MIN occurs in 15% of colorectal cancer triggered bygermline mutations in MMR genes associated with somatic inactivation of functioning allele (Salovaara R et al., 2010). Besides colorectal cancer, MIN-positive tumors has been observed across various other types of cancers such as endometrial, gastric, ovarian, gallbladder, prostate, and gliomas (Gelsomino et al., 2016). Studies have reported that oxidative damage can yield frame-shift mutations leading to MIN (Jackson, Chen and Loeb, 1998).

1.2.2 Nucleotide instability (NIN)

Nucleotide instability (NIN) embodies base substitutions, deletions and insertions of one or a few nucleotides (*figure. 5*). These occur when the repair machinery malfunctions, ensuing dramatic changes to gene structure and expression. For e.g.: a missense mutation in the *K-ras* gene is the cause for over 80% of primary exocrine pancreatic tumors (Chen *et al*., 2015). NIN is also seen in mitochondrial DNA (mtDNA), and instabilities of mtDNA has been seen in a variety of human cancers (LEE *et al.*, 2005).

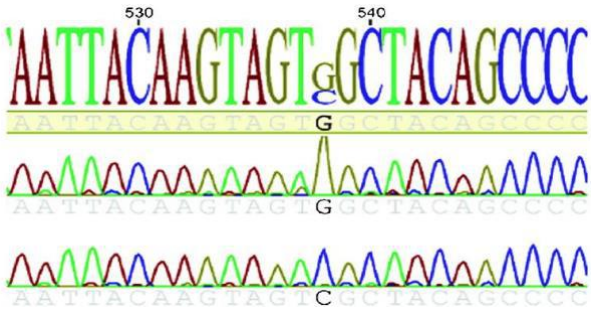


Figure.5: Nucleotide instability (NIN): Sanger sequencing in two lung cancer cell lines where G>C variant encoding a Gly>Arg amino acid are detected exhibiting NIN phenotype (Pikor *et al.*, 2013).

1.2.3 Chromosomal Instability (CIN)

Chromosomal instability (CIN) refers to an elevated rate of chromosome mis-segregation due to errors durig mitosis. Aneuploidy is the principal outcome of CIN, where gains or loss of whole chromosome are observed. Several factors have been associated to CIN: multipolar spindles, improper chromosome condensation/cohesion, inefficient chromosome compaction, defects in mitotic spindle assembly/dynamics, defective mitotic checkpoint and telomere attrition, replication stress, and improper kinetochore-microtubules attachments (Geigl *et al.*, 2008). CIN is the most prevalent form of genomic instability and has been observed in over 90 % of all malignancies. CIN, unlike MIN or NIN, is known to alter the expression of thousands of genes, thus making its outcomes for less predictable.

As MIN is characterized by defects in the DNA repair pathway, it is of prime importance that CIN and MIN are not mutually exclusive. Analysis on cell lines from the NCI 60 panel (60 cell lines derived from cancers of 9 tissue origins used for cancer research and anti-cancer drug screening) showed to have both, CIN and MIN (KM12, DU-145, SK-MEL-2, and IGROV-1 cell lines) (Roschke *et al.*, 2003). It is conceivable that these cell lines harbor mutations both on CIN-related gene(s) and MIN-related gene(s). Several studies on mitotic errors revealed a CIN-phenotype; causing mitotic arrest and mis-segregation in the chromosome (Janssen *et al.*, 2011; Crasta *et al.*, 2012). CIN and MIN displays a rather intricate interplay, where they work in different patterns triggering DNA damage (*Figure 6*).

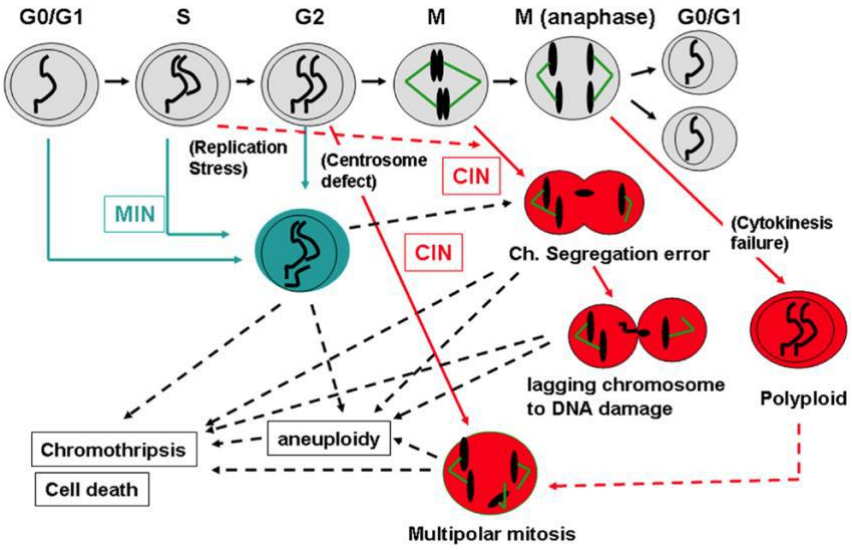


Figure 6: MIN and CIN impact on genomic instability. Cell cycle progression charts G0/G1, S, G2, and M phases (gray-shaded cells). MIN (blue shaded cells) is known to occur during G0/G1, S, or G2 phases leading to aneuploidy and chromothripsis (McGranahan *et al.*, 2012)*. CIN is caused by an event in mitosis leading to a chromosome segregation error (e.g., a kinetochore defect, a spindle challenge, a mitotic spindle checkpoint defect, a chromosome cohesion defect; shown in red)*(Rao and Yamada, 2013).

With a CIN phenotype, tumor cells are categorized depending on numerical and structural chromosome changes, where *Numerical CIN (nCIN)* is associated with gain or loss of whole chromosomes and *Structural CIN (sCIN)*, to structurally abnormal chromosomes (*Figure 7*) (Gollin, 2005).

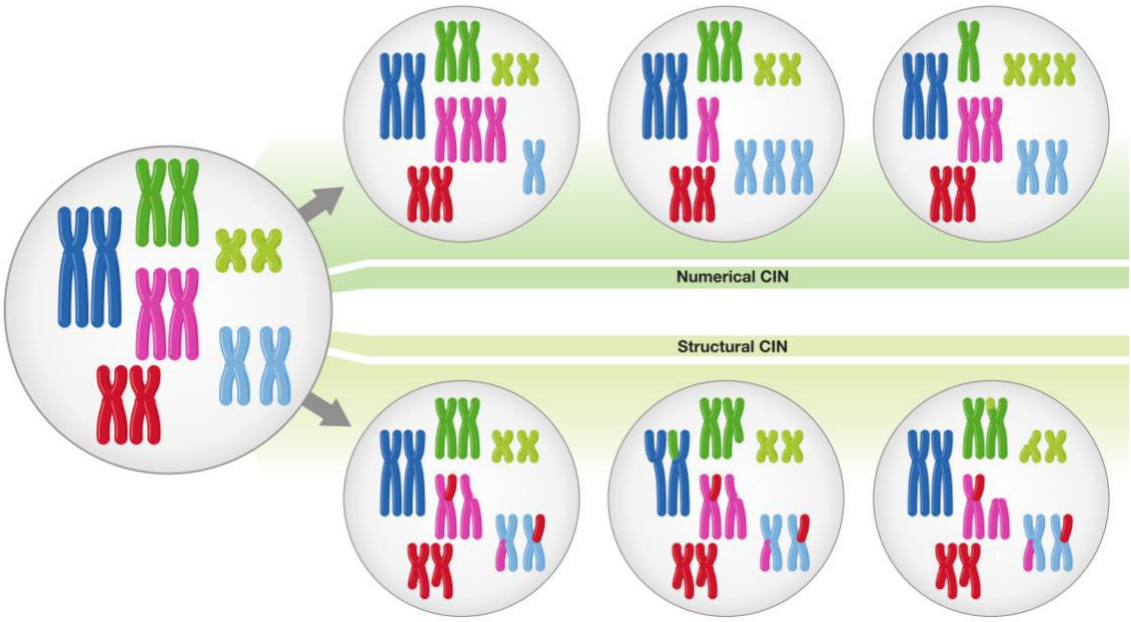


Figure 7 : Numerical and structural chromosomal instability (Bayani *et al.*, 2007)

Numerical CIN (nCIN)

nCIN is characterized by elevated levels of gains and losses of chromosomes, resulting in abnormal chromosome number or aneuploidy. These alterations can exist in a very stable form and are easily targeted, while unstable forms of aneuploidy are more common and adapts itself depending upon stress (internal and external). These have varying consequences depending upon the changes in chromosome copy number, ploidy and presence of structural aberrations. Moreover, erroneous mitotic processes involving anaphase failure, centrosome duplication or aberrant DNA repair may lead to tumors with nCIN-karyotypes (*Figure 8:a-d*). (Cahill *et al.*, 1998; Nowak *et al.*, 2002; Bayani *et al.*, 2007).

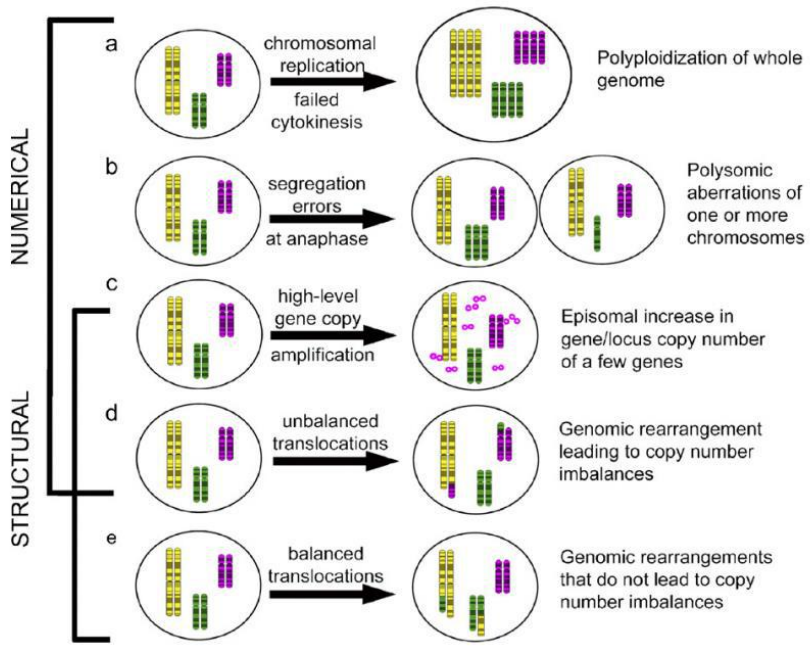


Figure 8 : Factors leading to Numerical and structural chromosomal instability: The normal cell is depicted with three chromosome pairs (yellow, red and green): (a) failed cytokinesis; (b) segregation errors during anaphase leading to formation of monosomic and trisomic chromosomes; (c) gene amplifications resulting in structural changes; (d) translocations causing unbalanced structural alteration; (e) balanced translocations with no changes in copy number variations(Bayani et al., 2007).

Structural CIN (sCIN)

Translocations, insertion, deletions and amplification of DNA lead to structural unbalanced chromosomal aberrations. These aberrations involve gain or loss of small regions of chromosome and the phenomenon is primarily referred as Structural CIN (sCIN). sCIN is largely seen in many cancers. These alterations are the cause for overexpression of an oncogene or repression/knockdown of tumor suppressor genes, as genes are either duplicated, fused amongst them or lost altogether (Mitelman, Mertens and Johansson, 1997; Cahill *et al.*, 1998; Bakhoum and Compton, 2012). Impairment in the double-stranded DNA repair machinery is known to aid sCIN as well (Venkatesan, Natarajan and Hande, 2015). Double strand breaks (DSB) generates non-specific chromosomal fusions as non-homologous end joining (NHEJ) comes into play, leading to sCIN (*figure 8 c-e*) (Natarajan and Palitti, 2008). Fusions among dysfunctional telomeres is the major cause of di-centric or ring chromosomes where formation of chromatin

bridges during anaphase occurs. These unstructured bridges result in cytokinesis rupture (McClintock, 1941; Gisselsson *et al.*, 2001). The Breakage-Fusion-bridge (BFB) are associated with dysfunctional telomeres, replication stress (Bristow and Hill, 2008; Burrell *et al.*, 2013) causing instability and heterogeneity in tumor cells. High grade tumors have always been linked to elevated frequencies of structural chromosomal aberrations (Mitelman, Mertens and Johansson, 1997; Nishizaki *et al.*, 1997).

1.3 Mechanisms driving CIN

A rather distinctive feature of CIN is the abnormal distribution of replicated chromosomes to daughter cells during mitosis. These mis-segregation has been known to cause genetic heterogeneity among tumor cells at every cell cycle checkpoint (Cimini *et al.*, 2001; Burrell *et al.*, 2013). One of the most prominent segregation errors in CIN-cancer cells are lagging chromosomes and chromosomal bridges (Thompson and Compton, 2011).

Chromosomes that fail to attach to the microtubules are unable to segregate at anaphase and are left behind (Thompson and Compton, 2011); this failure leads to merotelic attachment with the kinetochore of the chromatids(either one of them) attached to both spindle poles. Kinesins come into action to reduce the stability of the microtubules at the kinetochore and prompts release of merotelic attachments to correct its orientation. This spindle assembly is the crucial point where any impediments in mitosis progression, from metaphase to anaphase can be regulated (*figure 9A*). When mis-regulated, it results in checkpoint errors{*Figure 9B*;(Hanks *et al.*, 2004; Holland and Cleveland, 2012)}, erroneous sister chromatid cohesion (*Figure 9C*) (Iwaizumi *et al.*, 2008; Barber *et al.*, 2008; Zhang *et al.*, 2008, 2015); centrosome amplification (*Figure 9D*)(Ganem, Godinho and Pellman, 2009) and ultimately dysfunctional mitotic spindle apparatus (*Figure 9E*)(Janssen *et al.*, 2011).

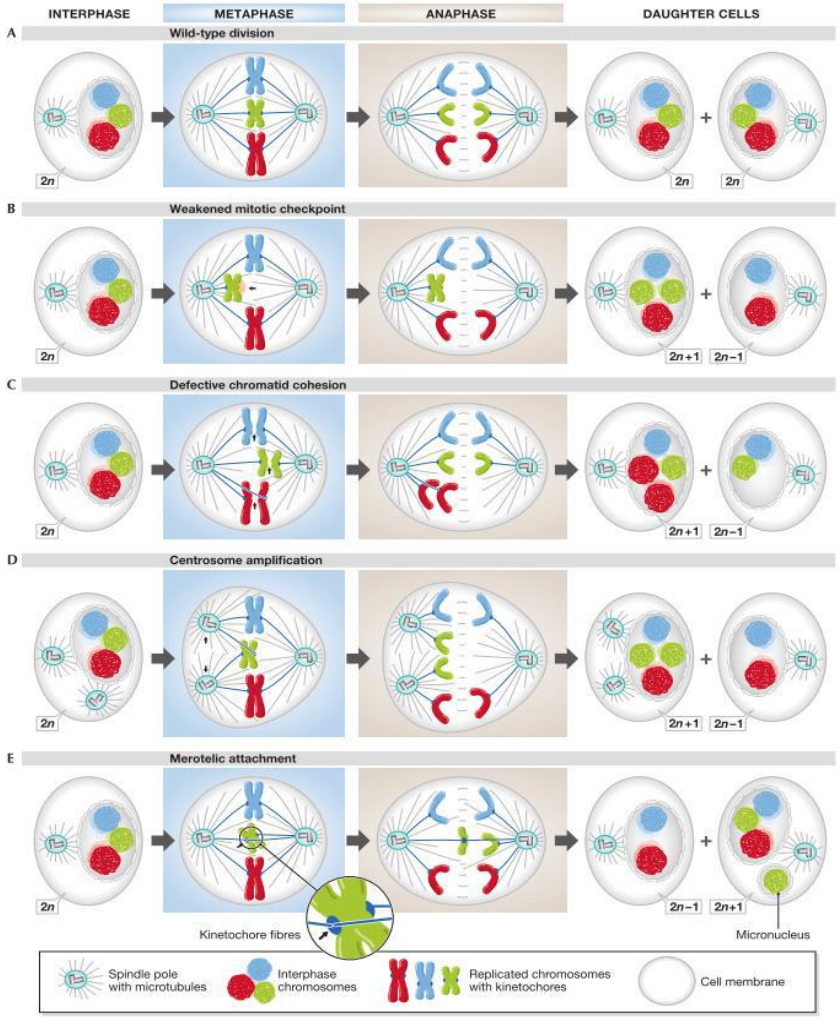


Figure.9: Mechanisms of chromosome mis-segregation in mitosis. (A) Normal cell division. (B–E) Mitotic errors resulting in aneuploid cells (McGranahan *et al.*, 2012).

Cancer cells are able to tolerate high levels of CIN as cell cycle checkpoints are unable to arrest themselves, resulting in loss of genomic integrity.

1.4 Micronuclei (MN)

Micronuclei (MN) are small, extra-nuclear chromatin bodies surrounded by a nuclear envelope. MN mainly originates from acentric chromosome fragments or whole chromosomes that fail to attach aptly to the spindle fibres and fails to segregate to daughter cells during the telophase (*figure 10*)(Savage, 1988). It has now been widely

accepted that micronuclei (MN) are an effective cellular indicator of CIN as high frequencies of MN are seen in most solid tumours and pre-neoplastic lesions that display CIN (Gisselsson *et al.*, 2001). MN are broadly assessed as an indicator of in vivo exposure for genotoxins (Bonassi *et al.*, 2011)

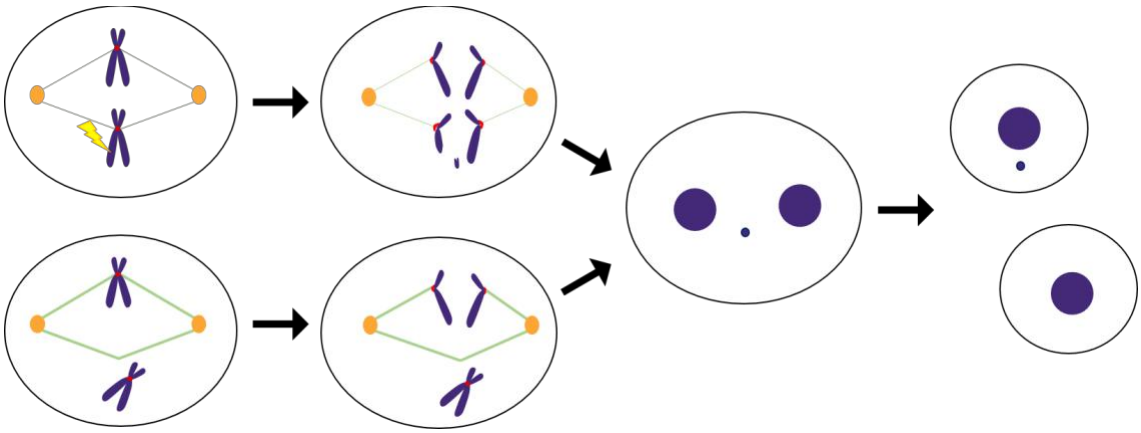


Figure 10: Schematic representation of micronucleus formation. Micronuclei are small particles in the cytoplasm consisting of acentric fragments of chromosomes or entire chromosomes, which do not integrate in the daughter nuclei during the cell division.

How MN is involved in the cellular transformation is still undetermined and there are several on-going investigations to address the fate of MN and MN-cells. Hypothetically there are four outcomes for the fate of MN: Degradation, Reincorporation, Extrusion and Persistence (*Figure 11 a*) (Hintzche *et al.*, 2017). Furthermore, for MN-cells, there are two additional outcomes: premature chromosome condensation or chromothripsis and eradication of MN by apoptosis (*Figure 11b*).

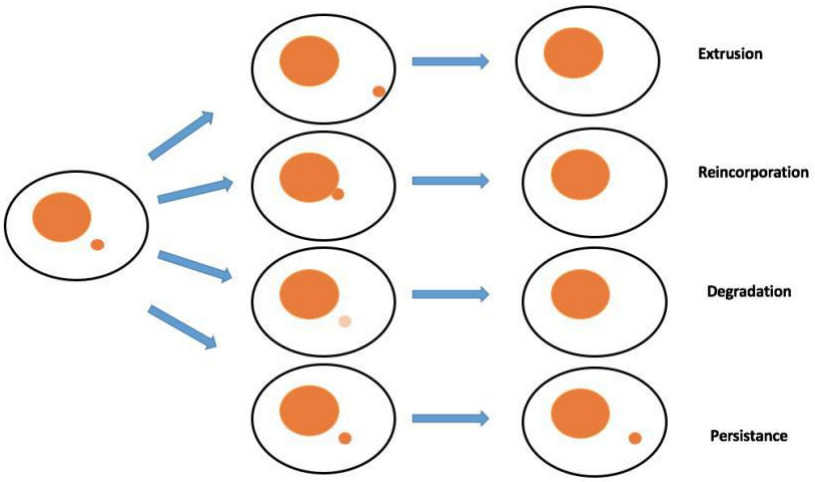


Figure. 11a: Fate of Micronuclei.

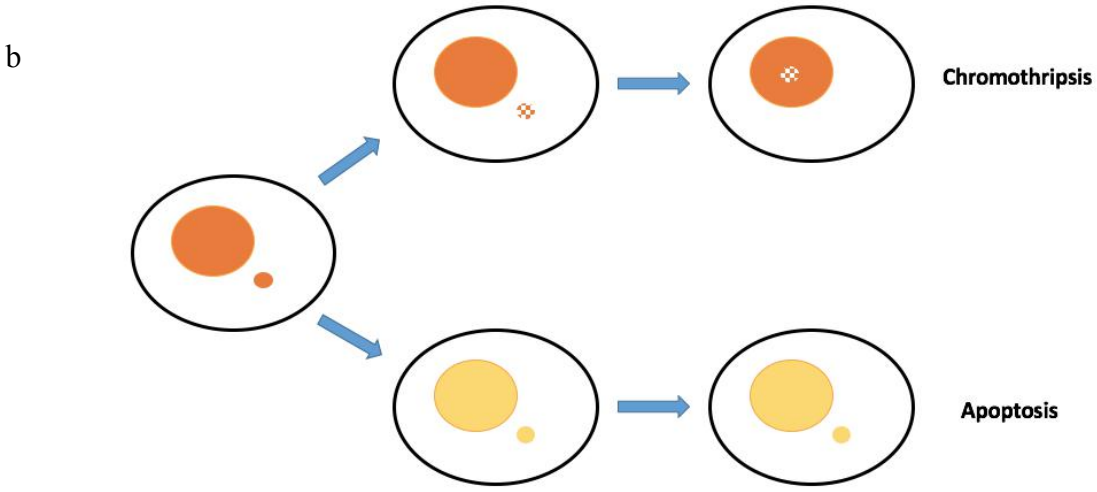


Figure 11b: Schematic overview of cellular fate MN cells.

Degradation of micronuclei leads to destruction of MN DNA and leads to its disappearance. Several works were undertaken to observe this phenomenon especially in aneugen-induced MN in RPE-1, U2OS and HCT116 cells (Huang *et al.*, 2012) but were unsuccessful to observe any degradation process. They underlined a probable mechanism of degradation might indeed be the outcome of the disruption of the MN envelope (Hatch *et al.*, 2013), highlighting that more than half of all spontaneously occurring MN are disrupted during cell cycle. These ruptures were also observed in the main nucleus where the MN-envelope failed in its capacity to repair (Hatch *et al.*, 2013). MN degradation was also found in primary fibroblasts when treated with clastogen and aneugen reagents,

where MN became p-53 positive over the time course hinting MN degradation by nucleases (Granetto *et al.*, 1996). Moreover, in irradiated fibroblasts, MN positive cells suggested apoptosis-like degradation process similar to what has been observed in chromothripsis (Terradas *et al.*, 2009).

Premature chromosome condensation(PCC) is known to occur in MN**.** It is a result of an asynchronous relationship between the MN and the main nucleus, as the chromatin condenses prior to the chromosomes of the main nucleus (Cancer *et al.*, 2014). PCC and MN formation had been a subject of debate lately until recent work by Pellmans's lab showed that ***chromothripsis*** is an outcome of MN, where the chromosome is shattered as a result of DNA breaks, owing to curtailed DNA synthesis. This leads to massive genomic rearrangements restricted to specific chromosomes (Zhang *et al.*, 2015). Several studies highlighted the role of MN using cellular models where MN are generated by nocodazole-induced mitotic arrest, or where conditional inactivation of kinetochore was used in association with a human artificial chromosome (Nakano *et al.*, 2008; Crasta *et al.*, 2012). In cell cycle studies, no DNA damage in MN was detected in G_1, while damages were acquired in G_2 and S phase. The MN replication was asynchronous with erroneous DNA replication and repair machineries. Moreover, it was observed that 10% of the pulverized chromosomes were micronucleated and the shattered chromosomes exhibited delayed replication (Crasta *et al.*, 2012). In human primary cells, several works link PCC with MN, for e.g.: in lymphocytes, PCC was seen in chromosomes and chromatids within MN (Terzoudi *et al.*, 2015).

MN formation and apoptosis have a rather complicated interplay (Whitwell *et al.*, 2015). Primarily, it still needs to be classified whether MN formation is the initial step of apoptosis or whether they are mutually exclusive. Initial studies on irradiated murine cells showed MN formation and exhibited a distinct morphology between MN and early stage apoptosis (Abend, Frombeck and Van Beuningen, 1999). The presence of micronuclei itself might represent a potential signal for apoptosis. Moreover, cells with centromere- positive MN were observed in the apoptotic fraction than in the viable ones. This implies that separation of whole chromosomes into micronuclei possibly signals for

cellular apoptosis, presenting a pathway for ***elimination*** of potentially aneuploid cells (Decordier *et al.*, 2002).

Reincorporation of MN is known to ensue during interphase, but it occurs more frequently during the next mitosis (Terradas *et al.*, 2009; Kirsch-Volders *et al.*, 2011). This puts MN onto the pedestal as not only an indicator for GIN, but also as a likely source when reincorporated into the main nucleus (Terradas *et al.*, 2012). MN in irradiated nasopharyngeal carcinoma cells were examined to elucidate the fate of MN cells. Cells were assayed for their proliferation and were assigned based on their ability to divide. Most of the MN positive and MN negative cells underwent successful cell division, besides the fact that MN positive cells had an increased percentage of cells in cell cycle arrest or cell death. Moreover, following the fate of MN cells it was observed that some were micro nucleated after mitosis while some were MN-free, suggesting that reincorporation is not always a persistent phenomenon (Huang *et al.*, 2011). A novel study was performed where MN reincorporated cells were analyzed using a technique called as Look-seq, a combination of cell imaging with single cell sequencing. This enabled them to follow the MN and help them analyze previous micro nucleated chromosome after reincorporation. They found massive chromosomal rearrangements exclusively in cells containing chromosomes originating from MN, thus proving a direct link to chromothripsis (Zhang *et al.*, 2015).

Extrusion of a MN from the cell is one of the many possibilites known to occur (Terradas *et al.*, 2009; Kirsch-Volders *et al.*, 2011). To understand this phenomenon, studies were performed to understand membrane blebing (protrusions of a cell membrane) and its impact on extrusion of MN; in the context of MN formation containing double minuts (DMs) during S-phase in tumor cells. Aggregation of DMs was observed during MN formation and they were found to be localized around the cytoplasmic membrane protrusions (blebbing), suggesting it as a mechanism by which MN is extruded from the cell (Utani, Okamoto and Shimizu, 2011).

What happens when MN ***persists*** after the next mitosis is an interesting question to address as in this case, degradation, extrusion and reincorporation does not occur. Some studies suggest no particular offset with MN persistance (Kirsch-Volders *et al.*, 2011; Utani, Okamoto and Shimizu, 2011; Bhatia and Kumar, 2013). Live imaging techniquies were principally used to follow the micronucleated cells, in human breast cancer cell line, MN- interphase cells were analyzed for presistance and found delayed cell cycle progression (Yasui *et al.*, 2010). Intrestingly in colorectal cancer cells, when the MN cells were looked at, the cells were found to enter mitosis rather normally, possbily suggesting that proliferation is not necessarily inhibited by enclosure of a chromosome into a MN (Huang *et al.*, 2012).

The Nuclear envelope (NE) is a vital component of MN. Upon NE reassembly, the lagging DNA is expelled from the nucleus and becomes encapsulated in its own NE, complete with a nuclear lamina and NPCs (Walker *et al.*, 1996). As MN has its own NE, the defects that derive from non-mitotic breakdown of the micronucleus NE leads to disruption of the micro-nuclear nucleoplasm transport (Crasta *et al.*, 2012) and eventually to loss of compartmentalisation, a key factor affecting the nuclear volume. A disrupted nucleoplasm transport results in a reduced levels of DNA repair factors, thus leading to impairment in the DNA repair capacity and localised accumulation of DNA damage (Hatch *et al.*, 2013). The accumulation of DNA damage in the MN has also been associated with replication-associated damage as it diminishes the recruitment of replicative DNA helicase and replication initiation factors (Crasta *et al.*, 2012). Replication associated damage might occur from replication fork collapse and/or processing of replication intermediates by cytoplasmic nucleases.

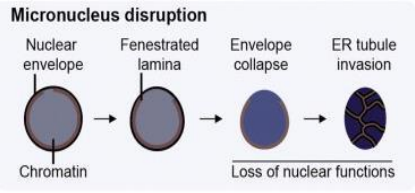


Figure 12: MN disruption with NE assembly(Hatch et al., 2013)

Funtional aspects of MN is a very important attribute to look into, considering thier invovlement in replication, transcription and DNA repair. Reports on MN containing

whole chromosomes showed to have competant trancriptional activity (Shimizu, Misaka and Utani, 2007). DNA damage response in micronuclei could help us unearth several meachanisms, as very little information is known about the activation of DNA repair mechanisms after damage. Initial reports showed abundance of p53 in MN once DNA breaks were induced by clastogenic agents(Granetto *et al.*, 1996). Elevated Rad51 expression along with replication protein A (RPA) were seen in radiation-induced MN, insinuating that these proteins were encapsulated within MN alongwith damaged DNA (Haaf *et al.*, 1999). Conversely, it was also reported that Rad51 and RPA might be recruited to MN with damaged DNA where there involved with DNA repair (Terradas *et al.*, 2009).

Recent advances have thwarted primal mis-conceptions of micronuclei, as they have evolved from passive indicators of DNA damage to active players in the formation of DNA lesions, thus unscrambling unforeseen roles of micronuclei in the origins of chromosome instability (Terradas, Martin and Genesca, 2016).

1.5 CIN: interplay with other instabilities

A combination of numeral and structural aberrations shapes a cancer cell. Current studies are focused on evidencing how various types of instabilities can influence one another in tumorigenesis.

There have been profound studies implicating GIN driving CIN. Double strand breaks (DSBs) are generated as a result of premitotic errors from DNA replication fork collapse leading to rearrangements and chromosomal breakage during aberrant DNA repair (Negrini, Gorgoulis and Halazonetis, 2010). These stress induced defects may result in acentric chromosome fragments which entails to chromosome bridges at anaphase (*Figure 13A*)(Gisselsson, 2008). These errors have been observed in colorectal cancer cells with CIN during anaphase. Moreover, DNA replication was found to occur at common fragile sites during mitosis, as an attempt to achieve complete replication prior to cell division, and botches in the machinery results in missegregation leading to onset of CIN (*Figure 13B*) (Minocherhomji *et al.*, 2015).

Telomeres are known to be sensitive to replication stress due to their association with the DNA damage machinery. Dysfuntional telemerases have been linked with genome-

doubling (*Figure 13A)* which generates tetra-ploid cells, a known CIN-precursors (Meeker *et al.*, 2002; Davoli and de Lange, 2012). Besides, dicentric chromosomes are prone to highly localized mutagenesis; associated with the phenomenon 'kateagis', demonstrated by the presence of the mutational signature of APOBEC-family of deminases (Stephens *et al.*, 2011; Leibowitz, Zhang and Pellman, 2015; Maciejowski *et al.*, 2015). Additonally RPA, suggesting single strand DNA exposure(APOBEC substrates), was detected around the chromosome bridges and was actively being processed by exonucleases (Maciejowski *et al.*, 2015).

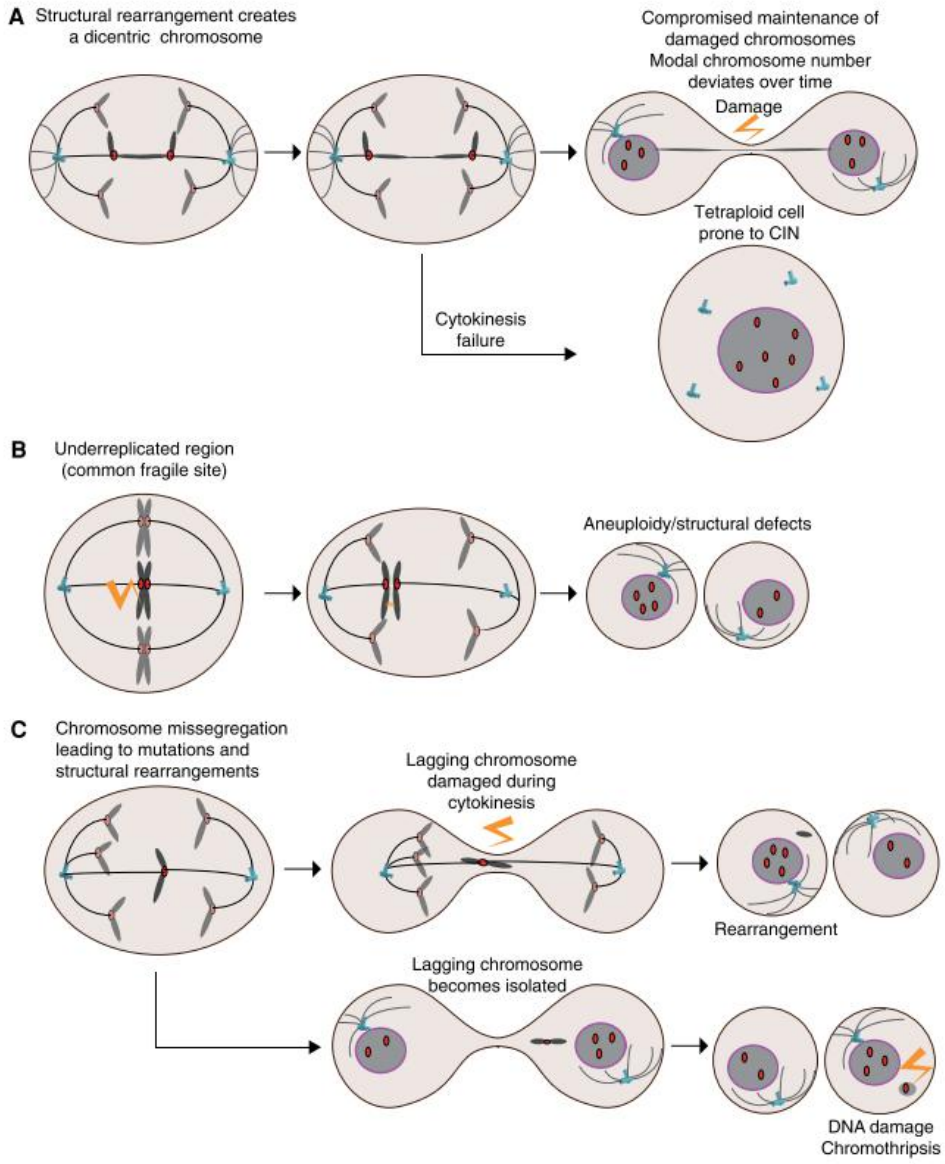


Figure.13:)CIN ad GIN . (A) Replication stress as a primary cause for structural rearrangements resulting bridges lacking nuclear membrane, leaving it exposed to extensive DNA damage and

rearrangements (B). Non-disjunction and aneuploidy as a result of DNA replication collapse. (C)Chromosome mis-segregation as a result of lagging chromosome entrapment. On the other hand, lagging chromosomes become isolated in MN, renders it susceptible to extensive DNA damage and chromothripsis. (Sansregret and Swanton, 2017).

Mis-segregation of chromosomes often leads to isolation of a few of them into MN which are prone to DNA damage as nuclear envelope collapses (*Figure 13C)* (Crasta *et al.*, 2012; Hatch *et al.*, 2013). DNA from MN once incorporated back into the main nucleus and transmitted to daughter cells during the following cell division, displays an intricate mechanism where a single chromosome segregation error causes mutations where no prior defects in genome maintainence were reported, thus sparking as a precursor of GIN (Crasta *et al.*, 2012; Zhang *et al.*, 2015). This higlights one of the many facets of CIN where it renders chromsomes prone to GIN (Hatch *et al.*, 2013).

1.6 Aneuploidy

Convincing evidence shows how intertwined aneuploidy and cancer are. Findings on 3,131 cancer copy-number profiles for somatic copy number alterations (SCNAs) in cancer where they analyzed for its prevalence and found that one-fourth of the genome was affected by SCNAs of aneuploidy (Beroukhim *et al.*, 2010). An analysis of the direct consequences of aneuploidy is infrequently performed, as aneuploidy is always accompanied by CIN or GIN (Cimini *et al.*, 2001). Genomic alterations, such as point mutation, rearrangements involving insertions, duplications, inversions, amplifications and translocations are the core aspects of cancer cells where the acute and chronic effects of aneuploidy are observed simultaneously (Durrbaum and Storchova, 2015).

Several models are being used to elicit the mechanism of aneuploidy (*figure 14*). A Cellular model with targeted mutation interferes with chromosome segregation producing a CIN phenotype (Sotillo *et al.*, 2007), aiding to analyze in changes and the physiological effects of aneuploidy. Another model to study aneuploidy is the Microcell-mediated chromosome transfer (MMCT) technique which allows the transfer of a specific chromosome from a donor cell line into a host cell line (Meaburn, Parris and Bridger, 2005).

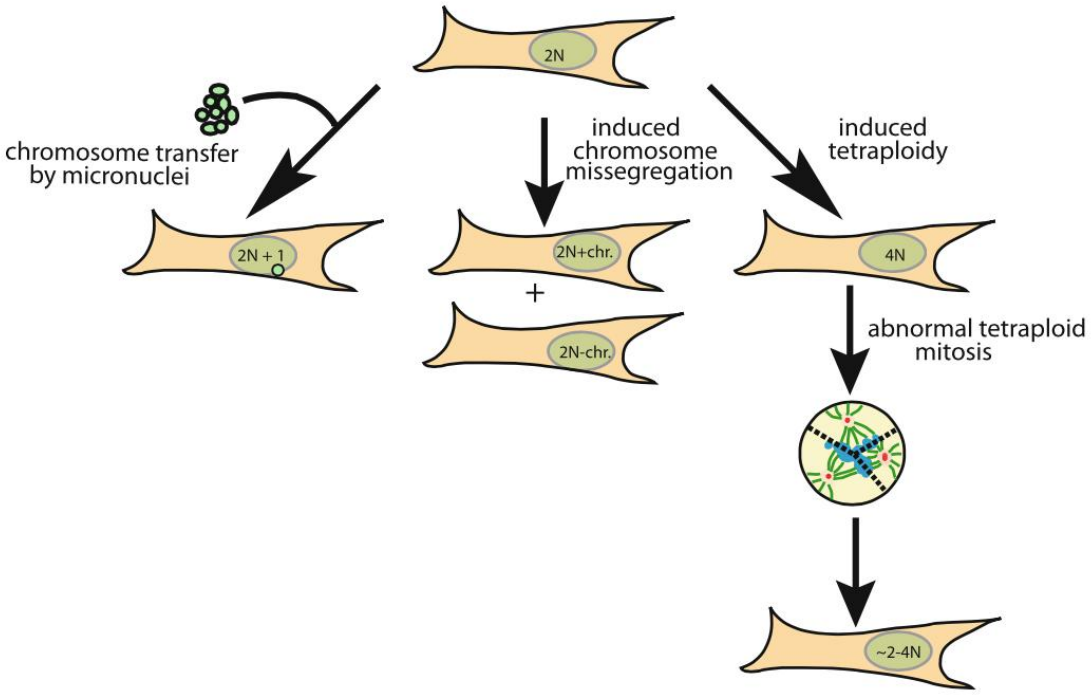


Figure.14 : Model to study aneuploidy: Mammalian cells lines with defined aneuploidy using micronuclei-mediated chromosome transfer (Ghadimi and Ried, 2015).

Cell responses to aneuploidy at the transcriptional or proteomic level needs to be determined, as it helps us comprehend gene expression imbalances caused by aneuploidy. Whether transcriptional changes affect copy number alteration of the DNA or if gene expression effects are beyond the chromosomes needs to be addressed. Gene expression has been largely investigated with copy number changes including trisomy's in patient samples, where trisomy in chromosome 21 had correlative expression change of gene in both humans and mice (Aït Yahya-Graison *et al.*, 2007; Vilardell *et al.*, 2011). However, genome wide studies showed that 20% of the proteins encoded in supplementary chromosomes did not exhibit enhanced expression (Torres, 2015). This shows how aneuploidy may induce a widespread transcriptional response not solely limited to alteration of chromosomes. Moreover, molecular pathways in protein homeostasis maintenance have been directly associated with aneuploidy (Torres *et al.*, 2007; Gemoll *et al.*, 2013). An extra chromosome might heavily alter the proteostasis where it fills in the cellular system with proteins causing proteotoxic stress (Oromendia and Amon,

2014). This in turn might activate protein degradation mechanisms to thwart overexpressed proteins or misfolded proteins.

Depending on the cancer type, variations are seen in copy number where genes are expressed as an outcome of aneuploidy. An efficient way to correlate them is by possible use of DNA copy numbers with gene expression data where fold changes are calculated with high/low copy number of genes with relevant mRNA expression changes; or by extrapolating the average gene expression with the copy number across a chromosome arm/segment (*Figure 15*) (Ghadimi and Ried, 2015).

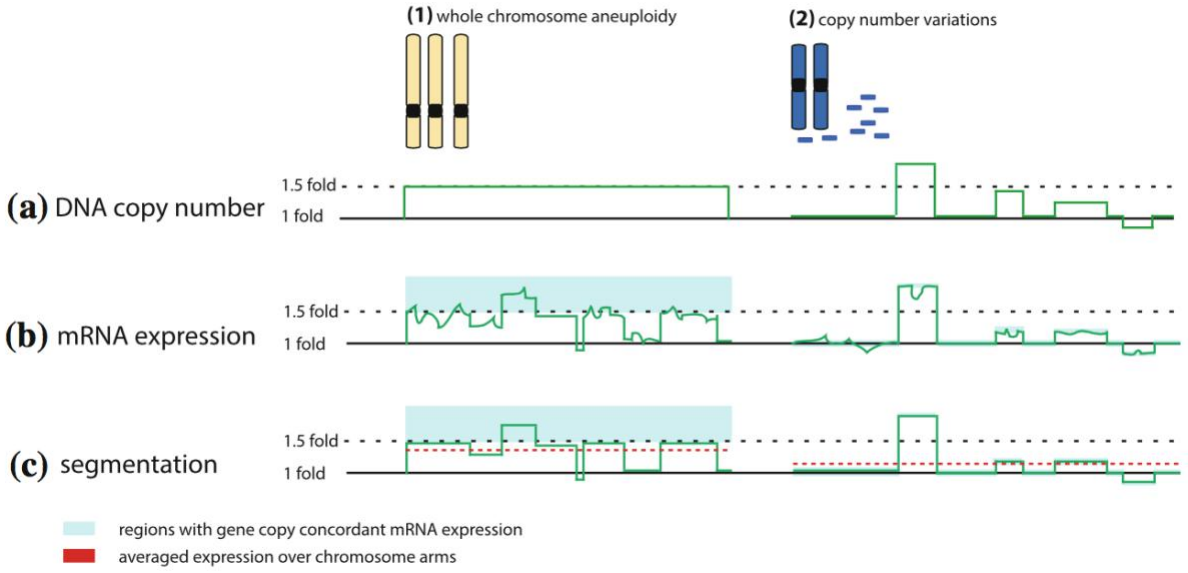


Figure 15: Correlation of mRNA and DNA copy number changes in aneuploidy. CIN or chromosome mis-segregation are the principal reasons for aneuploidy. This results in proteotoxic stress, growth defects and DNA damage with probably leading to cell cycle arrest (Ghadimi and Ried, 2015).

mRNA expression in trisomy chromosomes showed a twofold increment in the amplified regions compared to the normal diploid case , also, genes in the trisomy chromosomes presented elevated expression (Schoch *et al.*, 2005). Similar observations were also reported in breast and colon cancer (Pollack *et al.*, 2002; Tsafrir *et al.*, 2006). However, there is certain disparity, as studies from Taylor and colleagues reported the expression changes in prostate cancer and observed no distinct decrease in gene expression of all

genes in their corresponding regions (Taylor *et al.*, 2010). This suggests the role aneuploidy at the transcriptome level as indeterminate and it might be one of the factors affecting gene expression.

Thus, genomic instability, chromosomal instability and aneuploidy are distinguishing hallmarks of human cancers, the molecular basis and its stages in the cancer development has to be determined and we are just in the beginning to solving this enigma.

1.7 Genotoxic agents, DNA damage and aberrations

The genomic integrity of the cells is often subjected to physical and chemical agents which may modify bases of nucleotides and subsequently alter the genome. The human cells is subjected to at least 70,000 lesions a day (Lindahl and Barnes, 2000). Most of the lesions are single-strand DNA (ssDNA) breaks (70%), which are arise from ***exogenic*** (UV, ionizing radiation, alkylating agents etc.) (Hoeijmakers, 2001) and ***endogenic*** agents (by-products of metabolism or inflammation). The fact that genetic information , even if subjected to this astonishing daily barrage of endogenous and exogenous DNA damage, is transmitted virtually error-free highlights the strait of repair machinery acquired during evolution that helps repairing erroneous DNA and restoring genome integrity (Lindahl and Barnes, 2000; Friedberg, 2008).

Exogenous damage sources may be physical or chemical. UV radiation is a common example of a physical agent which causes covalent modifications among neighboring pyrimidine nucleotides resulting in pyrimidine dimers and photoproducts (Cadet, Sage and Douki, 2005; Pfeifer, You and Besaratinia, 2005) . Moreover ionizing radiation has both direct (damages DNA back bone, breaks in DNA) and indirect (production of ROS leading to base deamination and abasic site) effect on DNA (Jackson and Bartek, 2009). UV induced damage has a characteristic pattern, where there is prevalence of C•G to T•A mutations near two adjacent pyrimidines and a bulk of CC•CG to TT•AA double substitutions are observed as well, seen in UV-associated carcinomas (*Figure 16*) (Hendriks *et al.*, 2010; Alexandrov *et al.*, 2013).

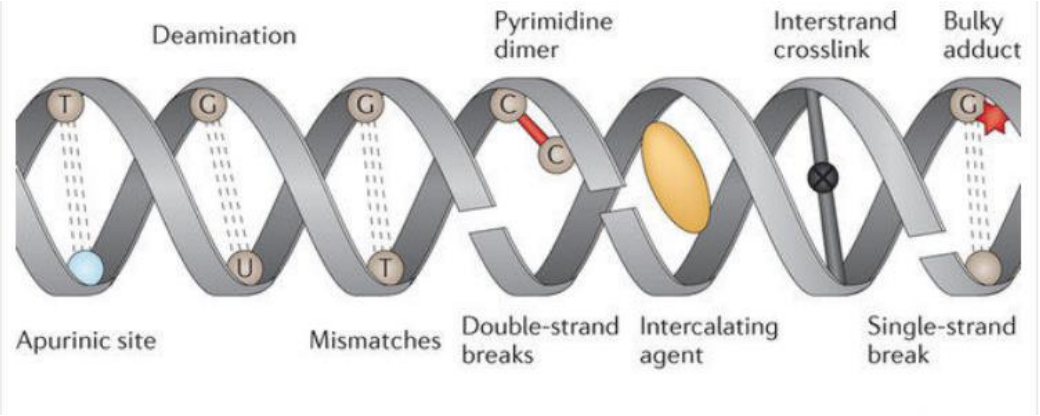


Figure.16: Types of genotoxic agents causing varying DNA damage (Helleday, Eshtad and Nik-Zainal, 2014).

Alkylating agents, antimetabolites, topoisomerase inhibitors and platinum drugs are some resources of chemical agents used in chemotherapy or anti-cancer drugs resulting in C•G to T•A transitions or G•C to T•A transversion depending on the type of agents used (*Figure 16*). Thereby indicates the effect of exogenous agents to generate mutations initiating tumor progression (Alexandrov *et al.*, 2013). Endogenous agents have been known to contribute to various DNA modifications ranging from adduct formation to SSBs and DSBs or DNA-protein crosslinks (Jackson and Bartek, 2009). Mitotic catastrophe and chromosomal aberrations has been suggested as the principal path of cell death (Dewey, Ling and Meyn, 1995; Swift and Golsteyn, 2014).

Recent breakthroughs in uncovering DNA damage has shown vast and diverse types of ***endogenous*** damage. Analogous to chemical agents, there is baseline damage which may overthrow the DNA repair machinery, thereby generating mutations and affecting genomic organization. It has been postulated that futile DNA repair genes leads to genomic instability and high mutational load associated with the DNA repair machinery has been found as an underlying cause of cancer (Loeb, Springgate and Battula, 1974; Nowell, 1976). Fractions of hereditary breast and ovarian cancer were accounted due to mutations in genes *BRCA1* and *BRCA2*, known to control DSB repair by homologous recombination (HR) (*Figure 17 E*). Moreover, whole genome sequencing studies has revealed widespread errors in most cancers involving HR and MMR (Morganella *et al.*, 2016; Nik-Zainal *et al.*, 2016).

Distinctive mutations and rearrangements yields definite pathways associated with cancer. During the course of malignancy, there may be ***specific mutations*** shaping up towards cancer, for instance, Alexandrov and his colleagues found specific thymine to guanine (T>G) base mutations at the immunoglobulin (Ig) genes. The Ig genes are the dominant region involved in chronic lymphocytic leukemia where the activity of error-prone DNA polymerase η in somatic hypermutation (SHM) is reflected (*Figure 17 C*). Also, underlying mutation(s) in dysfunctional DNA repair machinery has yet to be fully understood, where additional process de-regularizing the repair pathway have to be uncovered. One such example is chromothripsis, a one-off event cause chromosomal catastrophe and leading to multiple chromosomal rearrangements (Zhang *et al.*, 2015). These cataclysmic events may be facilitated by replication stress or due to spontaneous deamination.

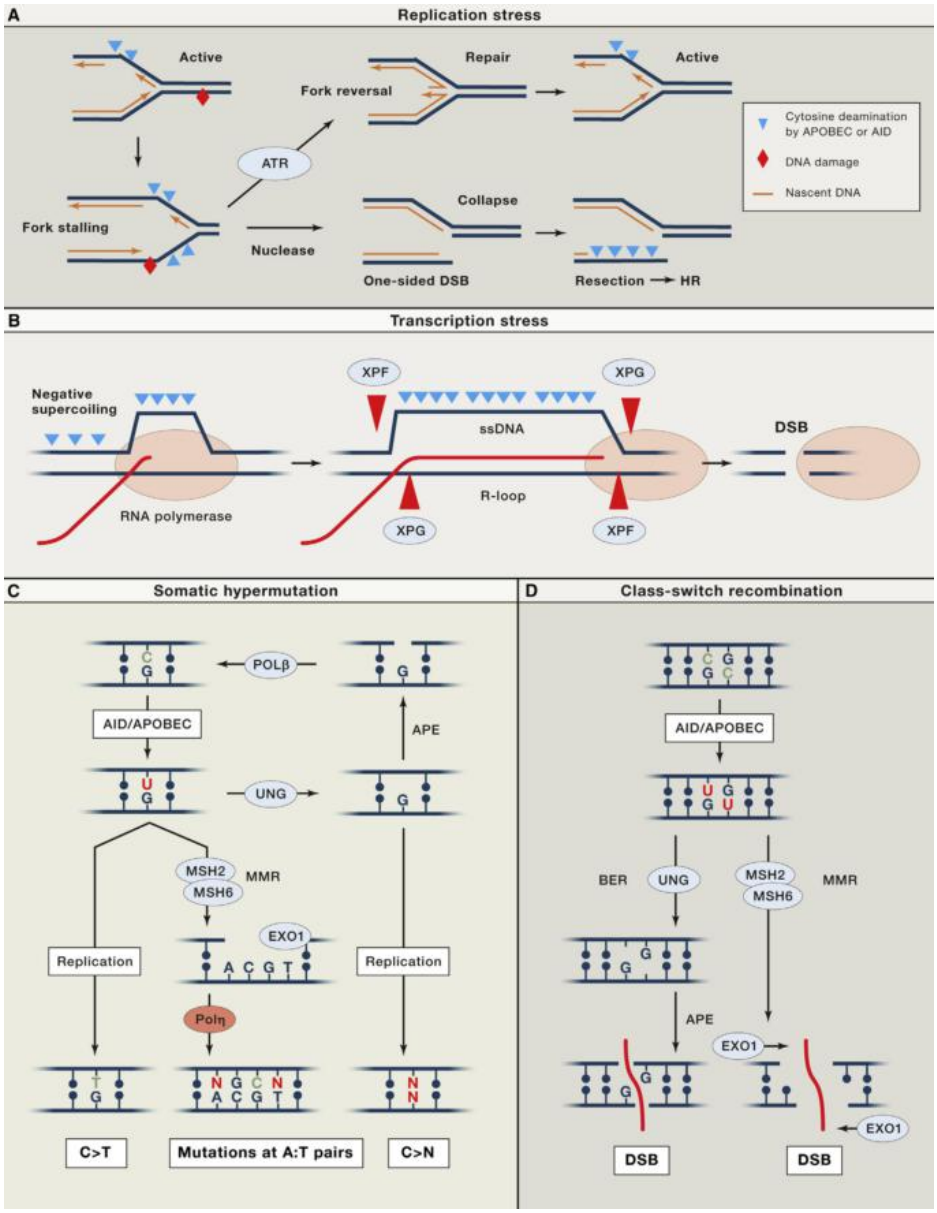


Figure 17: Overview of consequential DNA damage of cytosine deamination. A) ssDNA as a substrate for cytosine deamination during replication for progression. B) Transcription stress induced ssDNA exposure. C) Overview of SHM where AID/APOBECs act. D) Underlying mechanisms of cytosine deaminases in CSR (Tubbs and Nussenzweig, 2017).

During cell division, 6 x 10^9 nucleotides are replicated by DNA polymerases. These polymerases are known to work when mutated bases are excised during NER, BER and MMR and they fill the gaps in the ssDNA. Replication is known to generate mutations constantly, but are a very low level, depending on the tissue types (Tomasetti and Vogelstein, 2015). One of the most common base substitution signature in cancer was

also discovered to be accumulated in stem cells with elevated division rate. This was correlate with patient age, and was predominantly C>T transitions at CpG dinucleotide motifs, associated with 25% of somatic mutations in TP 53 codons driving tumorigenesis (Lindahl, 1993; Olivier, Hollstein and Hainaut, 2010; Alexandrov *et al.*, 2013). These deaminated products lead to transition mutations when the U:G mismatch is not recognized and replication is carried out (*Figure 17 A*). Active DNA editing is a contributor as an endogenous agent. It is a well-documented fact that ssDNA is a substrate for activation-induced cytidine deaminase (AID) and APOBEC family of deaminases (Swanton *et al.*, 2015). Many studies have suggested that APOBEC mutational burden as an indicator in cancer (Alexandrov *et al.*, 2013; Burns *et al.*, 2013); APOBEC signature was observed at elevated rates on the lagging strand, as extended exposures of ssDNA in DNA synthesis tolerates compared to the leading strand (*Figure 17A)*. These localized APOBEC signatures have been correlated to genomic rearrangements possibly due to increasing amounts of ssDNA produced due to replication and transcription stress (*figure 17A and B)*(Kanu *et al.*, 2016). During SHM and CSR, *Ig* genes are subjected to be targeted deamination by AID, with a preference for cytosine residues flanked by a 5` purine, resulting in U:G mismatches thereby generating C>T transition mutations (*Figure 17C*) (Di Noia and Neuberger, 2007). Instead, these mismatches are processed into DSBs leading to CSR (*Figure 17D*). Numerous other mechanisms have been hypothesized to act as endogenous agents for differential repair. For example distinct chromatin structures may be more prone to recruit error-prone repair than high fidelity repair (Liu and Schatz, 2009).

Endogenous and exogenous damage shapes the mutational landscape of cancer resulting in base substitutions, indels and chromosomal rearrangements.

1.8 The AID/APOBEC family of cytosine deaminases

Mutations, the driver of evolution, lead to variations within organisms and within species. They usually occur artlessly without an immediate need or cause, but this is not the case when AID/APOBECs are involved. The APOBEC family comprises of APOBEC1, activation-induced deaminase (AID), APOBEC2 (A2), APOBEC3A–H (A3A–H), and APOBEC4 (A4) proteins. This family has been originated from AID gene. Gene duplication in bony fish from AID and A2, were postulated as the probable events

leading to the evolutionary process of APOBEC proteins(Rogozin *et al.*, 2007; Conticello, 2008). However, purposeful mutations due to DNA deamination of cytosine to uracil are known to function in diverse cellular processes, ranging from antibody gene diversification to mRNA editing (Conticello, 2008). As part of their mechanism of targeted deamination, AID/APOBEC enzymes engage cytosines in the context of its neighboring nucleotides within ssDNA, resulting in APOBEC-induced mutations. In the context of cancer APOBEC induced mutations are often seen in a strand-ordinated fashion and present a specific mutagenic signature, TCW→TTW or TCW→TGW (mutated nucleotide underlined, W = A or T) (Stephens *et al.*, 2011; Roberts *et al.*, 2012; Alexandrov *et al.*, 2013).

AID

AID was the first member of the AID/APOBEC protein family to be characterised as a DNA mutator and its activity is exerted in activated B cells. AID is encoded by a gene localised on the chromosome 12 and it is primarily expressed in germinal center. AID triggers distinct downstream process: CSR, SHM, and, in some vertebrates, gene conversion where it edits the immunoglobulin (Ig) loci (Muramatsu *et al.*, 2000; Arakawa, HauschiLd and Buerstedde, 2002; Harris *et al*., 2002). AID induces different outcomes CSR and SHM: in CSR, it induces DSBs, leading to rearrangement of the Ig locus occurs (thereby, increasing the efficiency of immune response); in SHM, C-to-U DNA editing creates U:G mispairs resulting in somatic hypermutation process. Further, the U:G sites are subjected to BER or MMR, and errors if any, leads to transitions or transversions (Di Noia and Neuberger, 2002, 2007; Petersen-Mahrt, Harris and Neuberger, 2002). Moreover, at the Ig heavy and light chain loci, it was reported that AID may induce recurrent DSBs in non-Ig loci especially within B cell super enhancers (*Figure 17D)* (Meng *et al.*, 2014; Qian *et al.*, 2014). As some of them might be off-target sites of AID, its promiscuous activity might damage oncogenes, leading to chromosomal translocations and tumorigenesis (Ramiro, Nussenzweig and Nussenzweig, 2006). The most striking example of the association of AID with cancer is the c-myc/IgH translocation in Burkitt lymphoma, where the c-myc gene was translocated directly on the switch region of the *Igh* gene (Ramiro *et al.*, 2004, 2006).

APOBEC 1

APOBEC 1 was the first member of the APOBECs to be identified (Navaratnam *et al.*, 1993). APOBEC1 is known to be expressed in the human small intestine where it edits the *ApoB* pre-mRNA, thus forming a truncated form of *ApoB* called apoB-48, principal component of chylomicrons (Baum *et al., 1990*). APOBEC1 forms part of a complex which recognises a sequence localised at 3' of the cytosine to be deaminated called 'mooring sequence'. The complex binds to an AU-rich region and APOBEC1 deamination of cytosine 6666 changes a glutamine codon into a stop codon (Chester *et al.*, 2000). A1 does not edit the *ApoB* mRNA in amniotes (Harris, Petersen-Mahrt and Neuberger, 2002) but interestingly, A1 is also known to act as a DNA mutator and also restricts retroelements, which were reported in ex vivo assays, hinting towards an ancestral function of A1. Phylogenetic studies suggest that APOBEC1 appeared un tetrapod, while ApoB RNA editing evolved in mammals (Severi, Chicca and Conticello, 2011a).

Similar to AID, its mutagenic ability can be dangerous and its overexpression may lead to cancer(Severi, Chicca and Conticello, 2011; Saraconi *et al.,2014*)

APOBEC 2

APOBEC2 is expressed in skeletal muscle and heart and does not exhibit catalytic activities compared to its paralogues in bacterial assays (Lada *et al.*, 2011). Most of the molecular function id elusive, and it has been related to muscle development and repair (Sato *et al.*, 2010).

APOBEC 3

The APOBEC3s (A3s), consists of seven human paralogs in the human genome which were initially identified as paralogues of APOBEC1 (Jarmuz *et al.*, 2002). There are eight genes that encode for the APOBEC3s, seven of them form a cluster inside the chromosome 22, and an additional gene on human chromosome 12q24.11, which is likely to be a pseudogene possibly originated from duplication of the APOBEC3G gene

(Jarmuz *et al.*, 2002; Conticello *et al.*, 2005). The APOBEC3s can be classified in two groups with respect to the number of zinc-finger domains: APOBEC3A, APOBEC3C and APOBEC3H have a single zinc-finger domain, while APOBEC3B, APOBEC3DE, APOBEC3G and APOBEC3F have two zinc-finger domains (Conticello, 2008; Vieira and Soares, 2013). The APOBEC3s play an important role in innate immunity where it inhibits retroviruses via deamination of the cytosine in retroviral DNA intermediates (Harris and Liddament, 2004). APOBEC3G was identified as a factor involved in HIV restriction, where APOBEC3G is packaged into retroviral virions, acting on the first strand of DNA produced by inverse transcription in the host cells(Sheehy *et al.*, 2002; Harris *et al.*, 2003; Mangeat *et al.*, 2003). All the primate paralogues of APOBEC3 are capable of retroviral restriction against other viruses, with rather different efficiencies. There are evidences suggesting that APOBEC3 proteins can also interfere with the movement of retrotransposon elements (Chiu and Greene, 2008) and are likely to act for the defence system against foreign DNA (Stenglein *et al.*, 2010). APOBEC3 mediated mutagenesis was also reported due to replication stress which exposed ssDNA, as substrates for endonuclease cleavage (Kanu *et al.*, 2016). Most of the APOBEC mutation pattern has been linked with the activity of two members of the APOBEC3 subfamily—APOBEC3B (A3B) and APOBEC3A (A3A) (Alexandrov *et al.*, 2013; Burns *et al.*, 2013), implying that ubiquitous APOBEC-mediated mutagenesis is carcinogenic. Several reports suggested upregulation of A3B in almost 50% of cancers and is implied with chemotherapy resistance in estrogen-receptor-positive breast cancer (Burns *et al.*, 2013; Law *et al.*, 2016). This is particularly interesting as some of the AID/APOBEC genes are induced by estrogens (Pauklin *et al.*, 2009). The APOBEC3 mutational load may occur either as strand –coordinated clusters resulting in mutational showers, termed “Kateagis” which was earlier reported by Nik-Zainal in 2012. These localized A3 mutations exhibit colocalization instigating genomic rearrangements (Burns, Temiz and Harris, 2013).

1.9 DNA repair

"*We totally missed the possible role of ... [DNA] repair although ... I later came to realise that DNA is so precious that probably many distinct repair mechanisms would exist.*" Francis Crick, writing in Nature, 26 April 1974 (Crick, 1974). This is how Francis Crick had misconceived that DNA, being highly robust and stable did not need to be repaired, hence significantly delaying the idea of mutation and repair. Thousands of subsequent studies displayed the dynamic state of the DNA. Two determine the integrity of the genome: the balance between DNA damage and ***DNA repair.***

As described earlier, there are many types of DNA lesions and in order to counter them, there are several damage repair pathways (*figure 18)*. DNA repair pathways act on the various forms if DNA damage through multiple and distinct mechanisms specific for lesions: base excision repair (BER), mismatch repair (MMR), nucleotide excision repair (NER), and double-strand break repair, which includes both homologous recombination (HR) and non-homologous end joining (NHEJ) (Roger *et al.*, 2006).

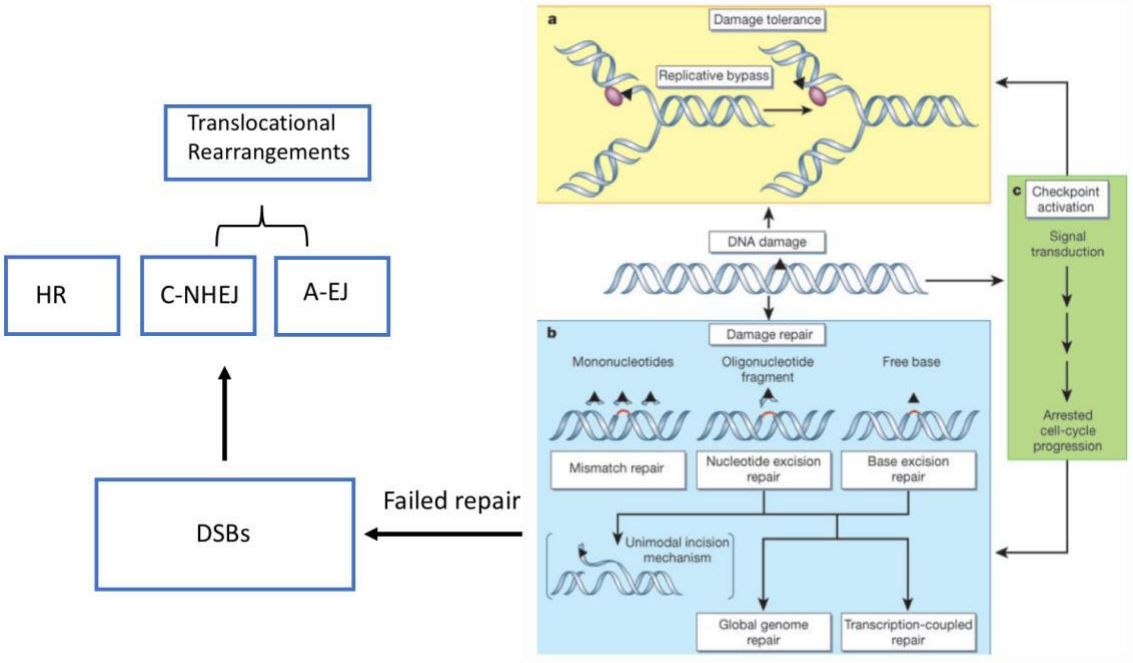


Figure .18: Different types of DNA lesions (black triangle) and their pathways. a. During damage tolerance, damaged sites are recognized by the replication machinery before they can be

repaired. b. DNA repair encompasses the excision of bases and DNA synthesis (red wavy lines), which requires double-stranded DNA. Error during DNA replication yields mis-paired bases which are excised via MMR. A damaged base is excised as a single free base (BER) or as an oligonucleotide fragment (NER). c. The cell has a network of complex signaling pathways that arrest the cell cycle and may ultimately lead to programmed cell death. Failed repair generates DSBs which has to be repaired, that's where C-NHEJ and A-EJ come into play (modified from Friedberg, 2003).

Base excision repair (BER), is the predominant mechanism responsible for the repair of damaged DNA bases. Abasic sites, uracil bases, 8-oxoguanine and single strand breaks are efficiently repair by BER, which remove or replace a piece of DNA including the damaged site. The enzymes that remove the bases are the glycosylases and we know about at least twelve DNA glycosylases acting upon a single or small number of partially overlapping base lesions (Bedard and Massey, 2006).The glycosylase cuts the bond between damaged base and deoxyribose (Zharkov, 2008).The resultant AP site is both an intermediate product of BER and a highly prevalent DNA lesion produced by spontaneous base loss. Regardless of mechanism, incision of the phosphodiester bond results in a BER intermediate strand break harboring 3- and 5-blocking lesions. This repair process comes to an end when these blocked termini are restored to conventional 3-OH and 5-phosphate ends, which are essential for DNA synthesis and subsequent DNA ligation reactions (Krokan and Bjørås, 2013).

Nucleotide Excision Repair (NER), primarily repairs bulky DNA adducts and pyrimidine dimers. Single strand nicks on either side of the lesion are introduced by multi-protein complex to the altered DNA strand. They function in diverse manner in eukaryotes and prokaryotes; where nicks are spatially introduced; 12 nucleotides apart in bacteria, while it is more than 24 in eukaryotes. When the damaged structure is recognized by an endonuclease that cuts the DNA strand on both sides of the damage. Later, an exonuclease removes a damaged piece of helix. Followed by synthesis, where, the resulting single-stranded region works as a template for a DNA polymerase that synthesizes a stretch that replaces the excised sequence. Mutations in the NER pathways,

especially the ones involved in recognition of the damaged site, in humans causes *xeroderma pigmentosum*, susceptible to skin cancer (Shuck, Short and Turchi, 2008).
NER and BER transpire simultaneously arising from different mechanisms, depending on the damaged site in the genome where active genes are constantly expressed (transcription-coupled repair) or are transcriptionally silent (global genome repair) (Friedberg, 2003; Le Page *et al.*, 2005).
The DNA repair machinery explained above is focused more toward SSBs; any errors in the repair pathways results in DSBs, which transpires into chromosomal translocations. When DSBs occur, the cells try to rectify them by using its DSB repair machinery which can either be error-prone or in most cases error free (Ghadimi and Ried, 2015).

Mismatch Repair (MMR), is involved in post-replication repair when mis-incorporated bases have escaped the proofreading activity of replicative polymerases. It corrects insertion and deletion loops that result from polymerase slippage during replication of repetitive DNA sequences. Once the alteration is detected, endonuclease nicks the synthesized DNA strand, while the endonuclease *exoI* deletes several nucleotides including the mismatches. DNA polymerase δ starts filling the gaps and the DNA ligase comes into action by repairing them (Li, 2008). MMR is regulated by about 10 proteins and mutations in these genes coding for the proteins have been found in hereditary cancer syndromes. Moreover, cells with dysfunctional MMR, display a mutator phenotype, which is characterized by microsatellite instability and an elevated mutation frequency. It is important to add that germline mutations in MMR genes predispose to a variety of cancers such as hereditary non-polyposis colon cancer (Lynch syndrome) (Peltomaki, 2001; Müller and Fishel, 2002).

Homologous Recombination (HR), is an error free DSB repair pathway. HR mainly functions in the S and the G2 phase of the cell cycle as it requires an intact template for efficient repair. The MRE11/RAD50/NBS protein complex is an intrinsic part of HR (Daley and Wilson, 2005; Williams, Lees-Miller and Tainer, 2010; Lammens *et al.*, 2011). This complex detects the DSBs once the exonuclease excises the nucleotides from the free dsDNA end, the protein complex comes into action where it works together with.

This yields to single-stranded 3′ DNA overhang which are further processed by replication proteins A (RPA), RAD51, BRCA2 and several other proteins into the RAD51-ssDNA-nucleoprotein filament. This nucleoprotein affects the dsDNA of the homologues site on the sister chromatid, serving as a template for error-free synthesis of DNA (Popp and Bohlander, 2010). DSB repair by HR is sterically complicated as it works several proteins also such as REQL2, BRCA, RAD54, etc. Alteration in most of these proteins are associated with various oncological syndromes (Daley and Wilson, 2005; Popp and Bohlander, 2010). GIN-associated tumors typically bear alterations in this pathway (Jones *et al.*, 2009; Bunting and Nussenzweig, 2013).

Classical Non-Homologous End Joining (C-NHEJ), is a form of error prone repair, implying that after resection of the broken ends, there are small insertions or deletions where the DSB was located (Roth and Wilson, 1986; Rassool, 2003). Its mode of repair doesn't involve an intact template, and therefore is the pathway of choice to repair DSBs occurring during the G1 and S phase of the cell cycle. C-NHEJ is more prone to merge DSBs that are not positioned with each other as it does not require a homologous sequence to guide its repair process. During V(D)J somatic recombination and CSR at the immunoglobulin loci, C-NHEJ is involved (Ghadimi and Ried, 2015). In the C-NHEJ, KU70/80 proteins are attached to the broken end in order to prevent the ends for drifting apart (Soutoglou *et al.*, 2007). This is followed by recruitment at the break of DNA-PKCs and the MRN complex (MRE11-RAD50-NBS), are involved in resection of broken DNA strand which are finally joined by the DNA ligase IV-XRCC4 complex(Roth and Wilson, 1986).

Alternative End Joining Pathway (A-EJ) or Microhomology-Mediated End-Joining (MMEJ), usually occurs in absence of KU70/80, XRCC4 or Ligase 4 and principally relies on the MRN complex (Wang *et al.*, 2003; Daley and Wilson, 2005; Popp and Bohlander, 2010). A-EJ mediated DSB repair leaves it trademark process with stretches of micro-homologies of 6 to 8 base pairs. This mechanism has been responsible for DSB repairs resulting in chromosomal translocations (Boboila *et al.*, 2010). Additionally, experimental models for chromosomal translocations displayed micro-homology based

mechanisms for a few de novo translocations (Wang *et al.*, 2003; Daley and Wilson, 2005). Similar observations were seen in human germline for chromosomal translocations. The breakpoints revealed complex structures with fragments of local DNA sequences, tiny inversions and deletions hinting how both C-NHEJ and the A-EJ works towards a substantial fractions adding human germline chromosomal translocations (Chiang *et al.*, 2012). Although DNA repair pathways have been extensively studied, sources of DNA damage which structure the mutational landscape of the cancer genome remain unclear.

1.10 DNA repair and cell cycle progression

As the DNA is continuously damaged, an efficient repair system must be chosen depending upon the type of lesion and the cell-cycle phase it is in. In case of DSBs, which usually are in S and the G2 phase are readily repaired by HR; while in the G2-M, where the chromosomes are condensed the repair becomes more difficult due to inability to find homologous regions. There are the cyclin-dependent kinases (CDKs) which controls the coordination of DNA repair (Vermeulen, Van Bockstaele and Berneman, 2003). Cells integrate DNA repair process with transcription and apoptosis forming the DNA-damage response (DDR) regulated by checkpoint proteins.

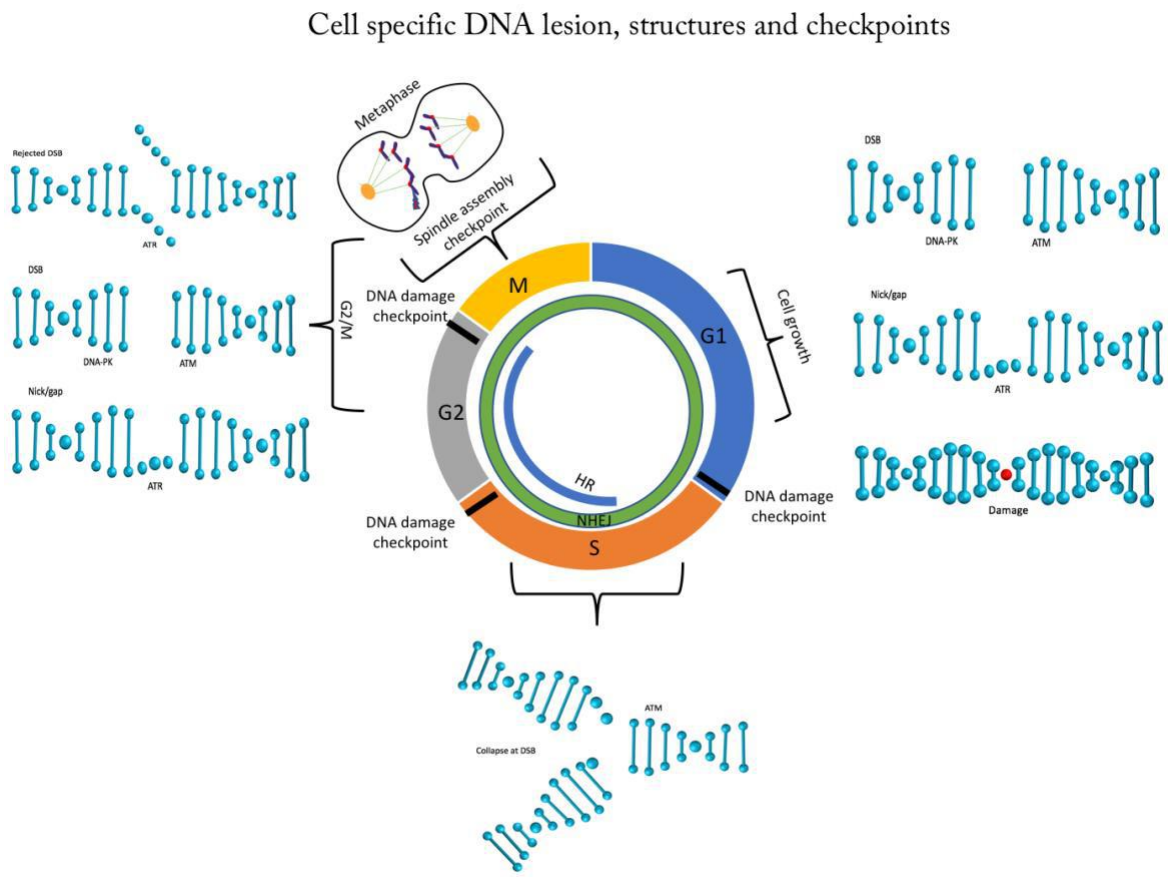


Figure.19: DNA repair and cell cycle: In G1 phase, double-strand breaks (DSBs) lead to activation of the phosphoinositide 3-kinase related kinases DNA-PK and ATM, while damages UV lead to ATR activation. Unrepaired DSBs and nicks in G1 causes collapse of replication forks, activating checkpoint. In the S phase, activation of ATR is due to stalled forks or gaps that are generated during replication.

Checkpoints intervene to co-ordinate DNA repair with chromosome metabolism and transition through the cell cycle (Giannattasio *et al.*, 2004). As described earlier with mechanisms of damage and repair, it is implied that stability and activity of checkpoint involved proteins regulates repair. Core elements of the checkpoint machinery primarily involve phosphoinositide 3-kinase related kinases ATM (DSBs), ATR (ssDNA and stalled replication forks and DNA-PK (DSBs) (Bartek and Lukas, 2007). These kinases work during different phases in cell-cycle progression. These kinases are activated once the DNA lesion occur and interact with specificity factors for e.g. NBS1 (for ATM), ATRIP (for ATR) and Ku80 (for DNA-PK) (Falck, Coates and Jackson, 2005; You *et al.*,

2005). ATM and ATR are recruited towards the damaged sites and forms substrates for checkpoint kinases, namely CHK2 and CHK 1 respectively (Matsuoka *et al.*, 2007).

In ***G1 phase,*** cells are exposed to damages from ROS species, UV/IR or by chemical agents. Depending upon the type of damage, SSB or DSB, different DNA repair pathways comes into play; starting from the BER and the NER pathways (Russo *et al.*, 2004; Sancar *et al.*, 2004). While DSBs induced by IR activates the NHEJ, due to compaction of chromatin and absence of sister chromatids in G1 (Takata *et al.*, 1998).

Nucleotide mis-paring, nicks and fork collapse are generally observed during ***S phase,*** when replication-fork transitions puts the stability of the chromosomes at risk if left unrepaired, leading to topological modifications leading to supercoils (Branzei and Foiani, 2005). This followed by topoisomerase mediated resolution allowing completion of S phase, further entails condensation and segregation during the M phase. Errors are corrected pre-dominantly by MMR, which identifies the error site and corrects it. BER is also known to be involved here where its main role is removing mis-incorporated uracil's (Sancar *et al.*, 2004). Gaps and nicks during replication are the primary source for HR, these are dealt by damage tolerance or by-pass replication machinery (figure 19). These replication machineries are regulated further by translesion synthesis (TLS) polymerase (error-prone) and template switch (TS)(error-free) (Fabre *et al.*, 2002; Jiricny, 2006; Lettier *et al.*, 2006; Branzei and Foiani, 2007; Lehmann *et al.*, 2007).

Strand breaks if not dealt with during replication, must rectified prior to mitosis. In case of HR, during ***S*** and ***G2 phases***, it uses sister chromatid as a template for repair, hence the need for the chromatids to be in close proximity, which is controlled by cohesin. Cohesin proteins works with sister chromatid cohesins and structural maintenance of chromosome (SMC) proteins together to form a complex and must be established during the S phase (Uhlmann and Nasmyth, 1998; Hirano, 2006). There are instances where DSBs triggers cohesion post replication, which are crucial for sister chromatids repair in G2 phase (Sjögren and Nasmyth, 2001; Ström *et al.*, 2004). The replicated DNA must be fixed during the S-G2 to prevent chromosomal breakage during segregation. NHEJ repair

is often observed during segregation errors and its likely to occur in the subsequent G1 phase especially when the checkpoint systems have not placed the cell under cell-cycle arrest during ***G2 and M phases*** (Deming *et al.*, 2001, 2002; Franchitto, Oshima and Pichierri, 2003).

Thus, cell cycle progression and DNA repair needs to be constantly monitored by the surveillance mechanisms mentioned above as precise DNA repair is the pivotal form for genome integrity.

1.11 Models for chromosomal instability

Several models *in vitro* and *in vivo* were developed to study CIN where most of them involves usage of a carcinogen. This is primarily due to the weak phenotypes obtainable in animal models. Accurate assessment for CIN, ideally must involve multiple repeated measurements of individual cells across cell populations in order to comprehend the variability present in the tumor (Geigl *et al.*, 2008). These are pretty straight-forward to analyze in cancer cell line, but determining the correct rate of CIN in various types get tricky with clinical specimens due to both cellular heterogeneity and to the availability of the specimens as well. Brief measures must be taken before interpreting the results due to the variability in calculating the frequency and extent of changes. As genomic instability is occurring at various genetic levels, detecting CIN serves as an adequate measure for GIN, but not vice-versa. These methods revolve around single cell or multiple cell approaches and includes karyotyping, flow cytometry, single nucleotide polymorphism (SNP) arrays, genome sequencing, and polymerase chain reaction (PCR) which are mentioned (Table 1).

Technique	Cellularity	Type
Karyotyping	Single-cell	CIN and Aneuploidy
Single cell sequencing	Single cell	CIN, translocations, insertions, deletions and mutations.
Flow cytometry	Multiple cell	CIN
SNP arrays	Multiple cell	CIN, SNP and loss of heterozygosity

Whole-genome sequencing	Multiple cell	CIN, translocations, insertions, deletions and mutations.

Table.1: Methods to analyze CIN (Pikor et al., 2013).

Karyotyping assesses chromosome number and structure in ***single cells*** to identify abnormalities and structural rearrangements (Beheshti *et al.*, 2001). The metaphase chromosomes are stained with Giemsa stain which are taken up by gene poor A and T rich regions resulting in a band pattern which are used to distinguish different chromosomes and ease the identification of abnormalities. Similar approaches include fluorescence *in situ* hybridization (FISH) coupled with spectral karyotyping which paints each chromosome with a different color allowing rapid identification (Bayani and Squire, 2001). Despite its limitations (labor intensive and impossibility to comprehend the whole cell population), it remains one of the most used technique to analyze non-clonal aberrations. With next generation sequencing and whole genome amplification technologies, tracing single cells has been much easier, and these novel approaches offer great insights at genomic alteration from cell to cell allowing simultaneous accurate detection of mutation and copy number variations (Zong *et al.*, 2012).

Flow cytometry is vastly used in ***multicellular approaches*** of CIN, where the cell suspension passes through scattered light, emitting fluorescence which is in turn assessed to measure aneuploidy. Comparative genomic hybridization (CGH) is yet another essential tool which permits quantitatively detects and visualizes whole and segmental chromosomal alterations (Pinkel *et al.*, 1998; Ishkanian *et al.*, 2004). Succinctly, reference genomic DNA and test DNA are labeled, pooled, and hybridized onto arrays comprising BAC, cDNA, or oligonucleotides; these are further analyzed based on fluorescence intensity. SNP array per se, warrants efficient mapping of copy number alterations and helps discriminate alleles at polymorphic sites (Zhao *et al.*, 2004; Gondek *et al.*, 2007; Heinrichs and Look, 2007). In general, sequencing-based technologies can detect a full spectrum of genomic aberrations, including single nucleotide variant (SNV), small insertion/deletion (indel), CNV, translocation, and novel mutations in CIN-tumors.

Cellular models have been developed to measure CIN and assess the agents and the processes associated with it. The use of the Human Artificial Chromosome (HAC) allows quantification of CIN in live cells where it measures mitotic chromosomal segregation in human cells (Lee *et al.*, 2013; Markossian *et al.*, 2016). There have been several quantitative techniques reported for CIN; the cytokinesis-block micronucleus assay (CBMN) looks for chromosome loss with the help of centromere-specific DNA probes in binucleated cells (Camps *et al.*, 2005).

With recent advances in DNA sequencing, chromothripsis was identified, displaying how powerful this state of the art genome analytical methods is.

2. Results and Discussion (Part I)

2.1 AID/APOBECs triggers DNA damage

Aiming to assess the effects of AID/APOBECs on chromosomal instability, I first evaluated whether the AID/APOBEC expression constructs were able to trigger a DNA damage response upon transient transfection in cells. To this aim I have used immunofluorescence to visualize the presence of γ-H2AX in HT 1080 HAC cells, a human cell line with a non-essential human artificial chromosome (HAC) with a functional kinetochore.

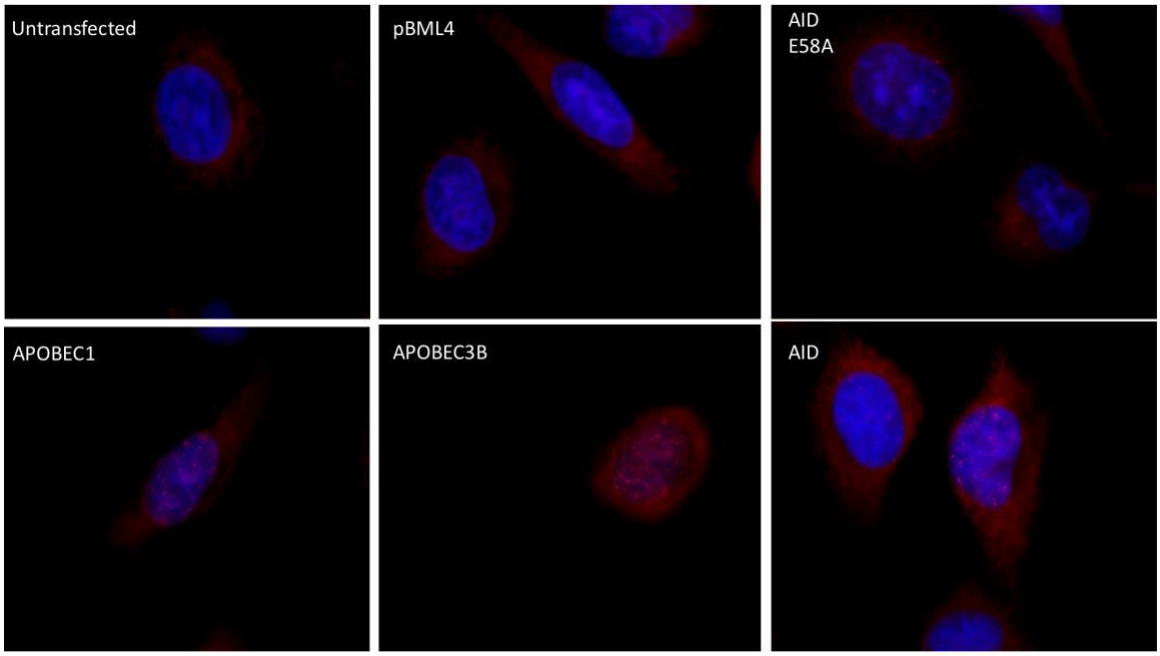


Figure.20: DNA damages induced by AID/APOBECs. Immunofluorescence of γ-H2AX where the HT-1080 cells were transfected with mock plasmid (pBML4), AID and its catalytically inactive mutant (AID E58A), APOBEC1 and APOBEC3B. Nuclei were counterstained with DAPI (original magnification 100×). Red foci represent DNA damage.

Essentially, H2AX is a known factor in the repair process of damaged DNA. The H2AX constitutes approximately 10% of the H2A histones in humans. The H2AX in presence of DNA damages is due to phosphorylated serine (139^{th}) residue (γ-H2AX). When SSBs and DSBs occurs, it results in immediate phosphorylation, aiding the recognition of DNA

damage almost within a few minutes from its occurrence (Kuo and Yang, 2008). Thus, by visualizing the accumulation of γ-H2AX will allow assessment of DNA damage caused by AID/APOBECs.

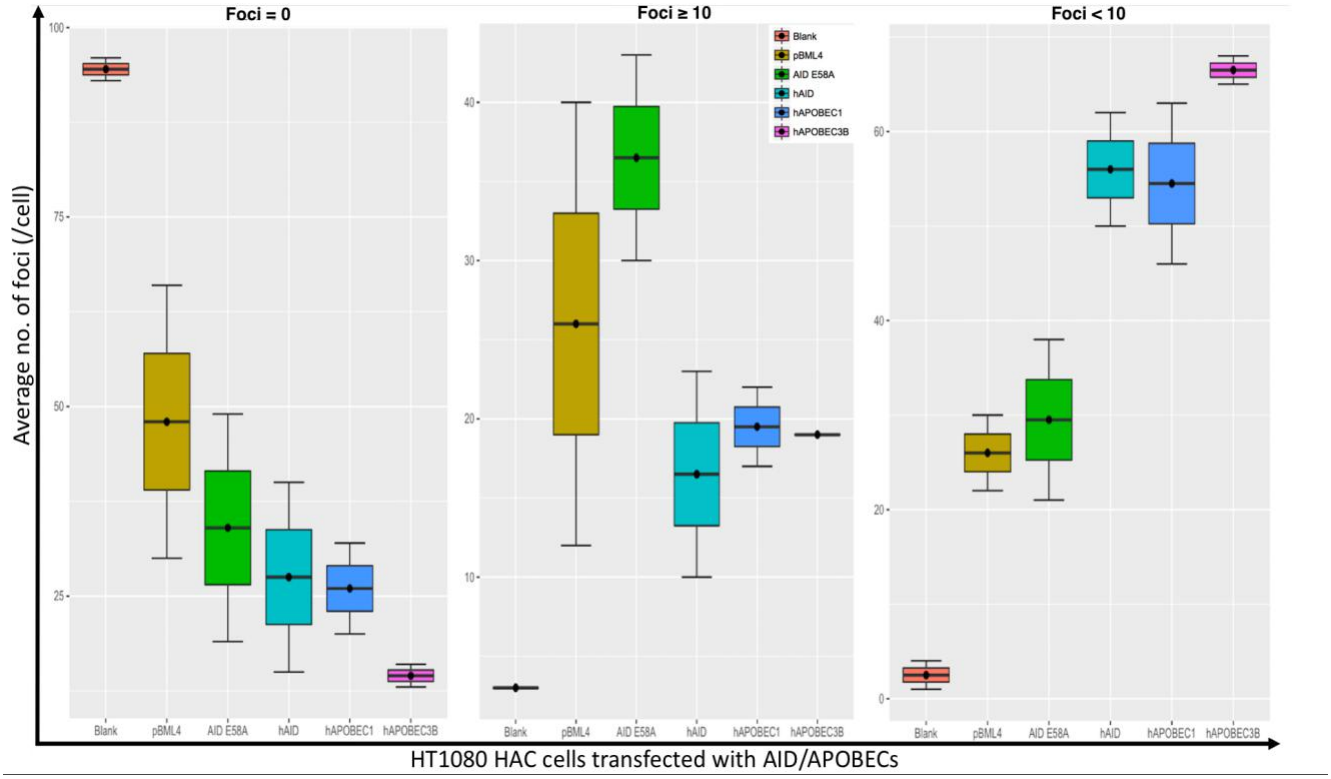


Figure.21: Production of the histone γ-H2AX foci: Graphical overview of γ-H2AX for 100 cells a foci range = 0, (foci range ≤10 and foci range >10. Box plots summarize foci of HT 1080 cells transiently transfected with pBML4 (empty plasmid), AID E58A (catalytically inactive deaminase mutant), APOBEC1, APOBEC3B along with untreated cells (blank). The combined data for two independent experiments in which a total of 100 nuclei were counted for each condition, p=0.005.

The HT 1080 HAC cells were transiently transfected with AID/APOBECs, followed by enrichments for transfected cells with puromycin. The HAC is continuously maintained as a non-essential 47th chromosome that replicates and segregates like a normal chromosome in human cells. To assess the levels the DNA damage *(figure 20)* quantification of γH2AX foci was performed manually, by individually counting the number of foci present in each cell nucleus. Due to the probabilistic nature of DSBs occurring, a systemic assessment was put in place for statistical robustness. I manually scored the foci and divided them primarily in three groups; *no foci, less than or equal to*

ten foci (aiding elimination of false negative) and greater than ten foci (*Figure 21 a and b*). In order to control for potential artefactual effects due to transfection, the HT 1080 HAC cells were transfected with a catalytically inactive mutant of AID (AID E58A) and an empty plasmid (pBML4). An increase in the nuclear foci for γH2AX signal was observed in cells transfected with AID, APOBEC1 and APOBEB3B. In case of AID and APOBEC1, there was around 55% of γH2AX foci intensity, while it was remarkably higher in the case of APOBEC3B (67%). Indeed it doesn't come as a surprise, especially in the case of APOBEC3B and AID, which previously have been demonstrated for its ability to promote γH2AX (Burns *et al.*, 2013; Daniel and Nussenzweig, 2013; Petersen *et al* ., 2001) ; while, these are new findings in case for APOBEC1, presenting its potential as a key player in DNA damage. Taken together, when the AID/APOBECs are overexpressed, there is rapid induction of strands breaks, indicating DNA-damage accumulation (*figure 21c*) and potentially causes genome instability. DNA strand breaks are one of the primal cause for lesions underlying the formation of chromosomal aberrations.

Considering that the AID/APOBECs have been associated to the onset of cancer through their ability to mutate DNA, I postulated that, mutagenic activity of these cytidine deaminases may have the capability of inducing chromosomal breaks resulting in chromosomal aberrations. With the aid of cytogenetic tools, it would help us understand downstream the DNA modifications of the cytidine deaminases.

2.2 Chromosomal aberrations fuelled by AID/APOBECs

Chromosomes are inherently uninemic i.e. each chromatid comprises of a single continuous DNA molecule. Thus, there is need to elucidate the molecular mechanisms of DNA damage response and in determining the link between the DSBs repair and its effects on chromosome. As described earlier there are four major types of chromosomal aberration which results in varying forms of structural aberrations

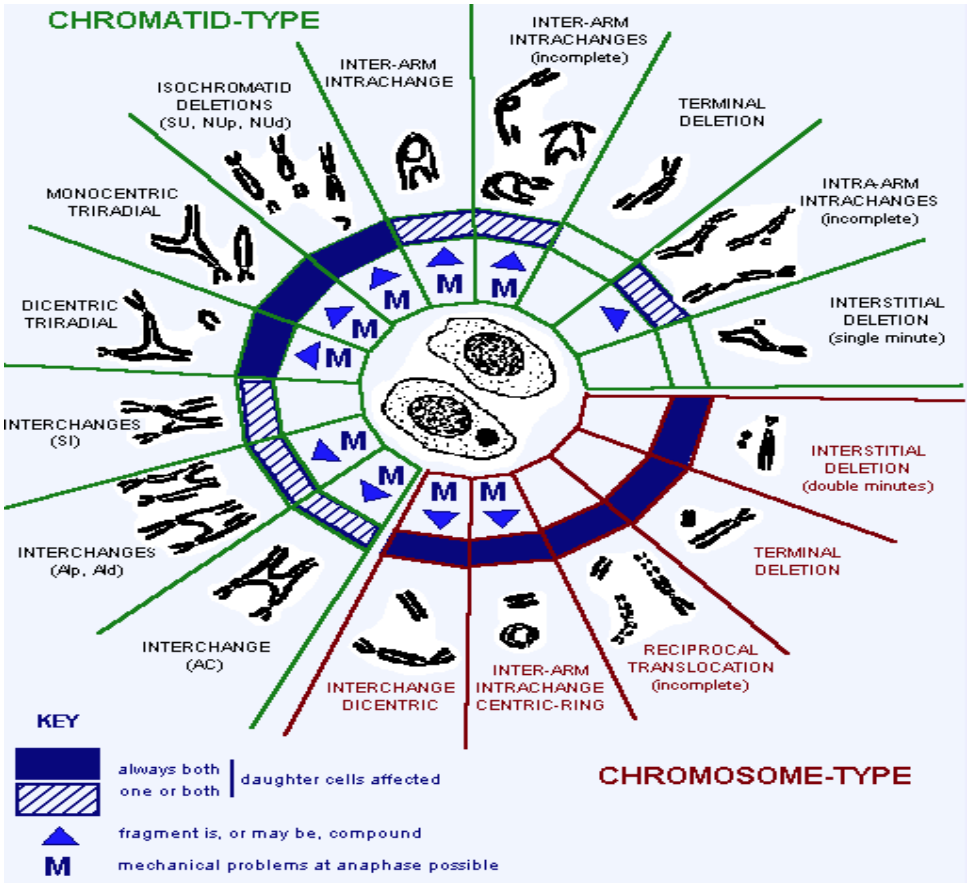


Figure.22: Different structural chromosomal aberrations resulting in acentric fragments (AF) leading to form micronuclei (MN) (John Savage, 2000).

Acentric fragments (AF) when duplicated results in MN. To determine whether AID/APOBECs induces aberrations resulting in acentric chromosome fragments during anaphase (*figure 22*), it seemed rather reasoned to understand its bearing thus I used cytogenetic tools to determine the same.

The HT 1080 HAC cells were transfected with AID, APOBEC1 and APOBEC3B, followed by colcemid treatment and lysing them in hypotonic buffer. These were later fixed, and the resulting chromosomes were stained with Giemsa. I examined 50 metaphase spreads per sample and evaluated chromosomal aberrations by scoring for chromosome aberration, chromosome exchange, and other abnormalities (dicentrics, rings, anaphase bridges, and pulverized chromosomes) (*Figure 23).*

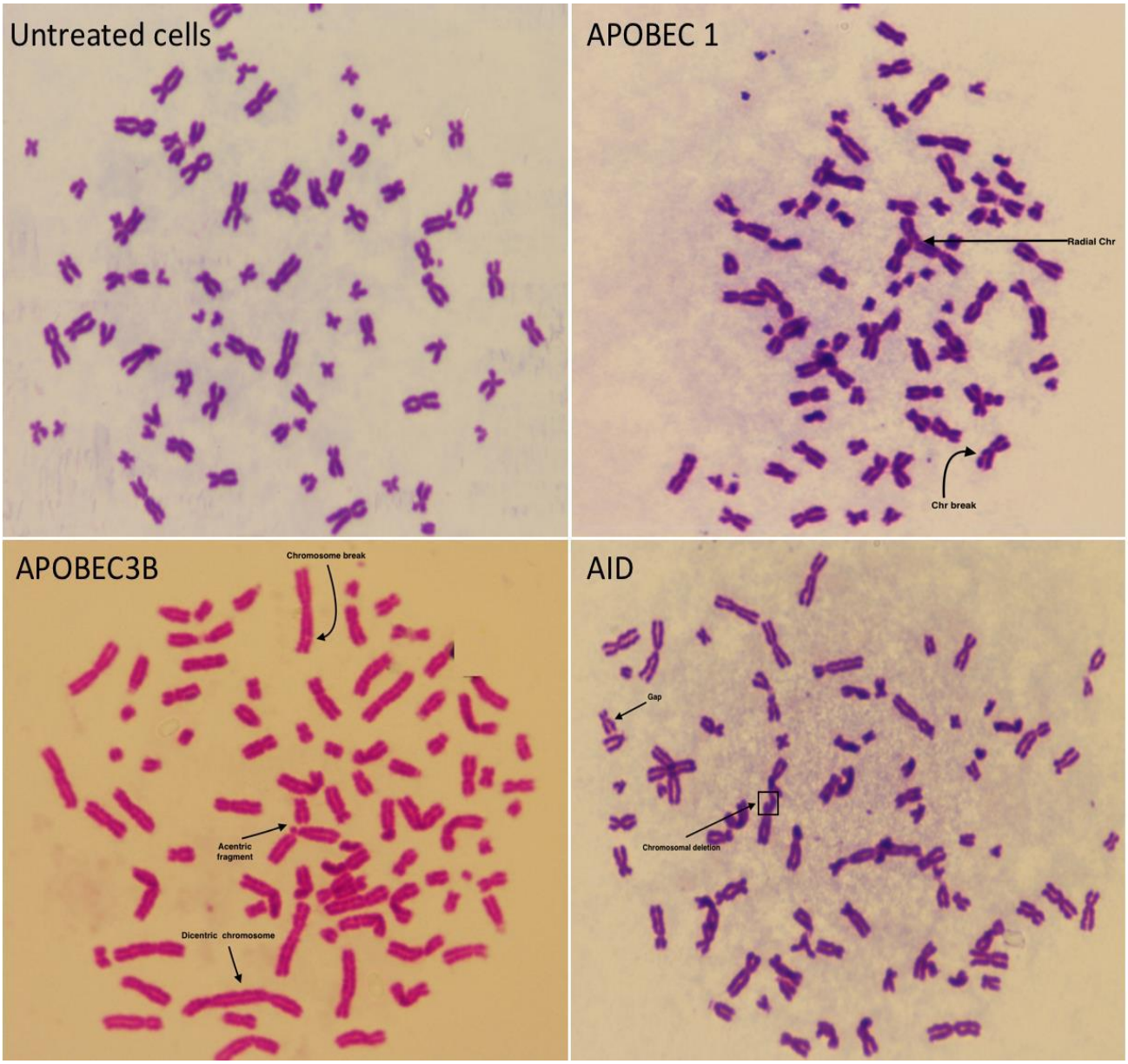


Figure.23: AID, APOBEC1 and APOBEC3B causing chromosomal breaks. Metaphase spreads were prepared from HT1080 HAC cells transfected with AID, APOBECs. Arrows indicate aberrations where radial chromosomes (APOBEC1), chromosomal break (APOBEC1 and APOBEC3B), di-centric chromosome(APOBEC3B) and gaps (AID), while box represent deletion (AID).

Overall, there was an increase in the number of structural aberration caused by the cytosine deaminases. Notably, in case of AID, the number of aberration were three times higher than in the control sample, while in the case of APOBEC1 and APOBEC3B,

a two-fold increase was observed (*figure 24a*). This clearly indicates the role of AID, APOBEC1 and APOBEC3B as DNA damage agents accounting for chromosomal instability. Additionally, there was a substantial increase in chromosomal exchange, an outcome from DSBs resulting from damage with aberrant repair within a chromosome or between chromosomes.

a)

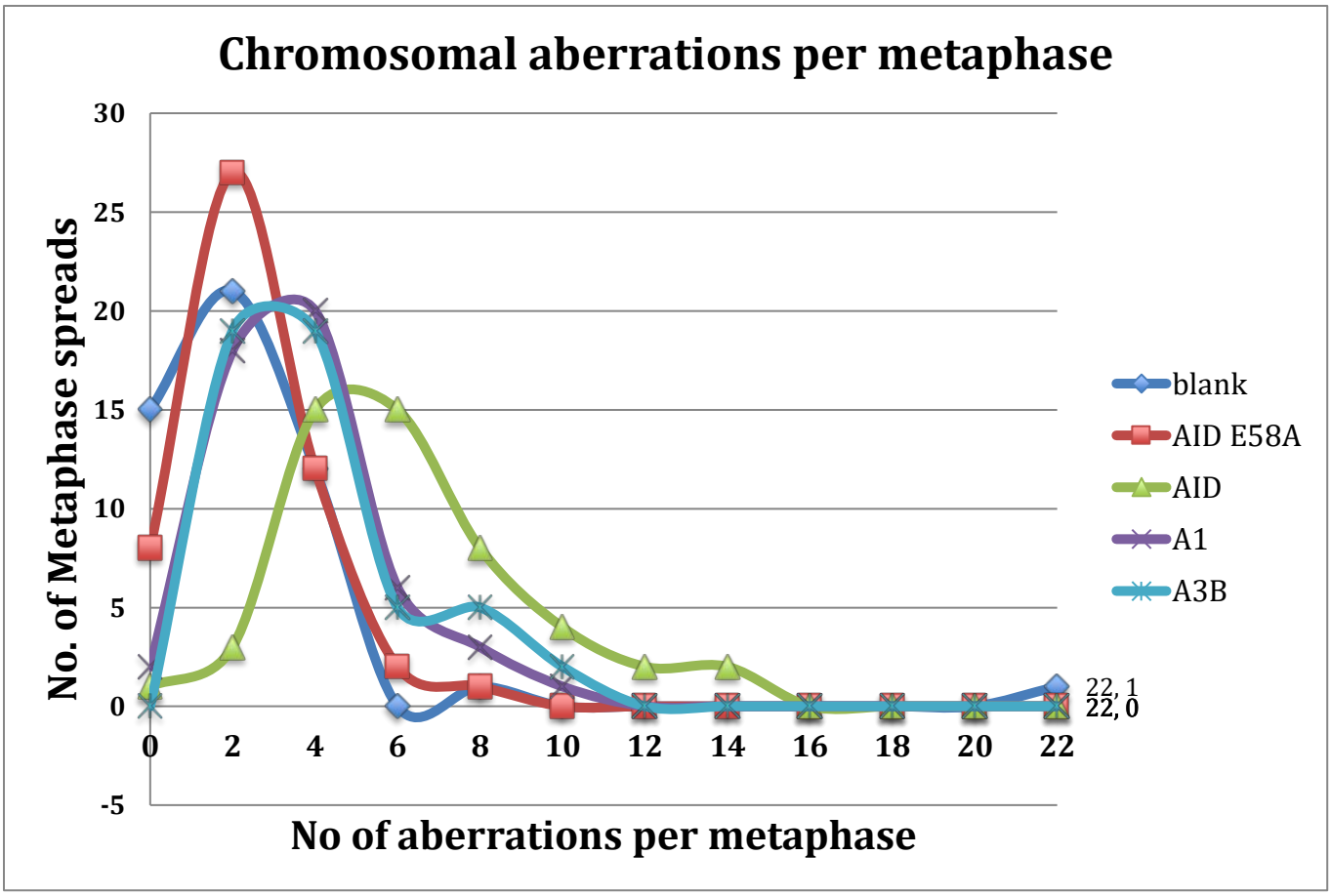


Figure.24 a: AID, APOBEC1 and APOBEC3 causing chromosomal instability. Metaphases were scored for chromatid-type, chromosome-type, and other (dicentrics, rings, anaphase bridges, and pulverized chromosomes) aberrations. The graph shows the total number of aberrations per metaphase spread in HT 1080 HAC cells. (t-test was performed for all the samples; AIDE58A = 0.87, AID = 2.2E-08, APOBEC1 = 0.01 and APOBEC3B = 0.0017).

b)

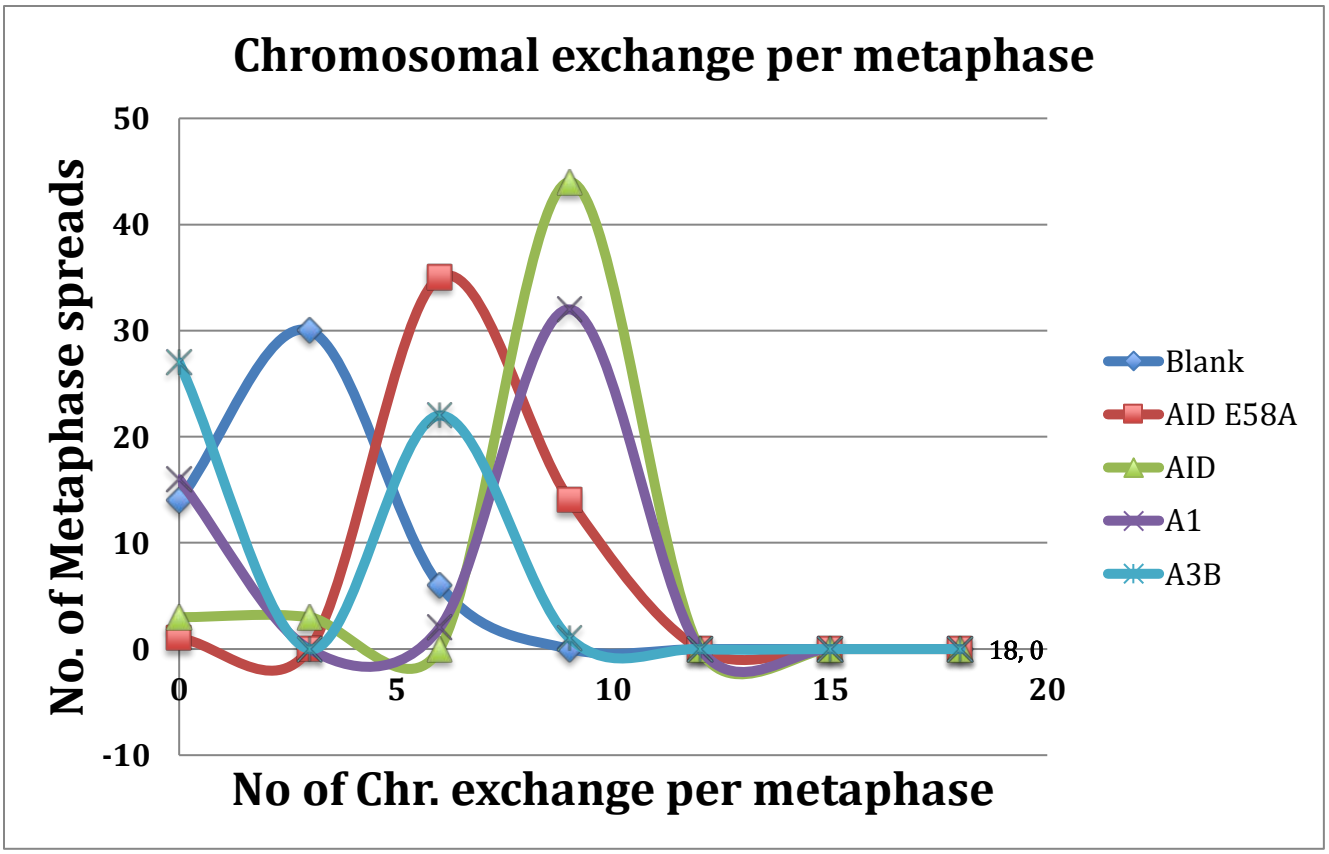


Figure.24 b: AID, APOBEC1 and APOBEC3 causing chromosomal instability. Metaphases were scored for chromatid-type and chromosomal exchanges were counted. The graph shows the total number of exchanges counted in each sample (t-test was performed for all the samples; AIDE58A = 0.2, AID = 1.4E-07, APOBEC1 = 1.2E-10 and APOBEC3B = 2.6E-5).

The persisting elevated levels of chromosomal aberrations by AID/APOBECS evidences how the cytosine deaminases may promote localized hypermutation resulting in massive chromosomal rearrangements. It is rather notable to see how AID overexpression results in massive chromosomal aberration and exchanges considering AID is the last active deaminase among those in mutator assays (*figure 24 a and b*). Moreover, as earlier works showed how AID is responsible for translocations at multiple genomic regions (Klein *et al.*, 2011); this work highlights the capability of the other cytosine deaminases as well in inducing chromosomal breaks.

Furthermore, there have been occurrences of radial chromosomes (RA), which have been previously described as an active indicator of CIN (Diffley *et al.*, 2000). This incidence may serve as a response to DSBs during replication. Moreover, RA maybe the result of DNA damage and stalled replication fork.

In addition to RA, a substantial number of acentric fragments(AF) were observed. These are chromosomal fragments which are deficient of spindle attachment organelles. In a dividing cell, most of the AF are excluded from the main daughter nuclei and are known to form micronuclei. Several studies have underscored MN as a consequence of DNA damage, errors in DNA replication or repair which generate acentric chromosome fragments. The cytological evidence reported here suggests a strong correlation of AID/APOBEC inducing DNA damage leading to AF and other chromosomal aberration. Based on these findings, I sought to test the possibility of AID, APOBEC1 and APOBEC3B as inducers of micronuclei.

2.3 AID/APOBECs induce micronuclei.

An ideal cell division calls for equal segregation of chromosomes in forthcoming daughter cells. Errors in mis-segregation have been largely associated to aid tumorigenesis. Moreover, the ability of AID/APOBECs to cause chromosomal breaks hints it progression towards chromosomal instability. Spindle assembly checkpoint (SAC) is one of the ideal targets to induce CIN (*Figure.25*). The MN are also seen as a precursor for DNA damage, causing mutations and chromosome rearrangements (Crasta *et al.*, 2012).

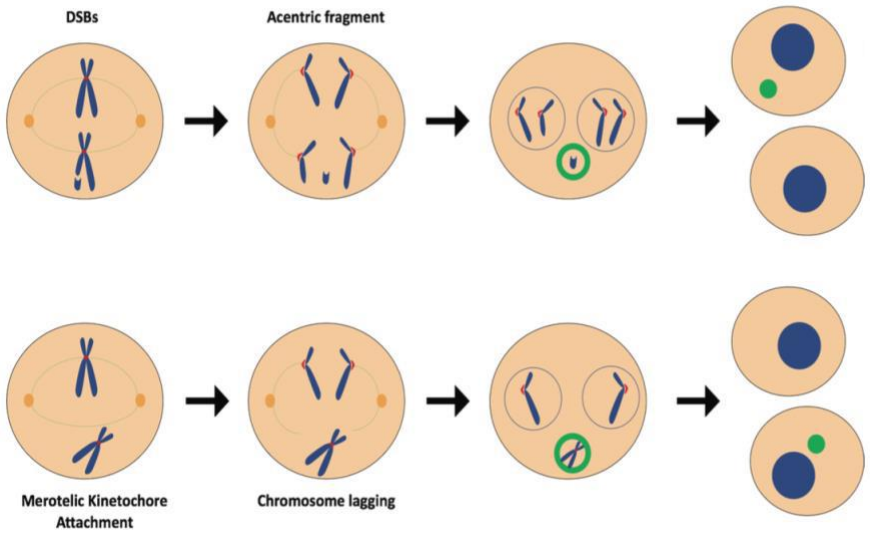


Figure.25: Micronuclei formation, an indicator for CIN.

To determine whether AID/APOBECs are inducers for CIN, I sought to use a quantitative assay to find micronuclei frequency. This quantitative assay for chromosome mis-segregation was developed by exploiting the human artificial chromosome (HAC) present in human fibro-sarcoma (HT1080 cells) where human kinetochore could be conditionally inactivated, leading to formation of micronuclei (Nakano *et al.*, 2008).

In this system, I used the cell line developed by Masumoto's lab; where the epigenetic state of chromatin within an active kinetochore could be controlled using a human artificial chromosome (HAC). The kinetochore assembly on the HAC can be impeded by simple doxycycline washout; thus, rendering the HAC in an inactive state where it fails to attach to the mitotic spindle and is left behind in the anaphase thus aiding the formation of micronuclei (*Figure.26*).

Initially, five cytidine deaminases (AID, APOBEC1, APOBEC3A/B/G) were evaluated for their potential to induce MN formation. These human paralogs were cloned in mammalian vectors under the control of the *β-actin* promoter bearing an internal ribosomal entry sequence for green fluorescence protein expression and puromycin resistance cassette This assay is a robust tool for scoring MN in bi-nucleated cells.

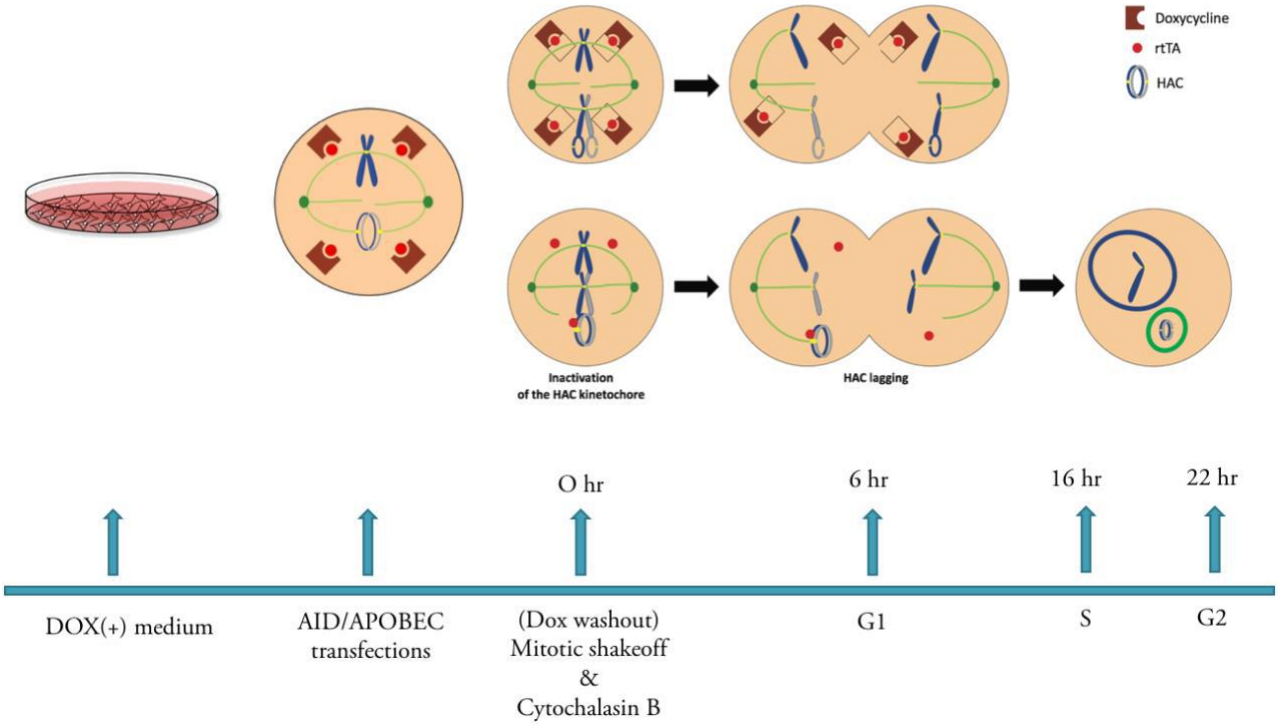


Figure.26: Workflow of the micronucleus assay with AB2.2.18.21 cells (HT1080) HAC cells.

The HT 1080 HAC cells were transiently transfected with AID/APOBECs, followed by enrichments for the transfected cells with puromycin while the presence of a functional kinetochore, the HAC is continuously maintained as a non-essential 47th chromosome that replicates and segregates like a normal chromosome in human cells. This is followed by doxycycline washout, leading to inactivation of the kinetochore, and treatment with Cytochalasin B (CytoB), an inhibitor of the mitotic spindle that prevents cytokinesis, this treatment synchronizes the cells and, due to the interference with mitosis, further aids MN formation. The cells were later washed with PBS and then fixed in ice cold (v/v) methanol: acetic acid (3:1) followed by Giemsa staining.

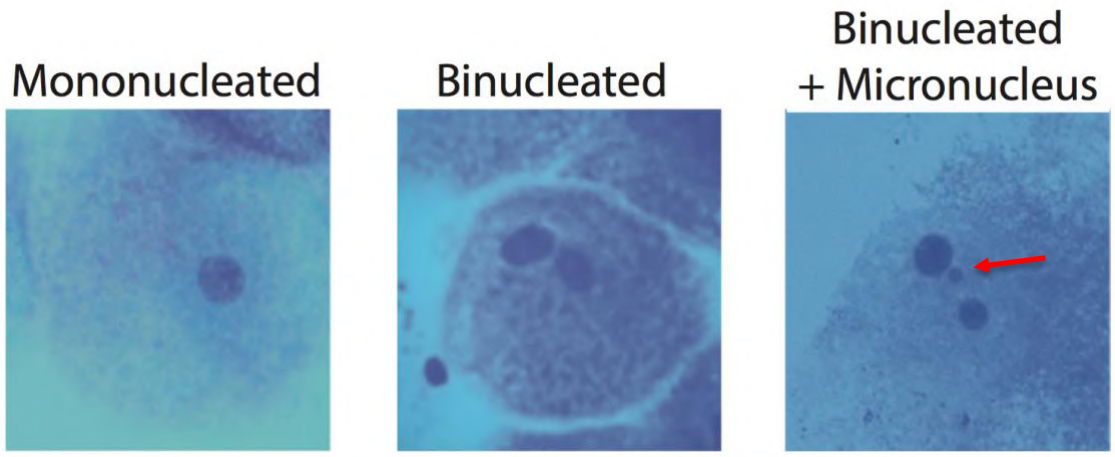


Figure.27: HT 1080 HAC cells illustrating variation amongst them. Red arrow indicate micronucleus.

Pilot experiments were performed on two sets of HT-1080 HAC cells with two different cytoskeleton destabilizers: Nocodazole ($100ng^{-ml}$) or Cytochalasin B ($100ng^{-ml}$) for 6 hours to synchronize for mitosis. The absence of doxycycline enables *TetR* binding to *TetO*, which induces inactivation of the HAC centromere by *tTS* and the subsequent formation of a micronucleus after release from mitotic arrest (*Figure 26*). Cells released from mitotic arrest were stained with Giemsa for micronuclei frequency analysis. Initially, 100 bi-nucleated cells were scored for MN. For the MN scoring, the following Criteria for was set: (1) Looking only for Bi-nucleated cells; (2) the diameter of the MN should be less than one-third of the main nucleus; (3) MN should be separated from or marginally overlap with main nucleus if there is clear identification of the nuclear boundary; and (4) MN should have similar staining as the main nucleus (*figure27*).

Interestingly, the set treated with nocodazole displayed heavy nuclear fragmentation, making it extremely challenging to score MN formation. This might have been due to effect of nocodazole resulting in possible disassembling of focal adhesion, making this treatment too toxic.

On the other hand, cells treated with Cytochalasin B, didn't display nuclear fragmentation, making them more accessible to score MN. As a control to measure the basal rate of spontaneous micronuclei formation, I used Dimethyl sulfoxide (DMSO) as a

vehicular control in un-transfected cells; while cells treated with CytoB, served as a positive control.

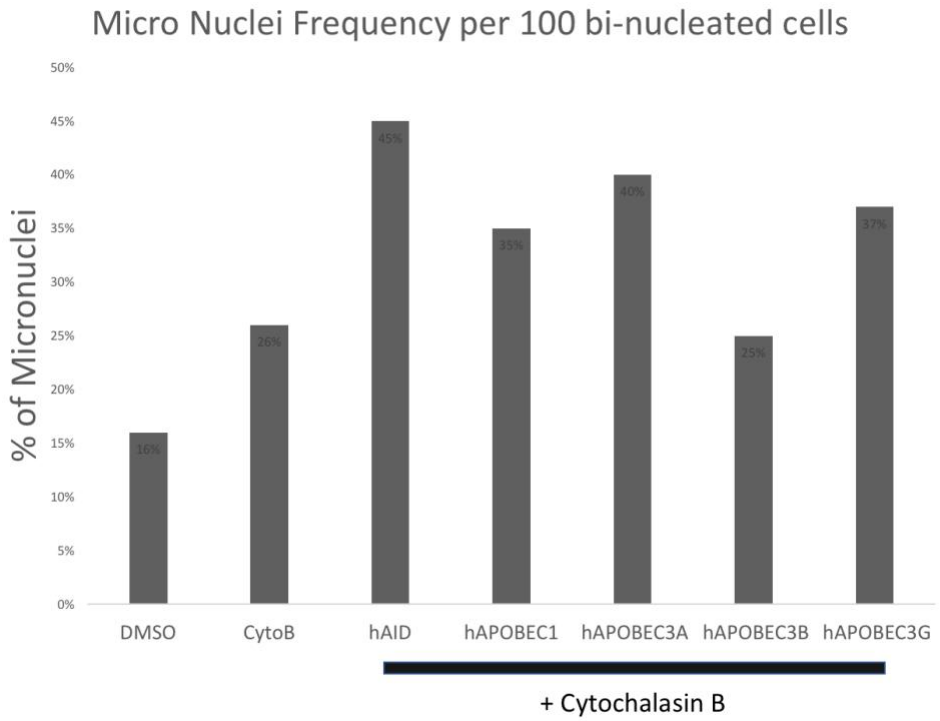


Figure.28: Micronuclei frequency for 100 bi-nucleated HT 1080 HAC cells transfected with AID/APOBECs followed by simultaneous doxycycline washout and CytoB treatment (S.D= 0.1).

As seen in the graph in *figure 28* above, it appears that the cytidine deaminases probably have the capability to induce MN, especially in case of AID, APOBEC1 and APOBEC3A and 3G, where they were nearly 50% higher than that to samples treated with Cyto B. Considering it was a pilot experiment, subsequent experiments tests were carried out through the double-blind method.

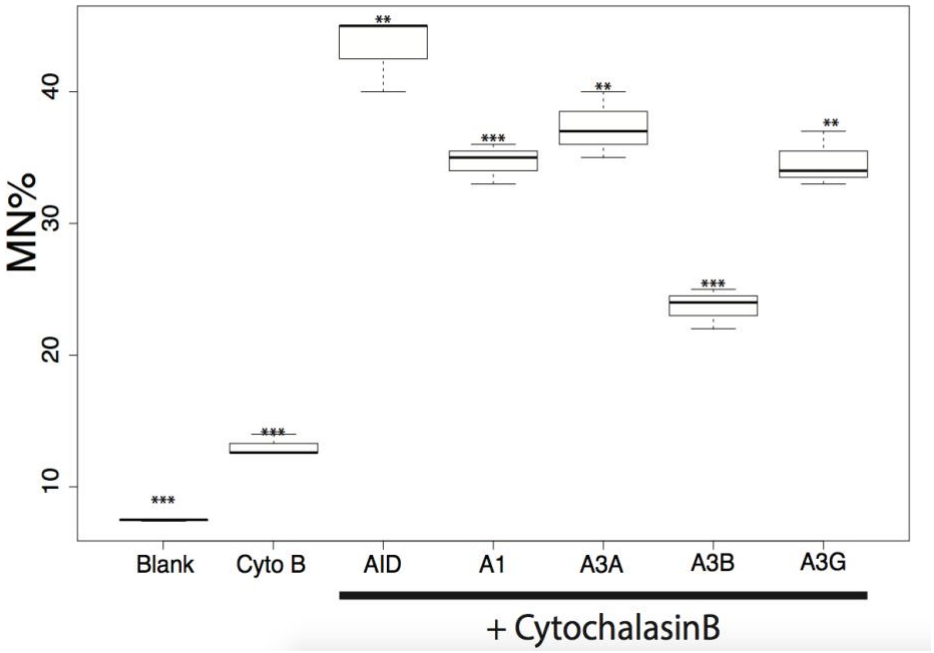


Figure.29: Micronuclei frequency observed for 200 bi-nucleated HT 1080 HAC cells transfected with AID/APOBECs, in presence of 100 ng^{-ml} *CytoB (P=0.0047).*

Here, as a control to measure the basal rate of spontaneous micronuclei formation, MN were counted solely in un-transfected (Blank) cells. While, 100 ng^{-ml} CytoB, was added to further purport the rate of lagging chromosomes. Remarkably, increased levels of micronuclei after transfection with AID, APOBEC1, APOBEC3A and APOBEC3G (in presence of CytoB) compared to those obtained in cells treated only with CytoB were seen (*figure 29)*. Thus, in order to fully assess the potency of AID/APOBECs, in inducing formation of MNs, it seemed rather coherent to solely express AID/APOBECs without treatment of CytoB.

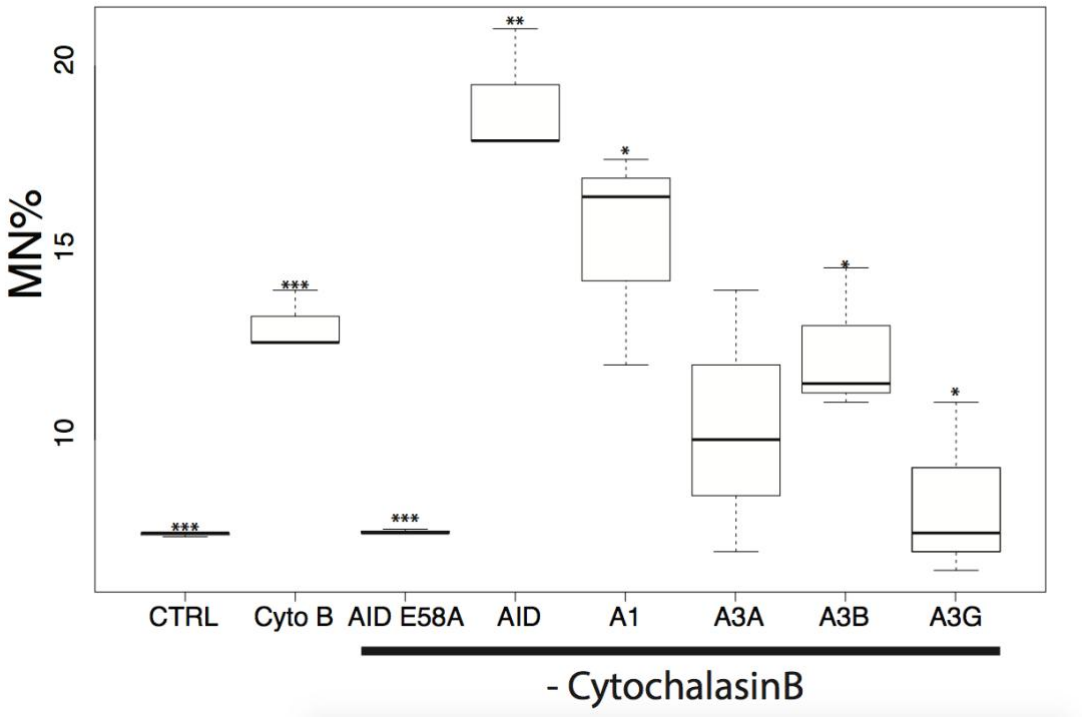


Figure.30: Micronuclei frequency observed for 200 bi-nucleated HT 1080 HAC cells transfected with AID/APOBECs, in the absence of CytoB (P=0.01).

Moreover, as a control for potential artefactual effects, I used either an empty plasmid or a catalytically inactive mutant of AID (AID E58A). Quite understandably, the micronuclei levels went down, due to the lack CytoB treatment hampering the actin filaments. Even in absence of co-treatment with CytoB, elevated levels of MN were observed especially in the case of AID and APOBEC1, where an increase of 50% was observed in the case of cells transfected with AID compared to those treated with CytoB (*Figure 30*). These results clearly demonstrate that AID and APOBEC1 are involved in the induction of MN and indeed prompt MN formation.

It could be that AID/APOBECs induce micronuclei, which may possibly cause DNA fragmentation as a result of erroneous DNA replication transcription, and nuclear envelope (NE) structure, causing random joining of DNA fragments fuelling CIN. Unprompted cytosine deamination is known to create uracil lesions which are repaired by BER, promptly initiated by uracil-DNA glycosylases. As BER primarily relies on UNG, the central cellular glycosylase to identify uracil's in SHM and CSR (Krokan and Bjørås,

2013), I investigated further to validate whether the induction of C to U changes and the downstream activation of the DNA repair pathways associated to this type of damage is related to MN formation. Further experiments were done using an inhibitor of the Uracil-DNA glycosylase (UNG) by using a mammalian vector expressing a bacterial UDG inhibitor (UGI).

UNG is an enzyme known to recognize the DNA damage induced by AID where it prevents mutagenesis by eliminating the uracil introduced by the deamination; thereby initiating Base Excision Repair (BER). Here, AID and the APOBEC 1 were co-transfected with a plasmid encoding for UGI (Uracil Glycosylase Inhibitor).

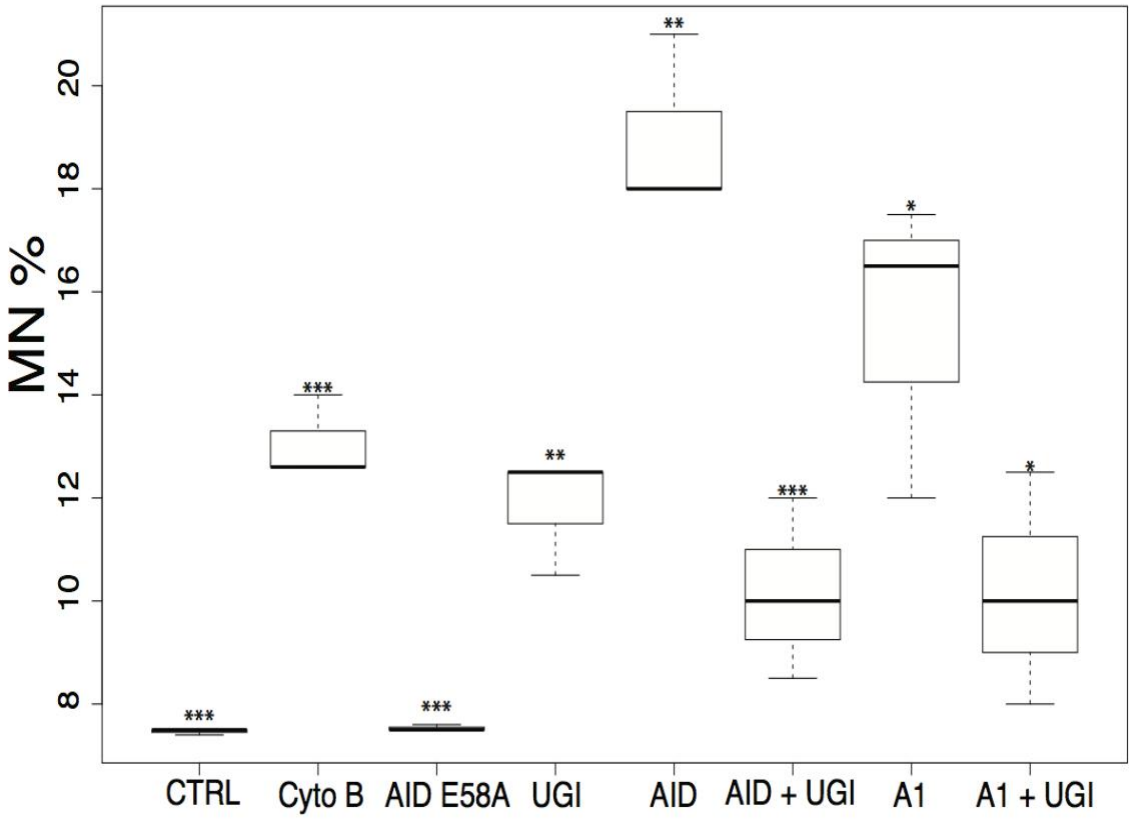


Figure.31: Micronuclei frequency for 200 bi-nucleated HT 1080 HAC cells transfected with AID/APOBECs the inhibitor for Uracil-DNA glycosylase(UGI) followed by simultaneous doxycycline washout and (P=0.002).

Diminution of the levels of MN formation was observed in cells expressing AID and APOBEC1 in the presence of UGI. This highlights the fact that AID and APOBEC1 promote DNA damage through C to U deamination, thus affecting the MN frequency in the presence of inhibitor of UNG. The result obtained confirms that AID/APOBEC1 induces MN via deamination. Intriguingly, elevated levels were also observed in cells transfected with solely UGI (*figure 31)*. This could have been due to possible interference between the inactivation of the kinetochore leading to inactivation of the artificial centromere, which leaves the centromere proteins (CENP) loose leading to possible MN formation. A study by *Zeitlen et al* in 2005 reported that in xenopus eggs, when inhibited with UGI, resulted in undetectable levels CENP A on the sperm DNA, suggesting a possible BER role in the assembly of the histone H3 variant (Zeitlin *et al.*, 2005).

Considering that the AID/APOBECs have been associated to the onset of cancer through their ability to mutate DNA, our finding shows another possible cancer-inducing effect of these deaminases: their ability to induce chromosomal instability. Addressing the conditions underlying CIN and the possible role of the AID/APOBEC1 as key factors is important to understand better the onset and progression of cancer. Moreover, their mutagenic activity might be relevant not only for MN formation, but also for the onset of genetic alterations related to MNs.

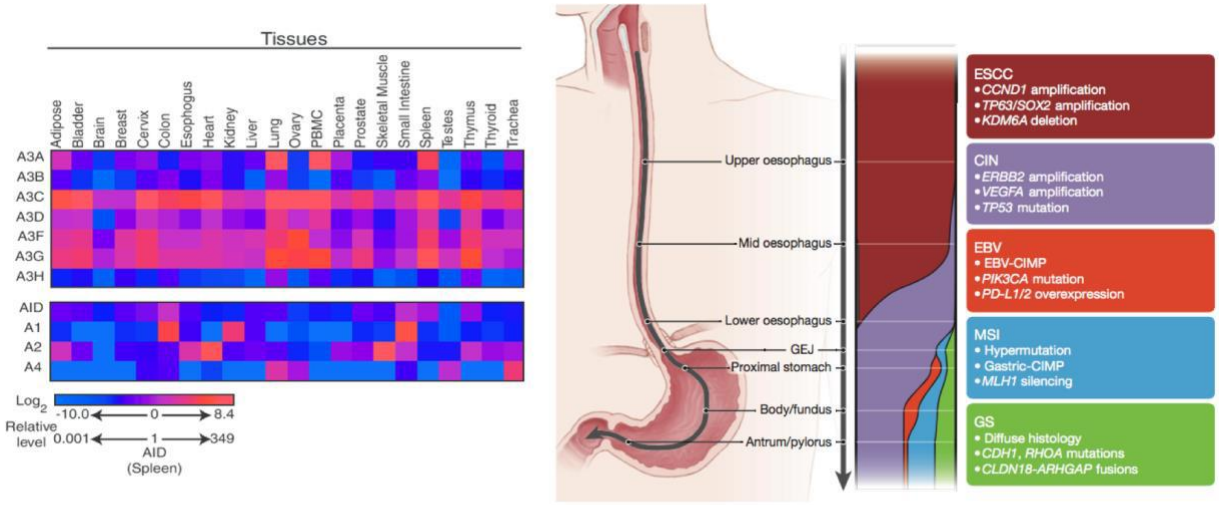
In order to illuminate further intrinsic roles, there is a need for further investigation at biochemical, genetic and the bioinformatics level.

2.4 AID/APOBECs and CIN in cancer: The *In-silico* approach

Considering the findings with the AID/APOBECs and CIN, an *In-silico* approach can be useful to identify the possible role of deaminases as key factors in onset and progression of cancer. In order to address the role of AID/APOBECs as chronic source for CIN-positive tumors, I considered the 'The Cancer Genome Atlas (TCGA) project' which catalogues genetic alterations responsible for cancer using genome sequencing and bioinformatics tools, in order to understand the genetic aspect of the disease.

AID/APOBECs are known to be expressed across various tissues, more evidently in bladder, lungs oesophagus, small intestine etc. (*figure 32a*). Lately, the TCGA

consortium had focused on characterizing gastric adenocarcinoma, where they classified gastric cancer into four subtypes, namely: a) Epstein-Barr virus (EBV), B) microsatellite unstable tumours (MIN), C)genomically stable tumors and D) tumors with chromosomal instability CIN) (Cancer *et al.*, 2014).



A) mRNA expression of AID/APOBECs across all tissues

B) Types of gastroesophageal carcinoma from the proximal oesophagus to the distal stomach.

Figure.32: A) Heat map of AID/APOBECs according to tissue type.(modified from Burns et al.,2013) B) Distribution of esophageal carcinoma (Cancer et al., 2016).

This was followed up with another report on genomic characterization of esophageal carcinoma. Considering the fact that the AID/APOBECs are expressed in these tissues, it seemed as the ideal sample set to investigate. Esophageal cancers have been plagued with low survival rates and are primarily classified into adenocarcinoma (EAC) or squamous cell carcinoma (ESCC). They had performed a wide molecular analysis of 164 esophageal tumors, 359 gastric adenocarcinomas and 36 additional adenocarcinomas at the GEJ and characterized tumors into defined groups (*figure 32b*).

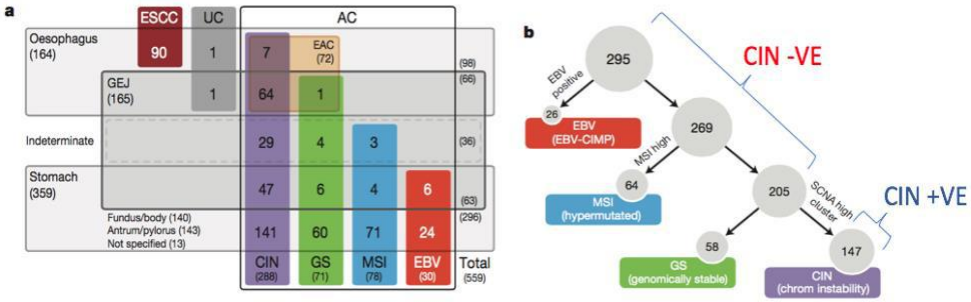


Figure.33: Major subdivisions of gastroesophageal cancer a) Division of patient tumor data according to molecular subtypes. B) Subdivision of tumors based on SCNA where EBV, MSI,GS positive tumors are labelled as CIN negative (modified from Cancer Genome Atlas Research Network et al., 2017).

Initially, they had evaluated EACs with gastric cancers and observed that CIN tumors were remarkably different from EBV, MSI or GS tumors. Moreover, in gastroesophageal adenocarcinomas (GEAs), they found CIN prevalence with 71 of 72 EACs classified as CIN, while being negative for MSI or EBV (Cancer *et al.*, 2014; Cancer Genome Atlas Research Network *et al.*, 2017).

Using this information, I carried out the analysis where I sought to evaluate the expression of the AID/APOBECs in tumor sample set from the TCGA consortium. I pre-defined the samples as CIN +Positive and CIN –Negative for gastric cancer (*figure 33)*.

Firstly, from the database, I tried to match the samples to the expression of AID, APOBEC1 and APOBEC3B. I manually curated a list of the said samples where the patient barcode was arranged according to tumor type, location, status of CIN, etc. (*supplementary data 2*). There was difficulty in extrapolating the data especially case of in patient retrieval, as there were no normal samples to match against (possibly due to the updates in the database). After re-observing at the workflow, I analyzed their data and linked the samples with the expression data set present in metadata to retrieve the sample ID of patients and curated another data list based on the ESCA and STAD metadata from GDC (https://portal.gdc.cancer.gov). With the obtained metadata, keys were generated for ESCA and STAD (supplementary data 3); and matched using a *perl* script (*supplementary data 3*). The resulting mutation dataset match, (Genome Reference Consortium build 37 (GRCh37) (GRCh38), enabled me to observe expression levels of AID, APOBECs based on tumor type, status, FPKM count and its histological type to obtain a curated dataset (supplementary data 3).

In stomach adenocarcinoma (STAD) samples the expression profiles of the entire deaminase family in CIN (+), CIN (-) were matched against the normal samples of the same. Here, a wide variation was observed amongst individual samples, reflecting different patterns towards CIN and tumor. There was very little to no difference for AID, and APOBEC 1. This is probably due to low expression levels, when compared to the

normal sample, hinting towards an unconstructive selection per se (*figure 34*). Interestingly for APOBEC3B, all the CIN tumor samples showed elevated expression when matched to the normal samples (*figure 34*). This is in agreement with previous reports suggesting an increased mutational load for APOBEC3B in stomach cancer (Burns, Temiz and Harris, 2013). This puts an interesting twist to the deaminases conundrum, where it leaves one wonder, if they work in a sort of assisted manner. Also, numerous factors including the availability of ssDNA substrate and the expression level of APOBECs could contribute to the extent of APOBEC mutagenesis in CIN tumors.

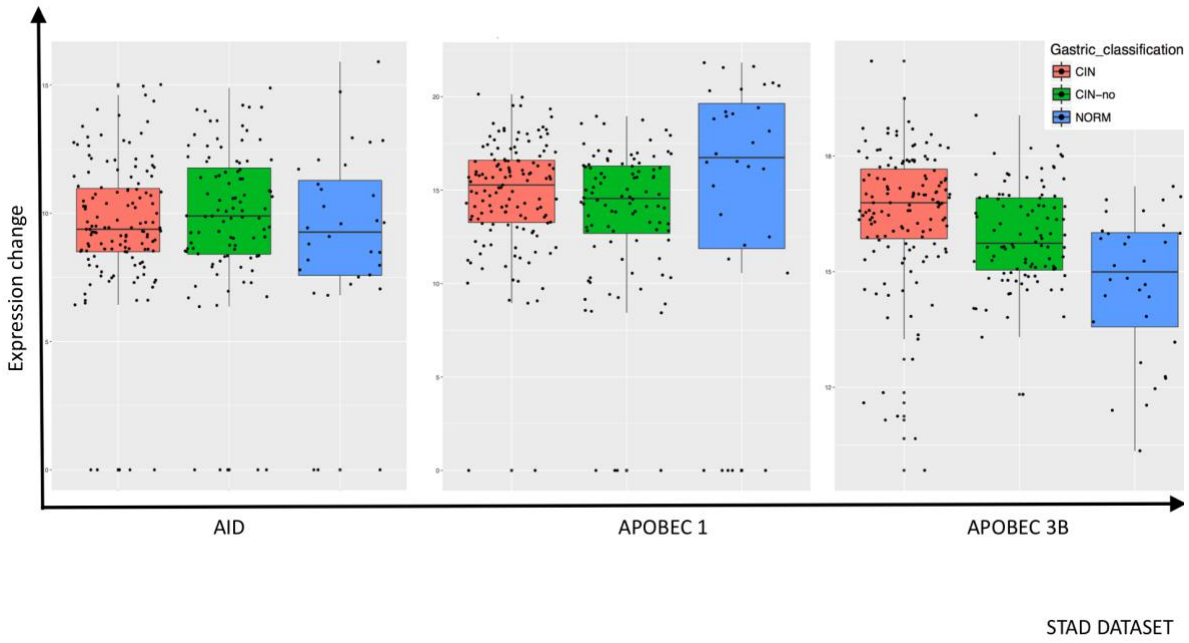


Figure.34: AID, APOBEC1 and APOBEC3B gene expression in STAD. The tumor samples have been classed into CIN positive(red) and CIN negative (green), these are matched against normal tissue samples.

While in case of ESCA, a positive correlation was observed when compared to the normal sample. Although not very significant, it does seem like the AID and APOBEC3B are upregulated in CIN tumor (*figure 35*). Correlations with APOBEC1 could be justified as the normal samples exhibited very low coverage making it difficult to fully comprehend its expression.

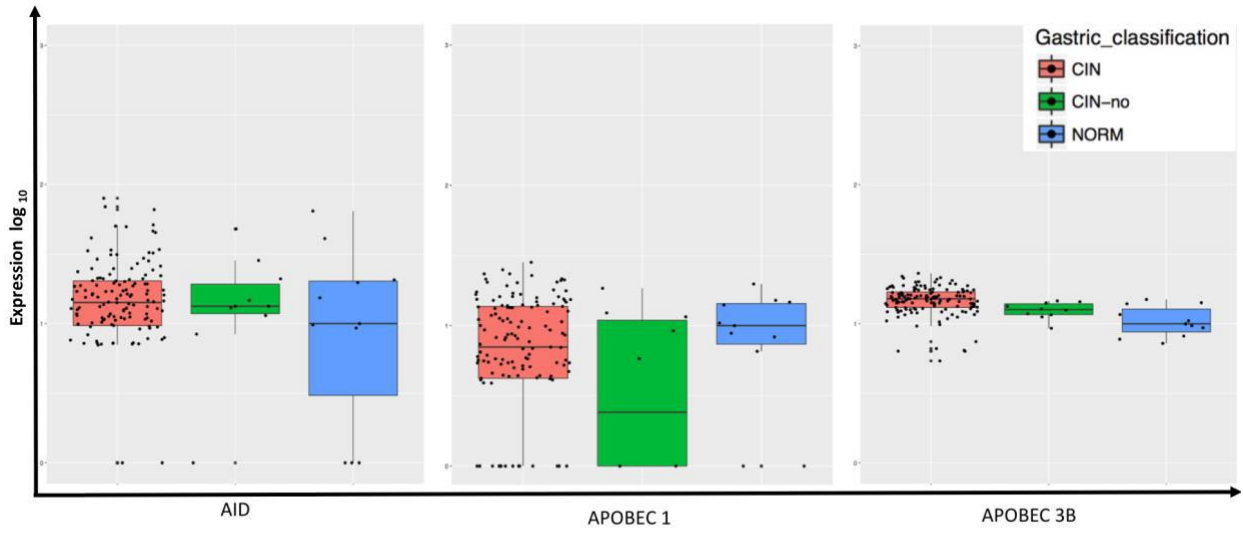


Figure.35: AID, APOBEC1 and APOBEC3B gene expression in ESCA. The tumor samples have been classed into CIN positive (red) and CIN negative (green), these are matched against normal tissue samples.

From the results obtained there is definitely an effect of deaminases seen in both CIN and vice-versa, underscoring how AID/APOBECs may have different functional roles in tumor progression. Earlier studies in TCGA samples has suggested that around 18% of them exhibited a strong APOBEC signature; in some cases up to 70% (Henderson et al., 2014; Nordentoft et al., 2014). Understanding CIN on the basis of expression levels is a tricky, considering the glitches faced with obtaining a normal set to match against. Currently, I am devising a workflow where I could exemplify the links better. Patterns of gene expression in tumors are result in genomic instability, which are defined by accumulation of somatic copy number alterations (SCNAs) and point mutations (PMs). Expression signatures have always been associated with always been associated with them as it gives an opportunity to explicate biological mechanisms. Ideally, across multiple cancer types I would investigate, expression of AID, APOBEC1 and APOBEC3B with the accumulation of PMs and SCNAs to CIN associated tumors.

3. Discussion

It has been demonstrated for the first time how AID/APOBECs promotes chromosomal instability by deamination. This highlights the intricate working of the cytosine deaminase in the cell cycle and how the family members work at different phases of the cell cycle to trigger DNA damage resulting in aberrations. AID is known to be fully functional in the G1 phase of the cell cycle, as the DSBs generated at the Immunoglobulin locus are strongly dependent on AID, while in the case of APOBEC 3B, it could very well be associated with delayed cell cycle arrest resulting in DNA fragmentation caused by its deamination activity (Petersen *et al.*, 2001; Burns *et al.*, 2013).

The reported results suggest a working model for MN formation induced by AID/APOBECs (*figure 36)*. Primarily, deamination by the AID/APOBECs at the genomic levels results deoxyuridine, which are immediately fixed by either BER or MMR, while erroneous repair leads to DNA mispairing. Whereas in case of staggered DSBs translocation of the chromosome is a common outcome. This leads to massive genomic rearrangements which directs for the formation of the micronuclei. Furthermore, when the staggered DSBs are carried over in the S phase, the HR-mediated repair resects the SSBs. This yields to vulnerable ssDNA being exposed to AID/APOBECs, further rendering higher mutational loads and fuelling MN formation.

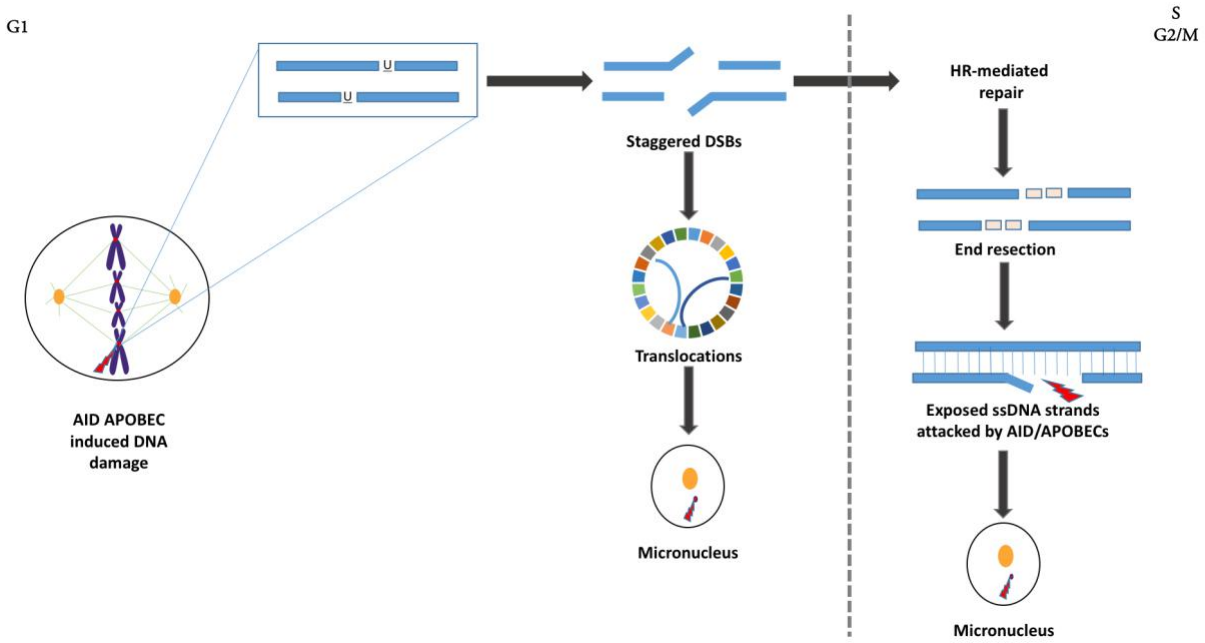


Figure.36: Working model for MN induction by APOBECs in cell phases.

Genetic alterations are the main cause of variability within organisms and the basis for genetic disease and cancer. Among the agents responsible for this variability are the AID/APOBECs, DNA/RNA editing enzymes active in innate and adaptive immunity pathways. Their action has been linked to the rise of mutations and chromosomal alterations found in cancers. Indeed, chromosomal mis-segregation resulting in Chromosomal instability (CIN) plays a major role in tumorigenesis, and -so far- only indirect evidence linked the AID/APOBECs to CIN.

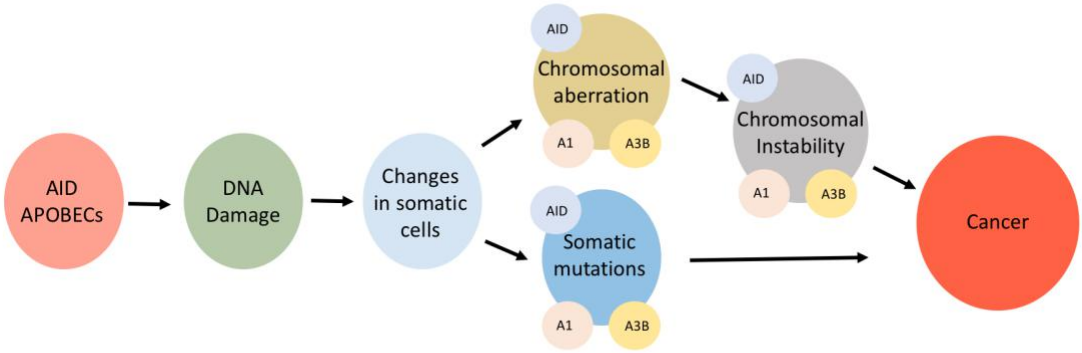


Figure.37: AID/APOBECs induces genetic alterations leading to cancer.

Micronuclei (MN) are an effective cellular indicator of CIN, as elevated frequencies of MN are observed in most solid tumors and pre-neoplastic lesions. Addressing the conditions underlying CIN and the possible role of the AID/APOBEC1 as key factors are important to understand better the onset and progression of cancer. Moreover, their mutagenic activity might be relevant not only for MN formation, but also for the onset of genetic alterations related to MNs. Considering that the AID/APOBECs have been associated to the onset of cancer through their ability to mutate DNA, these finding shows another possible cancer-inducing effect of these deaminases: their ability to induce chromosomal instability (*figure 37*).

4. Introduction (Part II)

4.1 Genome editing

Altering nucleotides of the genome with engineered nucleases in living organisms is genome editing. In the past decade, several nucleases have been developed and has radically changed the field. These enzymes include polymerases, restriction endonucleases and DNA ligases for *in vitro* DNA manipulations and recombinant DNA technology. The genome editing field has been vastly accelerated thanks to the use of zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and most recently the systems based on Clustered Regularly-Interspaced Short Palindromic Repeat (CRISPR) systems, holding great potential for clinical applications as gene therapy.

4.2 Zinc finger nucleases (ZFNs)

These nucleases were found way back in 1996 and have been subsequently used in gene editing experiments in mammalian cell (Kim, Cha and Chandrasegaran, 1996; Bibikova *et al.*, 2002, 2003). The ZFNs are known to function via a DNA/protein recognition motif which are comprised of a zinc finger mediated binding domain that recognizes for the site of interest, fused with the nuclease domain of *FokI* used for cleaving the target sequence (*Figure 38)* (Strobel and Dervan, 1990; Kim, Cha and Chandrasegaran, 1996; Gottesfeld *et al.*, 1997; Bibikova *et al.*, 2003). ZFNs causes DSBs which can be exploited to modify or delete a specific site at the genome. The insertion and deletions are induced by incompetent NHEJ-mediated repair, while in case of point mutations, donor templates are introduced by a more functional HDR-mediated repair (*Figure 38)* (Gaj, Gersbach and Barbas, 2013) . For a brief period of time, the ZFN protein technology was the only means available to create customized site-specific DNA binding proteins and enzymes. The ZFN-based technology was limited by disadvantages primarily related to engineer their design and, due to the necessity of creating a new ZFN for each specific locus to be targeted.

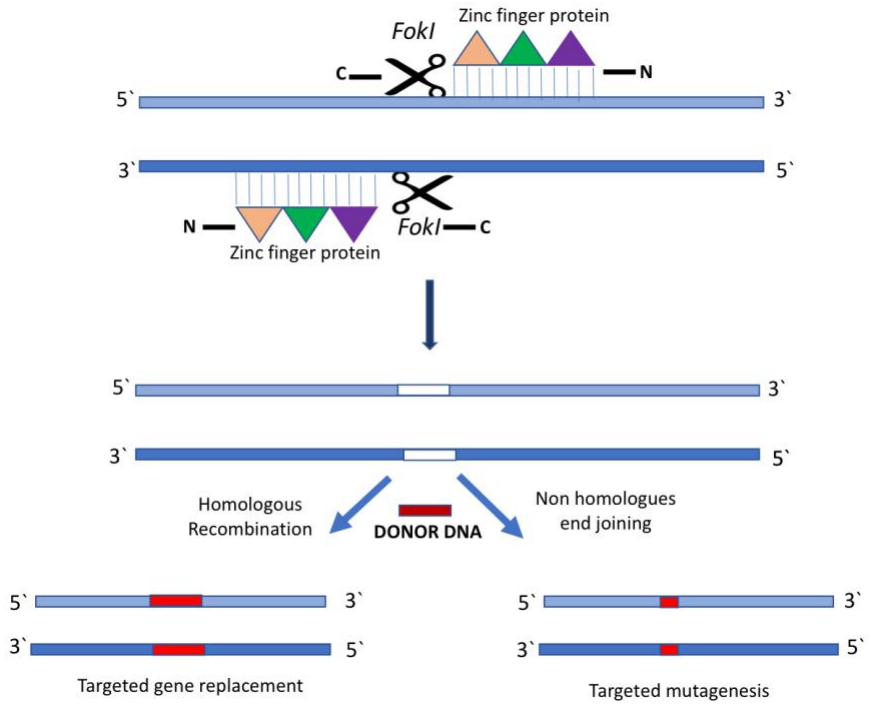


Figure.38: Schematic overview of the zinc-finger nuclease (ZFN) dimer bound to DNA. ZFN target sites comprises of two zinc-finger binding sites separated by a 5- to 7-bp spacer sequence recognized by the FokI cleavage domain.

4.3 TALEN

Compared to ZFN-based technology, the TALENs (Transcription Activator-Like Effector Nucleases) are fairly modular, easing feasibility in design and construct. TALEs were originally found in *Xanthomonas* bacteria, where they contain DNA binding domains which are approximately 33-35 amino acids and serial modules that recognizes a single base-pair (*Figure 39)*. The specificity of TALE is determined by two hypervariable amino acids, the repeat-variable diresudues (RVDs) (Deng *et al.*, 2012; Mak *et al.*, 2012). Similar to the ZFN, the modular TALE repeats can be customized to identify contiguous DNA sequence. With innovations in the of ZFNs, there have been numerous effector domains found to improve the targeting of TALEN; these include fusing the TALE repeats to nucleases, transcriptional activators and site specific recombinases (Christian *et al.*, 2010; Miller *et al.*, 2011; Zhang *et al.*, 2011; Mercer *et al.*, 2012). These modular proteins offer single base recognition making them much more flexible than ZF-proteins. Yet, the assembly of TALE arrays is cumbersome, giving them a technical disadvantage.

Several techniques have been developed in order to assemble them in a less-laborious fashion, namely : "Golden Gate" molecular cloning (Cermak *et al.*, 2011), high-throughput solid-phase assembly (Briggs *et al.*, 2012; Reyon *et al.*, 2012) and ligation-independent cloning (Schmid-Burgk *et al.*, 2012). The most promising work done in this field was the development of TALENs targeting every protein-coding gene in the human genome, serving as a pedestal for functional genomic studies (Kim *et al.*, 2013).The only limitation using TALEN *per se* is the need for 'T', which must be at the 5' of the target sequence.

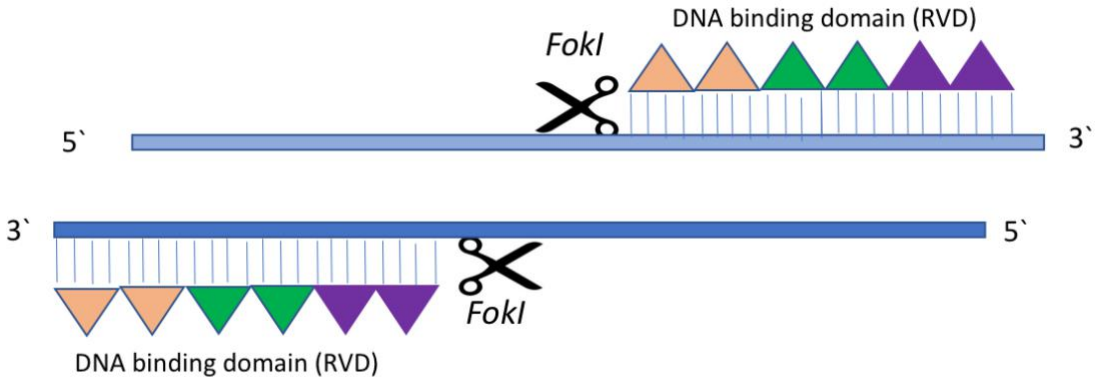


Figure.39: TALE nuclease (TALEN) dimer bound to DNA.

4.4 CRISPR/Cas9 system

Clustered Regularly Interspaced Palindromic Repeats CRISPR/ Associated (CAS) protein system prevails in prokaryotes as an adaptive immune defence against viruses or invading agents in bacteria. It was seen that CRISPR/Cas(when infected) conferred resistance against the lytic phages of *Streptococcus thermophiles* (Barrangou *et al.*, 2007). While mature CRISPR RNAs (crRNAs) were found to function along with the Cas proteins to combat with virus proliferation (Brouns *et al.*, 2008). Moreover, Doudna's lab revealed that the dual-RNA structure formed by the crRNA and the trans-activating RNA(tracrRNA) directs the *Streptococcus pyogenes type II Cas9 protein* (spCas9) to cleave specific target DNA sequences *in-vitro* (Jinek *et al.*, 2012); thus, showed it potential use as a genome editing tool. This was followed by tools developed from

Zhang's and Church's lab, where they set various RNA guided tool with an engineered CRISPR system (Cong *et al.*, 2013; Mali *et al.*, 2013).

In the CRISPR/Cas system, crRNA–tracrRNA or the guide RNA (gRNA) detects the target site on the genome and direct the Cas 9 protein for cleavage at that specific loci (Cong *et al.*, 2013). Here, the gRNA (which is a 20bp crRNA and tracrRNA) is driven by a U6 promoter *in vivo* or by phage RNA polymerase *in vitro* (Cong *et al.*, 2013; Mali *et al.*, 2013; Wang *et al.*, 2013). Once the gRNA are driven by these promoters, it requires a guanine (G) as the first nucleotide for its target site for the U6 driven transcriptions whereas for T7 it requires two G's (*Figure 40*) (Cong *et al.*, 2013; Jao, Wente and Chen, 2013; Mali *et al.*, 2013; Wang *et al.*, 2013). The protospacer adjacent motif (PAM) sequence is a short DNA motif and is the vital component of the Crispr/Cas system as it mediates the targeting of the Cas9. The Cas9 wouldn't be able to cleave the target site if it is not followed by PAM. PAM being absent in the CRISPR DNA, it discriminates bacterial self from non self DNA thus hindering the CRISPR locus from being subjected to the nucleases (Mali *et al.*, 2013).

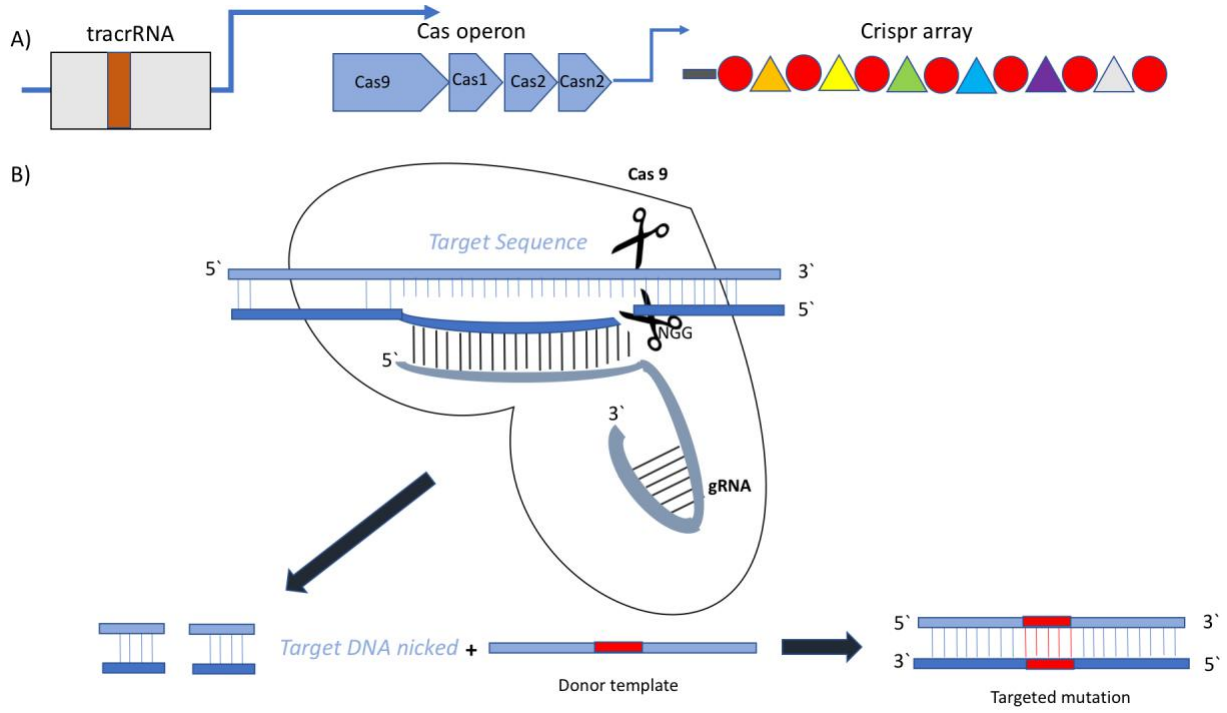


Figure.40: The Crispr/Cas9 system: A) Genomic CRISPR locus b) RNA-guided cleavage of DNA.

The feasibility to design the gRNA makes the Crispr/Cas system very attractive, and is widely used as a tool for genome editing in various organisms to study diseases by correcting them (Doudna and Charpentier, 2014). So far, CRISPR/Cas is being widely used in well-established model organisms. Most of the genome editing approaches have relied on the HR- directed repair where DSBs are induced in the DNA. Most of the time there are indels at the site of DNA cleavage via the NHEJ which leads to several outcomes in comparison to the desired one. This prompted researchers to increase the efficiency of HDR and supress NHEJ, thus highlighting the constant need for development of alternate routes for efficient gene editing and repair. Motivated by this argument, Base Editing was discovered, which involves correction of a point mutation in the genomic DNA.

4.5 Base Editing

This novel approach uses CRISPR/Cas9 to target point mutation without the need of inducing DSBs. This encompasses the usage of a Cas9 fused to a cytosine deaminase and a sgRNA directing the Cas9 at the site of base change (*Figure 41)*(Komor *et al.*, 2016). Initially, Liu's lab had developed three base editors namely, BE-1, BE-2 and BE-3. **BE1** (First generation base editor) a chimera of a catalytically inactive Cas9 fused to APOBEC1 (cytidine deaminase). These functioned with an average editing efficiency of 44% *in vitro,* where cytidines preceded by a T or a C was edited at the highest rate, while this drastically dropped to 0.8-7.7% *in vivo* (Komor *et al.*, 2016). They suspected BER to the culprit for low editing levels. In order to test if this indeed was the case, they fused a bacteriophage uracil glycosylase inhibitor to APOBEC1-dcas9 and called it **BE2**, and found the editing efficiency to be by three-fold. Besides BER, MMR is also known to correct U:G mismatches, while this could be easily stimulated with a nick on one DNA strand permitting the MMR machinery to discriminate the nascent DNA strand (Li, 2008) leading them to construct the **BE3**. Here, APOBEC1 is fused to a nickase Cas9 and a UGI, this allows MMR to be stimulated while keeping BER blocked. Unsurprisingly, higher editing efficiency was achieved (58-75%).

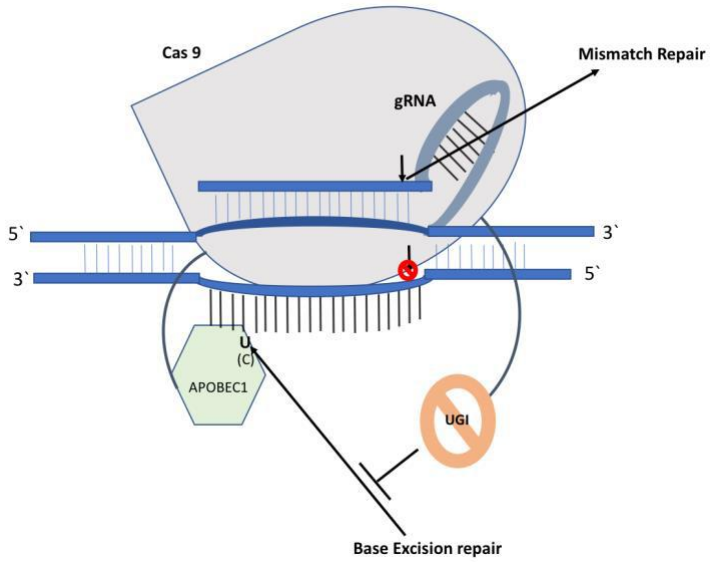


Figure.41: Base editor BE3: Schematic representation of the Base editor 3, where a cas9(nickase)is fused with APOBEC1.When nicked, they signal for the removal of the unedited strand to trigger MMR. The red symbol indicated inability to cleave the unbound strand while APOBEC 1 edits cytosine to uracil on the unbound strand and UGI bocks the activity of uracil glycosylase to prevent corrections by APOBEC1 to be edited.

This was followed by a barrage of Cas9 fusion constructs (where Cas9 was fused with other cytidine deaminase family members) (Nishida et al., 2016; Komor et al., 2017). Further programmable deaminases were generated by fusing them with the ZF or TALE-DNA binding modules (Yang et al.,2016). So far, most of the base editors reported have been known to mediated C:G to T:A conversion, thus insinuating the need of enzymes specifically targeting them and significantly expanding the scope of base editing across the genomic DNA.

4.6 The RNA-DNA editing enigma.

Evolutionarily, organisms are constantly badgered with endogenous and exogenous agents, putting them at risk. Living organisms need to devise a way to combat them and develop strategies for the same; these help them evolve/adapt. These transformations have changed the way we view the genome considering how a set of nucleotides yields RNA intermediate to the dynamic multi-dimensional interplay of RNA, DNA and protein interactions.

Alterations at the genomic or the transcriptomic level may have hazardous consequences. In spite of the innate pressure to guard against such effects, several biological entities developed distinct mechanism that perturb the DNA/RNA sequences and its analogous content. Primarily, there are two set of enzymes that can edit the polynucleotide sequences, namely: cytidine deaminases (known to deaminate cytidine to uridine in RNA/DNA) like APOBECs and adenosine deaminases (deaminates adenosine to inosine in RNA) ADARs (Adenosine Deaminases Acting on RNA) which bind to double stranded RNA (*Figure 42*) (Koito and Ikeda, 2012; Savva, Rieder and Reenan, 2012).

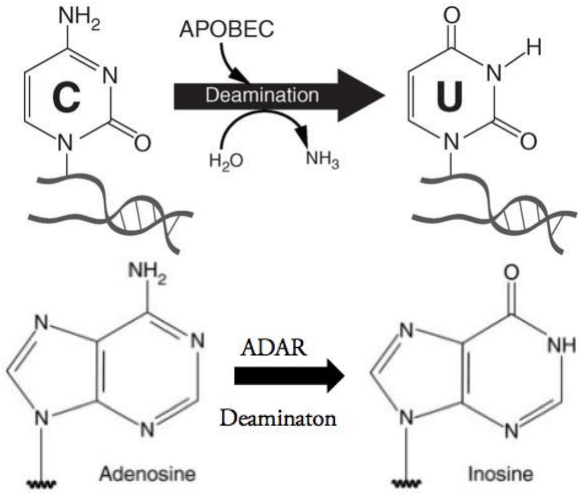


Figure.42: Deamination by cytidine and adenosine deaminases.

RNA editing

It is simple nucleotide change at one or several positions within a RNA transcript. Initial studies reported on RNA editing of the apolipoprotein B (apoB) transcript in mouse intestine and the glutamate-gated ion channel (GluR-B)1 transcript in mouse brain. It was shown that apoB mRNA possesses a UAA translational stop codon while in the genomic DNA points to a CAA glutamine one. While the edited GluR-B mRNA has a CIG arginine codon (I is recognized as G by decoding ribosomes), while the edited one has a CAG glutamine at the same position (‘A novel form of tissue-specific RNA processing

produces apolipoprotein-B48 in intestine', 1987, 'RNA editing in brain controls a determinant of ion flow in glutamate-gated channels', 1991).

Overall, RNA editing operates in two distinct mechanisms: Substitutional and Insertional/deletion editing (*Figure 43*) (Farajollahi and Maas, 2010).

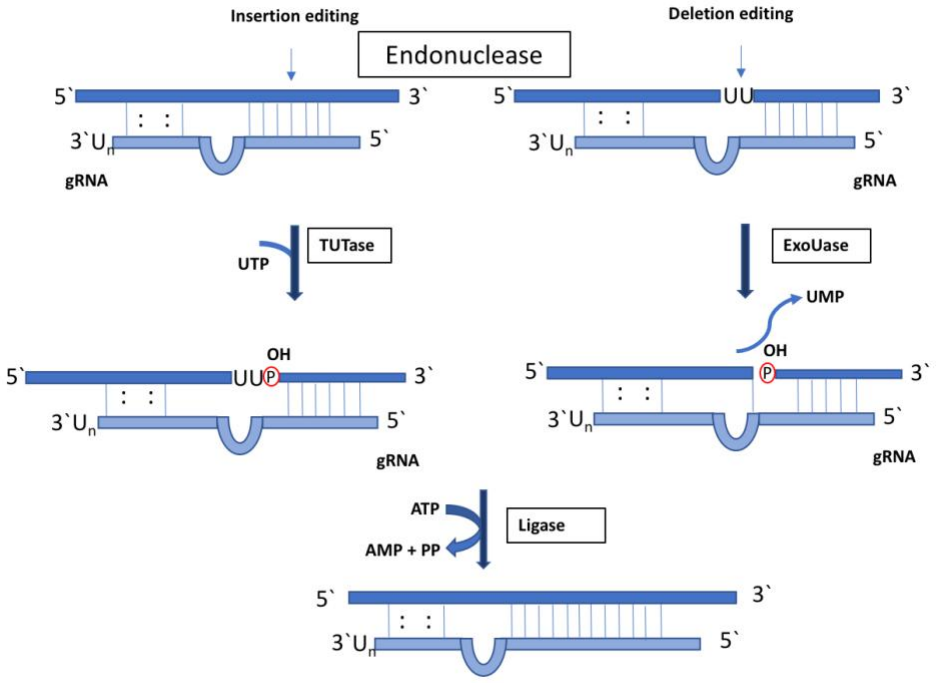


Figure.43: General overview of RNA editing: Insertion and Deletion editing.

This begins with the interaction between the RNAs by Watson–Crick base-pairs (unbroken lines) and G:U base-pairs (colons) determines the sites of cleavage and number of U nucleotides. Later, the gRNAs with 3` oligo(U) tails are added post-transcriptionally where it is facilitated from interactions with pre-mRNA 5` at the editing site. Editing occurs catalytically; endo-nucleolytic cleavage of the pre-mRNA by an endonuclease occurs upstream of the anchor duplex (8–10 bp) between the pre-mRNA and its 'cognate' gRNA(arrow). U are either added to the 5`-cleavage fragment by a *TUTase* in insertion editing or removed by an *ExoUase* in deletion editing, as specified by the sequence of the gRNA. The resultant 50 and 30 mRNA fragments are then ligated by an RNA ligase. Several cycles of coordinated catalytic steps occur until all the sites specified by a gRNA are edited, resulting in complementarity G:U, A:U and G:C base-pairing between the edited mRNA and the gRNA, except at the gRNA terminus. Editing

by each gRNA creates a sequence that is complementary to the anchor region of the subsequent gRNA to be used, thereby enabling the sequential use of the multiple gRNAs that are required to edit the mRNAs in full (Stuart *et al.*, 2005).

Transcriptomic analysis on the RNA editing levels suggested that A to I RNA editing and C to U editing in RNA are modest (Rosenberg *et al.*, 2011). This may be due to the targets being present in the retroviral/repeat containing regions making them harder to look it and the fact that cytosine deaminases seem to work at the DNA level as well (considering the C to U editing of RNA is APOBEC 1 mediated) (Smith *et al.*, 2012).

4.7 ADAR

Adenosine deaminases acting on RNA (ADARs) principally deaminates Adenosine (A) to Inosine (I) in double-stranded RNA (dsRNA) (*figure 44).* Moreover, it also deaminates A in pre-mRNAs which alters individual codon thereby affecting splicing of the untranslated regions (Keegan, Gallo and O'Connell, 2001). The number of genes and isoforms of ADAR varies between species, but in case of mammals it encodes for three ADARs: ADAR1, ADAR2 and catalytically inactive ADAR3 (Chen *et al.*, 2000; Nishikura, 2010).

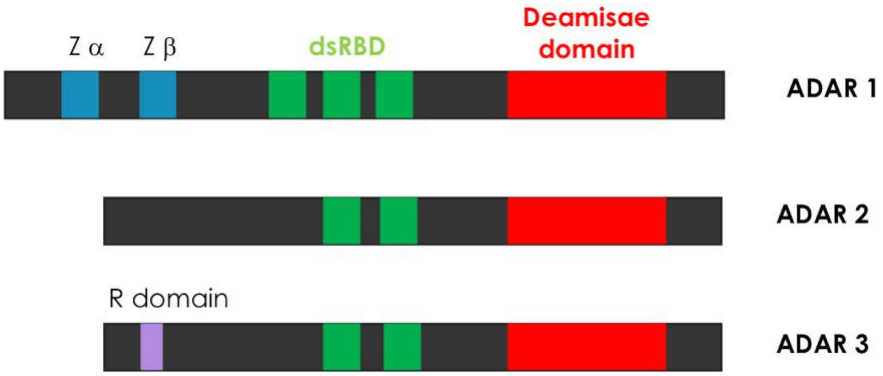


Figure.44: Domain map of active members of human ADAR family. Protein, domain and linker lengths to scale to demonstrate relative size (ADAR1: Z1 aa135–201, Z2 aa295–359, dsRBD1–3 aa504–569, 615–680, 727– 792, Deaminase aa837–1222) (ADAR2: dsRBD1 aa78–142).

The ADAR family shares a fairly common domain structure comprising of a variable number of amino terminal dsRNA binding domains (dsRBDs) and a carboxy-terminal catalytic domain (*Figure 44*) (Bass, 2002). The human ADAR1 has unique two Z-DNA binding domain which are known to pinpoint the left-hand helical variant of DNA in a

sequence independent manner and is known to confine the nucleic acid from adopting another conformation (Imre Berger *et al.*, 1998; Schwartz *et al.*, 1999). Very little is known about the interaction of the deaminase domain with the RNA. While the ADAR2 is active (in terms of catalytic activity) despite the presence of dsRBD *in vivo* (Herbert and Rich, 2001). With regards to the characteristic structural feature, ADAR1 and ADAR2 were reported to flip the target adenosine out of the DNA and into the enzyme's active site. This was demonstrated from the reports on a fluorescent adenine analog at the *GluR-B* editing site (R/G) being consistent with base flipping by ADAR2; this also bears resemblances to DNA methyltransferases which flip out the target base being found in the deaminase domain of ADARs (Hough and Bass, 1997; Stephens, Yi-Brunozzi and Beal, 2000; Yi-Brunozzi, Stephens and Beal, 2001).

ADAR 2

The Beal lab recently resolved the X-ray crystal structure of human ADAR2 deaminase domain (hADAR2d) bound to dsRNA (Matthews *et al.*, 2016).

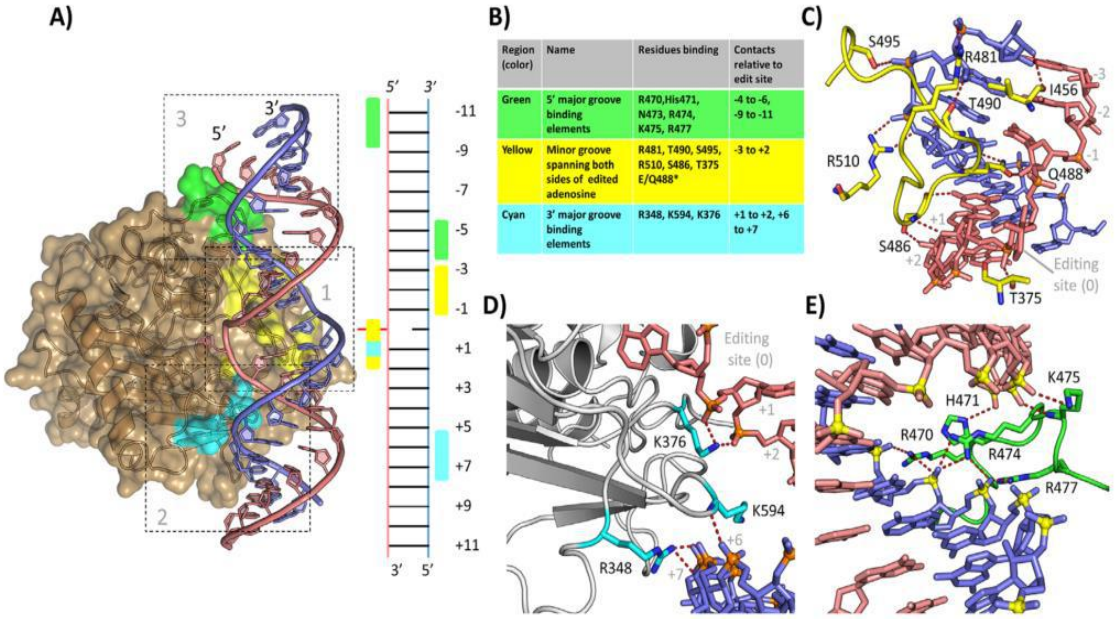


Figure.45: Interactions of hADAR2d with dsRNA. A: hADAR2d-RNA structure (5ED1, hADAR2d E488Q with Bdf2 derived 23-mer 8-azanebularane containing RNA) with three areas of interaction. (Edited strand in salmon, complementary stand in blue). Ladder diagram represents a secondary structure of the protein RNA contact. B: List of the protein residues and the

bound. C: Magnified view of region 1. D: Magnified view of region 2. E: Magnified view of region 3, crystal structure of hADAR2D (Matthews *et al.*, 2016; Thomas and Beal, 2017).

In the *figure 45 A and B*, three main regions of protein-RNA contact are seen which is spread across 20 base pairs of dsRNA on one side of double helix with interacting minor groove and adjacent major grooves. In the region 1 (*Figure 45A*), the Q488 residue (E488 in WT) accepts the hydrogen bond from the 2`-OH group from the edited strand (*Figure 45C)*; while the S486 H-bond together with the backbone amide nitrogen with the 2'-OH group at +1 and +2 positions and finally the T375 to the hydroxyl group of the edited base in the active site. This implies the importance of ADAR2 in determining preference for a 3` guanosine nearest neighbor (Matthews *et al.*, 2016). The interaction at the edited strand serves as a crevice for the active site and aids stabilization of the backbone involved for the edited adenosine to reach the catalytic core. The 486-491 residues at the minor groove aids the base flipping step for the deamination, also elucidating this loop role for ADARs preferential editing at adenosines (Matthews *et al.*, 2016).

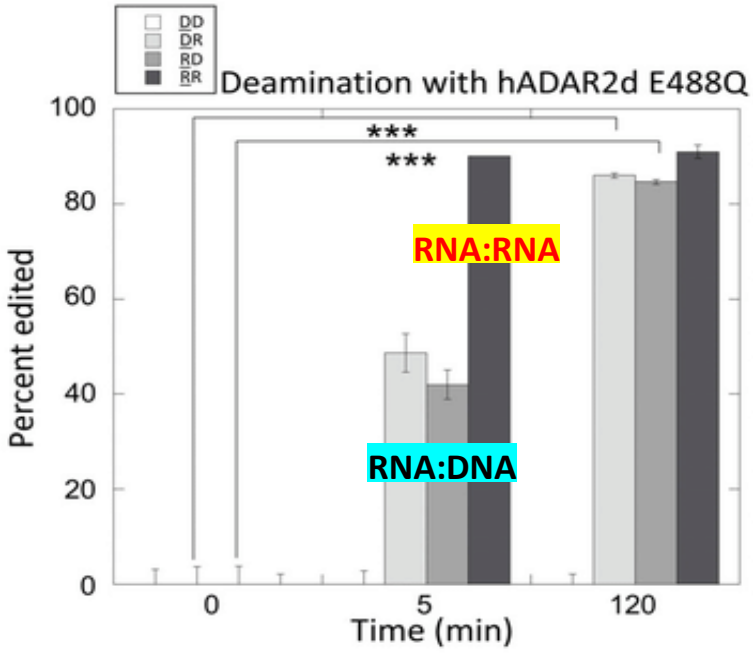


Figure.46: ADAR2 is able to edit the DNA strand in DNA:RNA hybrid (Modified from (Zheng, Lorenzo and Beal, 2017).

Moreover, in order to elucidate the role of the 2` hydroxyl contacts of the human ADAR2; it was seen that they are involved in editing but are not necessarily required for editing reaction hinting the possibility of reaction with DNA/RNA hybrids (Zheng, Lorenzo and Beal, 2017). Moreover, it was reported that overexpression of human ADAR1 lead to dA to dG mutation in the DNA (Tsuruoka *et al.*, 2013). Zheng and his collegues investigations on ADAR2, where they isolated the deaminase domain to target six different 2'-deoxyadenosines in the M13 bacteriophage ssDNA genome. DNA editing efficiencies varied depending on the sequence context of the editing site consistent with known sequence preferences for ADARs. These observations suggest the reaction within DNA/RNA hybrids may be a natural function of human ADARs. To test for reactivity in DNA/RNA hybrids and comparing them with similar RNA or DNA substrates, they created four new 24 bp duplexes with varying backbone structure of the component strands. Undoubtedly high efficiency of ADAR to deaminase all RNA substrates was observed, while the hybrid DNA/RNA were significantly deaminated (> 40%) after five minute reaction complete editing observed at 120 min (*Figure.46*) (Zheng, Lorenzo and Beal, 2017).This opens up new avenues for biological uses for ADAR2 especially in the field of genome editing.

5. Results and discussion (Part II)

5.1 ADAR as new tool for base editing

Recent advances in targeting genomic sequences changes has served great potential in the genome editing field for therapy. Using customized nucleases for targeted deletions, insertion and specific base sequence changes in a broad range of organisms and cell types. So far, Base editing depended on the cytidine deaminase to introduce changes; thus, it completely avoids the use of DBS and donor templates and offers a new prospective for gene editing. Targeting of C: G pairs has been demonstrated using these base editors, while targeting of A:T pairs is not currently possible.

Lately, the X-ray crystal structure of human ADAR2 deaminase domain (hADAR2d) bound to dsRNA had been resolved (Matthews *et al.*, 2016); Further studies from Beal's lab suggested that DNA editing efficiencies varied depending on the sequence context of the editing site consistent with known sequence preferences for ADARs. These observations imply the reaction within DNA/RNA hybrids may be a natural function of human ADARs.

This scrutiny of the structure led to the hypothesis that direct deamination of an adenosine in a DNA strand could be possible. I conjectured that engineered fusion proteins of CRISPR/dCas9 and deaminase domain of ADAR2 may induce a single dA to dG mutation, highlighting new possible role of ADAR2 as a DNA editor. Thus, I envisioned to design and build a CRISPR/dCAS9 fusion with the deaminase domain of ADAR2 (BE-ADAR2). This would be tested in a mammalian cellular model coupled with a fluorescent reporter assay.

5.2 Modelling the ADAR2 Base editor

The deaminase domain of ADAR2 was fused to the amino terminus of the catalytically inactive Cas9 using a 16 residue XTEN linker (*figure 47a*). To assess the ability of ADAR-BE in real time, I developed a reporter system, where a target site was specifically designed for ADAR2-Cas9/sgRNA complex to deaminate. This target site

design was based on Beal's lab work, where they assess the deamination ability of ADAR2 (Zheng, Lorenzo and Beal, 2017). The fluorescent reporter system was designed for quantification of real time editing in mammalian cells. This reporter system consists of dual fluorescent protein (mCherry-eGFP) which are separated by the target site for ADAR2 to deaminate, one a single A to G editing event occurs, there is a codon change resulting in the translation of the fluorescence activity (eGFP) suggesting gain of function (*figure 47b*).

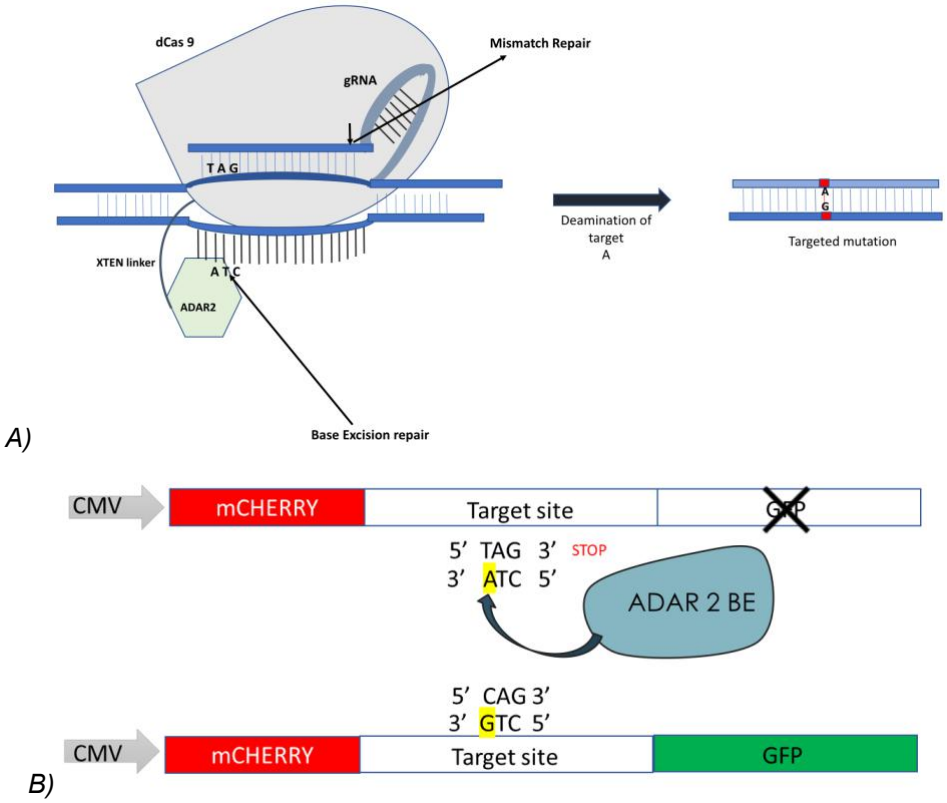


Figure.47 A) Construct design of the ADAR2 base editor. The guide RNA-programmed A→G conversion in vitro B)The mechanism of ADAR reaction in gene reporter: Schematic of the mCherry-eGFP) system in the context of a mammalian construct with a CMV promoter that drives expression of a bicistronic message encoding mCherry the target site with a stop codon and the wild-type eGFP

Next, in order to test the ability of the base editor *in vitro*, I designed a motif specific for the target site, and five different sgRNA with varying distance of C mispaired from the PAM sequence in order to test the better conditions for the editing (*figure.48)*.

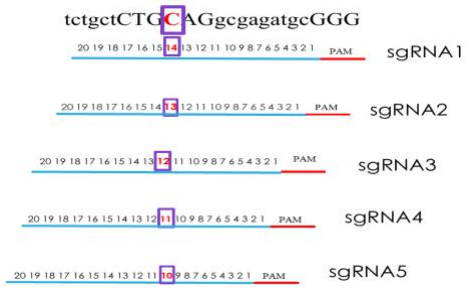


Figure.48: Sequence of the different sgRNA (Color in RED highlights the PAM; boxes in purple highlights bases to be edited)

5.3 Optimization and Validation of the ADAR2-Base editor

Protospacer adjacent motif (PAM) are very short nucleic acid sequences in *S. pyogenes*. In *Cas9*, there are two amino acids that specifically recognize the GG dinucleotide of a PAM and 'pulls' it away from the complementary CC of the other strand. This exposes the backbone phosphate of the upstream "N" nucleotide, which interacts with four more amino acids of Cas9. This slight destabilization of the DNA double helix is enough to promote formation of the necessary DNA-RNA hybrid if the sequence immediately upstream of the PAM matches that of the associated guide RNA. Thus, it is of prime importance that off-target effects of BE-ADAR2 should be avoided, that is, cutting at other, unintended sites in the genome and it is an important step in designing gRNAs. Without this function of the PAM, BE-ADAR2 would never be able to base-pair with genomic DNA, rendering ineffective to precisely target.

The ADAR2-BE constructs and controls (BE ADAR HAQ, catalytically inactive ADAR domain fused to a *dcas9*) were assembled and tested. These constructs were co-transfected into HEK-T cells with the mCherry-eGFP reporter and sgRNA to direct the editosomes to the target site. Transient transfections of HEK293T and HEK 293 cells were performed using PEI and were probed at various time points (24 hrs, 96 hrs and 140 hrs). Initial round of experiments showed that sgRNA1 functioned efficiently than the other sgRNAs (where the C is 12 bases downstream PAM) (*figure 49)*.

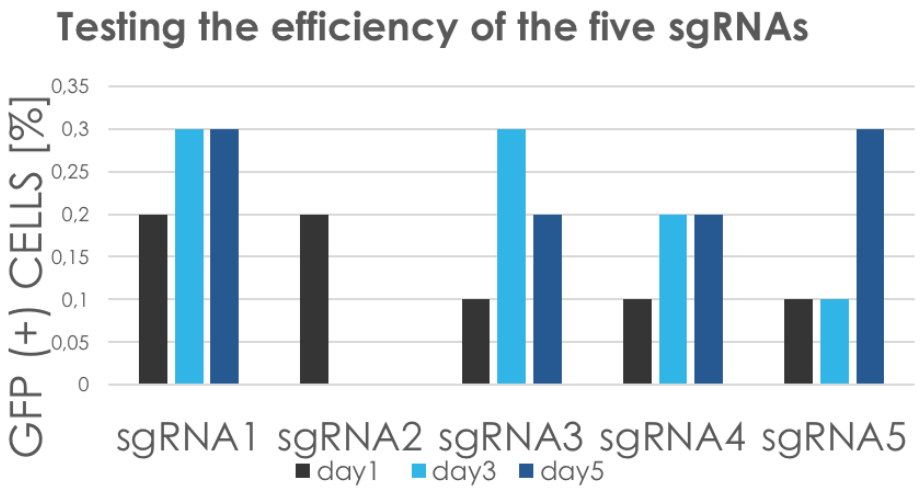


Figure.49: Effect of changing the sequence surrounding the target C on editing efficiency in vitro.

Overall, the editing efficiency of the system was quite low (~ 0.3%), prompting the need for optimization, where various combinations/concentrations of gene reporter, BE-ADAR2 and sgRNA were used. (*table 2*). Reversion of GFP was analysed at 24, 96, 140 hrs by FACS. Additional constructs were incorporated as negative controls, namely: BE w/o rat A1(cas9 alone) and BE ADAR (E308Q, catalytically inactive ADAR). FACS analysis showed abysmally low percentage of DNA editing when compared to the catalytically inactive mutants of ADAR2. This could have been due to the lack of time for the base editors to edit the base before translation. Thus, looking up at editing levels in the chromosomal DNA seemed like a logical step. Stable cells line of HEK-T and HEK incorporated with the mCherry- eGFP reporter were developed (*figure 50*).

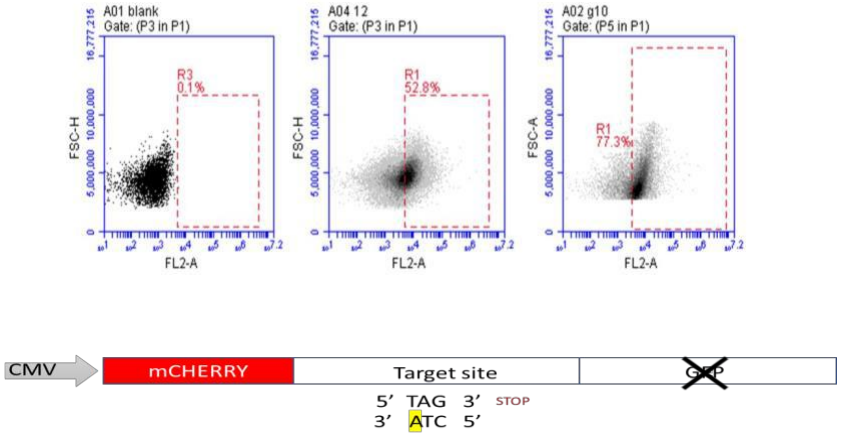
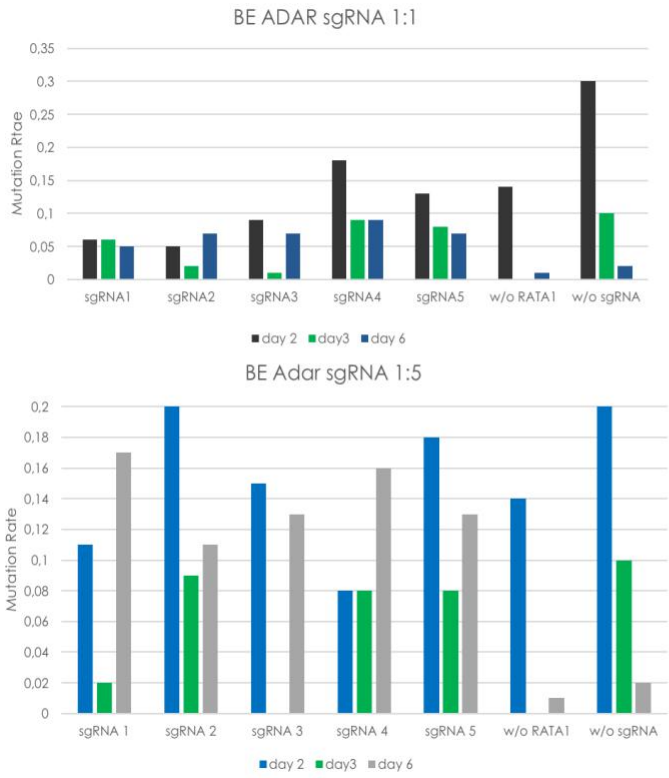
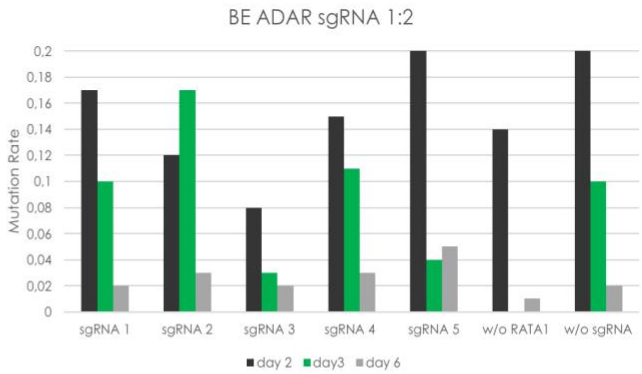


Figure 50: FACS analysis of the clone with a brighter red fluorescent population, and its difference from the negative control (blank).

The resulting mCherry- eGFP-reporter pools were co-transfected with the base editing constructs and a target site codon directed gRNA using PEI.





Vector	Amount of vector (ng) [1:1]	Amount of vector (ng) [1:2]	Amount of vector (ng) [1:5]
sgRNA 1	500	500	500
BE ADAR	500	250	100
sgRNA 2	500	500	500
BE ADAR	500	250	100
sgRNA 3	500	500	500
BE ADAR	500	250	100
sgRNA 4	500	500	500
BE ADAR	500	250	100
sgRNA 5	500	500	500
BE ADAR	500	250	100

Figure 51: Percentage of editing in the sample with the varying concentrations of BE ADAR and sgRNA on different day.

As seen above, initial round of experiments had shown signs of base editing, but these reports were thwarted as I was not able to replicate the same outcome (*figure 51)*. There could be several rationalisations, primarily, due to the location or the distance of the target site from a sgRNA anchoring motif (PAM), which prevents it to become a substrate for editing. Another possibility is the strength of the binding of DNA to ADAR2 and the possible presence of off-target sites. While the ability of the ADAR2 fusion to the dcas9 needs to be analysed, as fusing them to the carboxy terminus of dcas9 has yet not been tested. Another aspect could be the usage of the linker and its length as longer linker might aid ADAR deaminase to have a better access to substrate.

Furthermore, repair by Thymine-DNA glycosylase (TDG), which removes thymine moieties from G/T mismatches by hydrolyzing the carbon-nitrogen bond between the sugar-phosphate backbone of DNA and the mis-paired thymine, could be one of the several reasons for the failure of the system. It might also be worth looking up at testing various target sites for efficient deamination by ADAR2. Currently, additional constructs are being designed with keeping these aspects in mind.

As an alternative for the wild type ADAR2, I set to develop mutants of ADAR2 which may have a greater affinity to act on the DNA.

5.4 Modifying the RNA editor to target DNA

Recent reports have suggested that a double helical structure is required for ADAR substrates. Beal's lab reported the X-ray crystal structures of the human ADAR2 deaminase domain bound to substrate RNAs. Fascinatingly, these structures also identified five direct contacts to 2′-hydroxyls in the minor groove near the editing site with only four of these common to the two different RNA sequences crystallized (*figure 52)*(Matthews *et al.*, 2016).This work led me hypothesize whether the 2′-hydroxyl contacts which are required for the ADAR reaction could altered or modified to let ADAR2 react directly with a double stranded DNA.

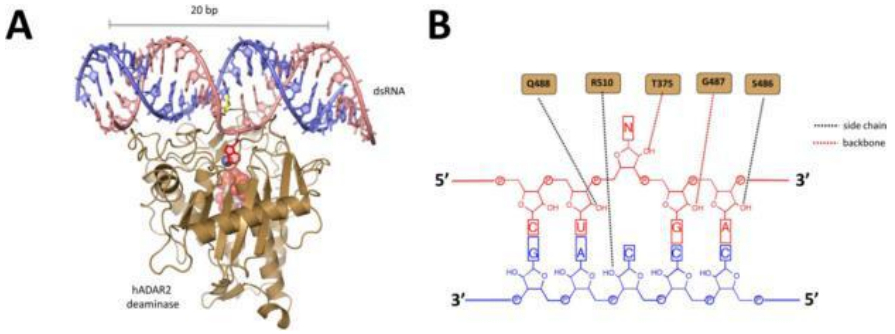


Figure 52: Interactions between hADAR2d and 2′-hydroxyl groups (Matthews et al., 2016; Zheng, Lorenzo and Beal, 2017).

The E488, T375 and S486 residues are involved in the binding with the edited base, while the residue T490 promotes the base flipping step of the deamination reaction. Keeping these in mind, I developed mutants of human ADAR2 which may possibly have an affinity to bind the DNA. The wild type ADAR2 was mutagenized via PCR with

degenerate primers aimed at the four sites involved with 2´-hydroxyl contacts (*Supplementary data table 1*), and were later cloned back into the bacterial vector; transformed in to *DH-5α E. coli* strains, where a plasmid library was obtained. Moreover it is known that ADAR2 requires inositol hexakisphosphate (IP6) for editing, whereas the inositol phosphorylation system is not known to be present in bacteria, and is only found in all eukaryotes (Macbeth *et al.*, 2005).Thus, phytic acid was added as a substrate initially to obtain active ADAR2 mutants which were selected through two bacterial screens to assay for DNA mutator phenotypes: Rifampicin and Nalidixic acid assay (Petersen-Mahrt, Harris and Neuberger, 2002).

These assays are based on the onset of resistance to Rifampicin and Nalidixic acid, where the mutants are transformed in *E. coli* strain *KL16*, sensitive to Rifampicin and Nalidixic acid. Acquisition of mutations in the RNA polymerase (*rpoB)* or gyrase *(gyrA)* genes can reverse the phenotypes. Rifampicin inhibits bacterial DNA-dependent RNA polymerase and binds to RNA polymerase at a site adjacent to the RNA polymerase active center and blocks RNA synthesis. Resistance to rifampicin is due to mutations occurring at rifampicin binding site on RNA polymerase (on the *rpoB* region, encoding RNA polymerase β subunit). For Nalidixic acid, *gyrase A* relieves strain while double-stranded DNA is being unwound by helicase resulting in the negative supercoiling of the DNA. Mutator phenotypes were obtained using these the bacterial assays, where colonies resistant to rifampicin as well as Nalidixic acid were individually picked and plasmid DNA containing ADAR2 mutants were obtained (*figure53)*.

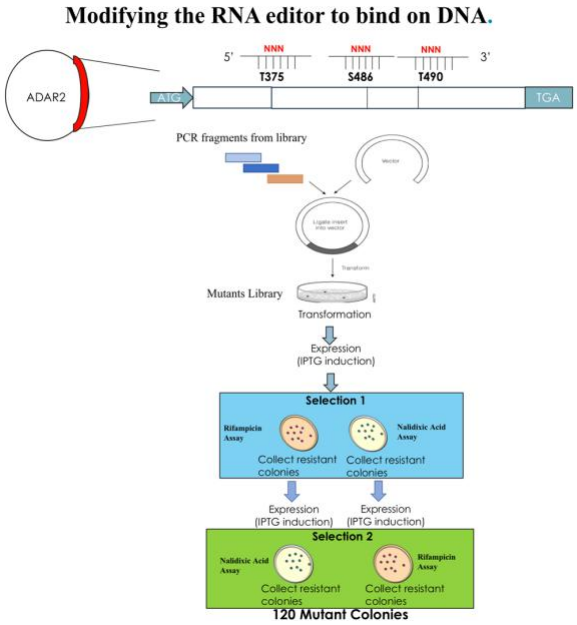


Figure.53: Schematic overview of the mutator screening assay to obtain ADAR2 mutants.

The notion for a dual selection of mutator assay was to eradicate weak mutator phenotypes. Overall, 120 ADAR2 mutants were obtained, where 60 of them were selected from colonies positive for rifampicin followed by nalidixic acid (RN); while the other 60 were Nalidixic acid followed by rifampicin (NR). These mutants were cloned into the bacterial vector pTRC99A, under control of a *trp/lac* (tac) hybrid promoter. The obtained mutants were tested for their ability to mutate the DNA. Here, the mutants were transformed into *E.coli* strain KL16, and were selected based on their effect on the frequency of mutation to rifampicin resistance and nalidixic acid. Human Activation Induced Deaminase (AID), was used as a positive control for the induction of a mutator phenotype in *E.coli* (Petersen-Mahrt, Harris and Neuberger, 2002).

With the initial screening of 120 mutants(in the absence of phytic acid), two mutants were selected for their ability to induce resistant colonies. These mutants NR 3.18 and NR 3.20, exhibited considerably higher mutation rate when compared to AID *(figure54)*. Subsequently, these mutants were retested in order to ensure reproducibility and were also tested along with catalytically inactive mutants of ADAR2 NR3.18 and NR3.20 (*figure 55)*.

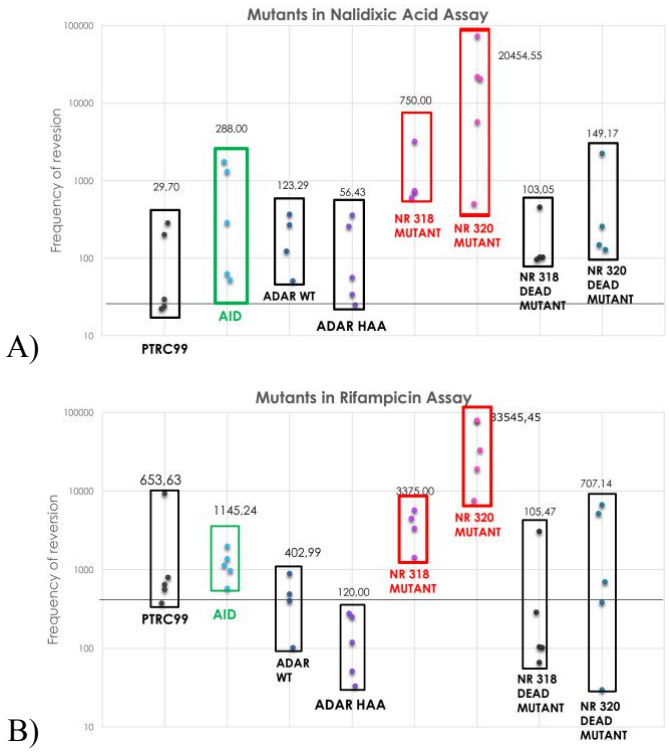


Figure .54: A) frequencies of Nal R mutants generated after overnight culture with IPTG of E. coli KL16 carrying either the ADAR mutant expression plasmid and the vector control. B) frequencies of Nal R mutants generated Each point represents the mutation frequency of an independent overnight culture.

From the fluctuation analysis, the mutants generated seemed quite potent. The ADAR2 NR3.18 and NR3.20 mutants displayed exceptional mutator phenotype capabilities (Figure53 A and B). Their mutation frequency was 100-fold higher than AID when tested in both rifampicin and nalidixic acid. While the catalytically inactive NR3.18 and NR3.20 displayed mutation frequencies similar to that of the empty vector. This indicates that ADAR2 induces DNA deamination on *E.coli*. These mutants were sent for Sanger sequencing for the characterization of mutations present on the plasmid DNAs.

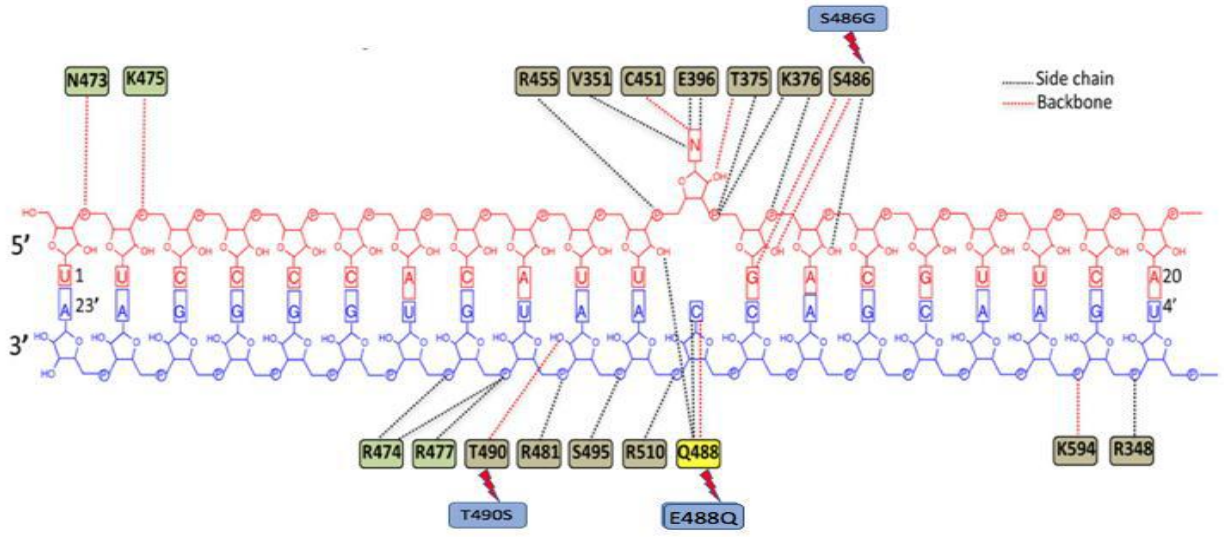


Figure.55: Summary of the contacts between hADAR2d and all the mutated amino acids (S486G, E488Q and T490S) indicated with red symbols and the Bdf2-C RNA duplex.

The mutants had been exactly modified at S486G, E488E and T490S *(Figure 55)*. These are the sites involved in base-flipping enzymes to stabilize the altered nucleic acid conformation by intercalation of an amino acid side chain into the space which is vacated by the flipped-out base which E488. While the S486, is known to accept an H-bond from the 2-amino group of the G on the 3' side of the edited nucleotide. Guanine is the only common nucleobase that presents an H-bond donor in the RNA minor groove suggesting that other nucleotides in this position would reduce editing efficiency, when mutated to S486G. These are all the essential residues which may alter base flipping.

Subsequently, the mutants were characterized to see if they indeed caused A:G mutation mediated by ADAR NR318 and NR 320. If this was true, we should notice a dA to dT transition in mutations pattern of rifampicin and nalidixic acid resistant colonies. Thus, the *rpoB* and the *gyrA* mutation spectrum were analysed in 12 independent colonies from each approach (*Figure 56 and 57)*.

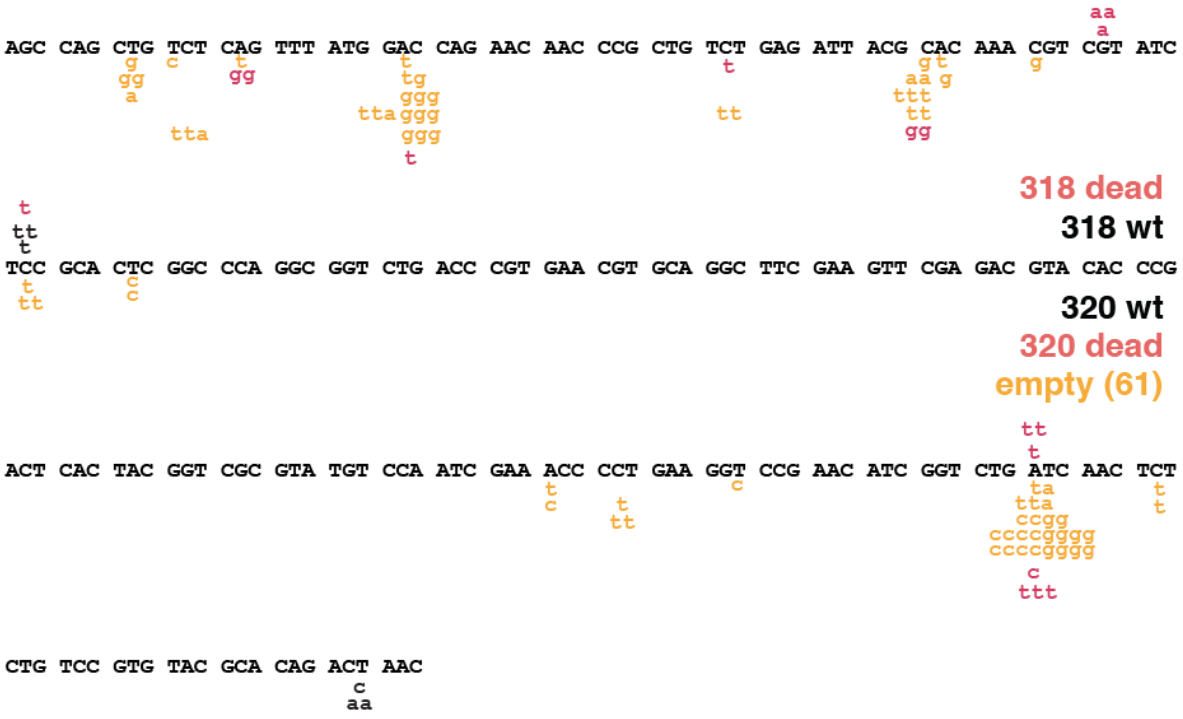


Figure.56: Spectrum of rpoB mutations in rifampicin resistant colonies from empty vector (in yellow) (ptrc99), mutants of ADAR2 NR3.18 and NR3.20 (in black) and the catalytically inactive mutants of the same (in red).

The spectrum of the *rpoB* gene from the rifampicin resistant colonies for NR318 and NR 320 displayed very few mutations on a part of the *rpoB* where it is generally known to carry maximum mutation conferring rifampicin resistance. Interestingly in the case of the catalytically inactive mutants of NR.318 and NR3.20 there were some dA to dG shift observed (*figure 56)*. This, together with lack of mutations in the colonies from the catalytically active mutants, suggests that mutation from the NR3.18 and NR3.20 maybe occurring elsewhere in *rpoB* gene. Given the extent of mutation caused by the mutants (seen in the fluctuation analysis); there is definitely a greater need to analyze the dA-dG transition, across the entire *rpoB* genome in order as there is only a limited number of base substitutions that could yield the selected phenotypes.

However, while analyzing the spectrum of *gyrA* mutations in the colonies resistant to Nalidixic acid; ADAR mutants showed a significant dA to dG and dG to dT transitions, suggesting deamination of dA. Considering the difference in the mutational distribution

between the active mutant and the catalytically dead mutants; it seems that the ADAR mutants has the ability to induce mutation in *E.coli*.

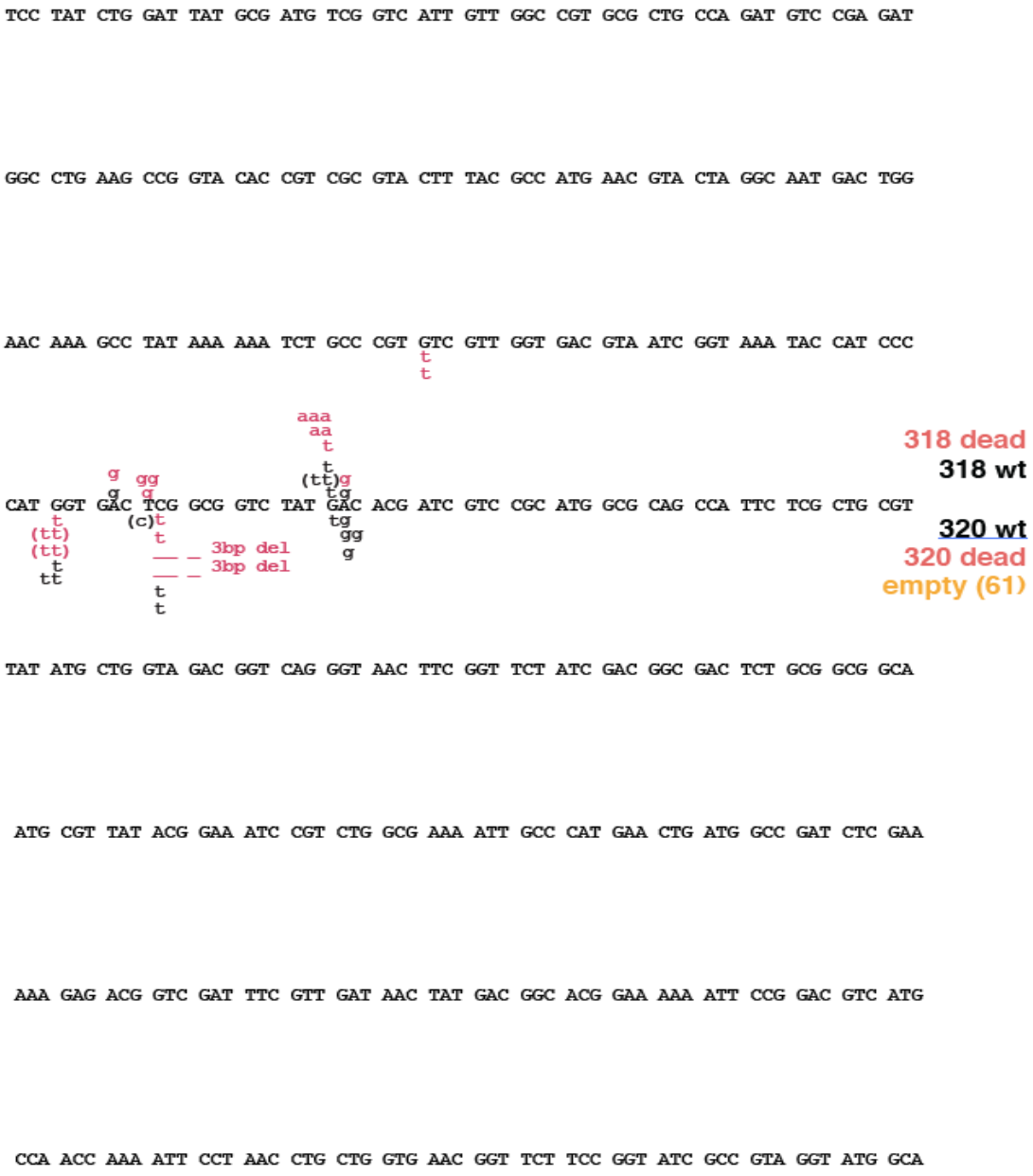


Figure.57: Spectrum of gyrA mutations in nalidixic acid resistant colonies from empty vector(in yellow) (ptrc99), mutants of ADAR2 NR3.18 and NR3.20 (in black) and the catalytically inactive mutants of the same (in red).

These are preliminary indications of NR3.18 and NR3.20 triggering deamination of dA residues in the DNA. Further analysis is currently being carried out in order to fully assess its potential as a DNA editor *in vitro.*

This functionality of the ADAR2 mutants in bacterial system put these deaminases in a whole new light. The results open several possibilities of DNA-targeted activity of the mutants and its potential biological uses.

The mutants are quite apart from the WT ADAR2 as they tend to display DNA mutator activity in *E.coli*. Moreover, WT ADAR2 requires inositol hexakisphosphate (IP6) for RNA editing, whereas the inositol phosphorylation system is not present in bacteria and is only found in eukaryotes (Macbeth *et al.*, 2005). The IP6 is known to be buried within the enzyme core contributing to the protein fold. Nonetheless, the NR3.18 and NR3.20 mutants do not seem to require IP6 to act as DNA editors. Additionally, the mutants might serve as a better alternative for human ADAR2 for its use as base editor targeting A:T pairs with potential usage in biotechnological applications such as gene therapy, antiviral and cancer therapy. Currently, I am working on constructing these mutant deaminases where I am fusing them with Cas9 and TALEN as a novel genome editing tool.

Conclusively, ADAR deaminase domain bearing mutations in the enzyme's base flipping loop may edit DNA. Fusion of ADAR catalytic domains with nucleic acid binding domains, particularly hybrid binding domains, and activation with additional specific mutations are likely to enhance reactivity with DNA.

During the writing of the thesis, a study has been published reporting the development of adenine base editors (ABEs), evolved from the bacterial tRNA adenosine deaminase, that mediate conversion of A•T to G•C in genomic DNA when fused to a catalytically impaired CRISPR-Cas9 (Gaudelli *et al.*, 2017).

6. Materials and methods

Bacterial Strains used

Strain	Genotype	Uses	Reference
DH-5α	F– Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1	Host strain for general cloning	Invitrogen
STBL-3	F– mcrB mrr hsdS20 (rB–, mB–) recA13 supE44 ara-14 galK2 lacY1 proA2 rpsL20 (StrR) xyl-5 λ– leu mtl-1	Host strain for lentiviral and retroviral cloning	Invitrogen
KL-16	Hfr (PO) relA1 spoT1 thi-1 ung+	Host strain for rifampicin and nalidixic acid assay	(Petersen-Mahrt, Harris and Neuberger, 2002)

Media

Bacterial media LB (Luria-bertani), SOB (LB medium + 10 % MgSO4), SOC (LB medium + 10% MgSO4 + Glucose 1M) were prepared using standard recipes (Sambrook and Russell, 2001). Media were sterilized by autoclaving at 120^{o}C for 20 minutes. Antibiotics were used at the following working concentrations.

Ampicillin: 100μg/ml

Kanamycin: 50μg/ml

Rifampicin: 40μg/ml

Nalidixic acid: 40μg/ml

Preparation of chemically competent bacteria

Bacteria were inoculated overnight in 250ml in SOB medium at 37°C and grown until it reached the OD_{600} of 0.6 (Biophotometer Eppendorf spectrophotometer). This was followed by 10 minute incubation in ice and centrifuged at 2500g for 10 minutes at 4°C. The pellets were resuspended in Transformation Buffer (pipes 10mM, $MnCl_2$ 55mM, $CaCl_2$ 15mM and KCl 250mM), Two washes were performed before incubating the bacteria for 15 minutes on ice. DMSO was added to reach 7% of final concentration, followed by a further incubation on ice for 5 minutes. The bacteria then were aliquoted, snap frozen in liquid nitrogen, and stored at -80°C.

Transformation of chemically competent bacteria

Bacterial transformations: Competent cells were thawed on ice and 100 µl of competent cells were mixed with 100 ng of plasmid DNA, or 7.5 µl of ligation product, for 30 minutes to allow adsorption of the DNA to the cells. The mixture was subjected to a heat shock at 42°C for 45 seconds and immediately cooled on ice for 2 minutes. After addition of 1.0 ml of SOC media the cells were left to recover and to express the antibiotic resistance at 37°C for one hour. Cells were pelleted gently and re-suspended in ~100-200 µl SOC before plating them on plates containing the appropriate antibiotics and incubated overnight at 37°C.

Rifampicin assay/Nalidixic Acid assay

The protocol for these assays was adapted from the work of (Petersen-Mahrt, Harris and Neuberger, 2002). The plasmids encoding for AID, ADAR2 and its mutants were expressed in *E.Coli KL16*, under the control of a trp/lac (tac). Individual colonies were grown overnight in 3 ml of LB, supplemented with 1µL IPTG(1M) and 3 µL of ampicillin, to induce the expression of the proteins. 400 µL of inoculate were plated on a medium containing rifampicin [50 µg/ml] or nalidixic acid [40 µg/ml] to select resistant

bacteria. Mutation frequencies were measured by determining the median number of colony forming cells that survived selection per 10^7 viable cells.

Sequence analysis: The nature of the Rif and Nal mutants was determined by directly amplifying and sequencing the relevant section of *rpoB* (627-base pair PCR product amplified using 50 - ttggcgaaatggcggaaaacc-3` and 5`-caccgacggataccacctgctg-3`) or *gyrA* (521-bp PCR product amplified using oligonucleotides 5`-GCGCGGCTGTGTTATAATTT-3` and 5` -TTCCGTGCCGTCATAGTTATC-3`).

Cell lines

Human fibrosarcoma HT1080 cells bearing the alphoid tetO human artificial chromosome and expressing the tetracycline repressor TetR fused to the transcriptional silencer tTS (a gift from A.Musio,CNR PISA). They were maintained at 37° C with 5% CO2 atmosphere in DMEM 10% FBS, 100 IU ml^{-1} penicillin, 100 µg ml^{-1} streptomycin (GE life sciences), 1 µg ml^{-1} doxycycline (Sigma), 0.5 mg ml^{-1} G418 (geneticin, Life technologies) and 4 µg ml^{-1} blasticidin S (Invivogen) at 37°C in 5% CO2.
HEK 293 and HEK 239T were maintained in medium that have been prepared starting from commercial media Dulbecco's Modified Eagle's Medium (Euroclone). DMEM medium was supplemented with 10% fetal bovine serum (FBS), 5 U/ml penicillin, and streptomycin sulphate, 1% L-Glutamine (200mM). Cells were grown in incubators at 37°C with 5% CO2.

Transfection of eukaryotic cells

Transient transfections for HT1080 HAC were performed using X-tremeGene HP DNA transfection reagent (Roche Diagnostics, Basel, Switzerland) and Liopfectamine 3000 (Invitrogen) according to manufacturer's instructions. Transiently transfected cells were selected through puromycin selection (2.5µl/ml).
Transient transfections in HEK 293T cells were performed using Polyethylenimine (PEI) (Sigma Aldrich), following manufacturers protocol.

For electroporation of HEK 293T $5x10^6$ cells were washed, re suspended in 800µl of PDB and electroporated with 7.5µg of vector in a 4mm electroporation cuvette (Biorad). These were electroporated using the Gene Pulser X-cell Total System (Biorad). The following parameters were set :Voltage = 250V, Capacity = 950µF; Resistance = ∞

Generation of Micronuclei

HT1080 cells carrying the HAC were cultured in doxycycline-free medium and treated with 100 ng ml−1 CytochalasinB (Sigma) for 6 h to synchronize cells in mitosis. The absence of doxycycline enables TetR binding to TetO, which induces inactivation of the HAC centromere by tTS and the subsequent formation of a micronucleus after release from mitotic arrest. Cells released from mitotic arrest were stained with Giemsa for micronuclei frequency analysis.

Micronuclei frequency, Microscopy and Image analysis

The transfected HT-1080 HAC cells were treated with and without CytochalasinB (as previously described) for 6 hours. The cells were later fixed with an acetic acid- methanol (1:3) solution and were added on microscope slides pre-treated with a mix of Hydrochloric acid and ethanol. The slide was rinsed briefly with small amounts of tap water, after which one small drop of mounting medium (70% glycerol in PBS) was added to the slide and covered with a coverslip. Micronuclei frequency was obtained by counting the number of micronuclei in bi-nucleated cells. The samples were imaged using a Nikon Eclipse 50i epifluorescence microscope with a Nikon CFI Plan Fluor 100× objective (Nikon Instruments Europe B.V). The microscope was equipped with a DS-Fi1 digital microscope camera head (Nikon Instruments Europe B.V)

Scoring Micronuclei

At least 200 cells were counted for each experiment and experiments were repeated three times in double blind conditions. Criteria for scoring micronuclei included the following: (1) presence of bi-nucleated cells; (2)

the diameter of the MN should be less than one-third of the main nucleus; (3) MN should be separated from or marginally overlap with main nucleus as long as there is clear identification of the nuclear boundary; and (4) MN should have similar staining as the main nucleus.

Standard statistical analysis such as ANOVA and tukey's test was performed to assess the micronuclei frequency.

Immunofluorescence Microscopy

Cells were grown in a 6-well (euroclone) post transfection on coverslip glass (Mensel Glaser) which was previously incubated with polylysine (sigma). The medium was aspirated and cells were fixed with 100% cold methanol for 20 minutes. The slide was then washed with PBS followed by blocking using PBS/BSA 1% for 15 minutes and 2x washings with 0.5% PBS/BSA. Primary mouse monoclonal antibody - (anti γ-H2A.X, Ser139, Cell Signalling technology) 1:1000 was added and incubated for 30 minutes at room temperature. Secondary antibody conjugated with Alexa Fluor 594 (abcam, 1:1000) was added and incubated for 30minutes at room temperature. Samples were washed thrice, incubated in DAPI for nuclei staining (5 μM for 5 minutes) and washed with PBS. Microscope coverslips were then were then mounted on the slides with a drop of fluorescence mounting medium (Dako). Images were acquired using a Nikon Eclipse 50i epifluorescence microscope with a Nikon CFI Plan Fluor 100× oil objective lens (Nikon Instruments Europe B.V). The microscope was equipped with a DS-Fi1 digital microscope camera head (Nikon Instruments Europe B.V)

Metaphase chromosome spread preparation

Transfected HT-1080 HAC cells were incubated for 2 hours 0.2 $\mu g\,ml^{-1}$ colcemid (Gibco). The cells were collected by trypsinization, resuspended in 0.075 M KCl at 37 °C for 30 min, and fixed overnight in methanol/acetic acid (3:1) at 4 °C. The cells were dropped onto wet glass slides (treated with ethanol: acetic acid, 1:3) and allowed to air dry. Chromosomes were banded by staining with 2.5% phosphate-buffered Giemsa (Bio-optica).

Reporter Assays: Semi-confluent stable clones with mCherry – eGFP reporter HEK-T cells in a 6-well plate format were transfected with 500 ng gRNA, and 100 ng of each base-editor [15 min, RT with 2.4 µl of PEI polyethylenimine (PEI)(Sigma) and 100 µl of serum-free DMEM (Euroclone)].
The cells were analysed at 48, 96, 144 hours after transfection by flow cytometry analysis FACS (Fluorescence-activated cell sorting).

Recombinant DNA techniques

Plasmid DNA preparation

Plasmid DNA were prepared from bacterial cultures using the eurogold plasmid miniprep kit (Euroclone) for minipreps, while for Maxi-preps, high-pure plasmid maxiprep kit (Invitrogen) and was performed according to the manufacturer's instructions.

Oligonucleotides

Oligonucleotides, including PCR and sequencing primers are listed in the supplementary table 1 (appendix), and were synthesized by Macrogen.

Restriction endonucleases

New England Biolabs endonuclease were used for restriction digestion, all the enzymes were used as per manufacturer's protocol.

Vectors and Plasmid Construction

CIN experiments:

All the plasmids expressing APOBECs were cloned in pAID puromycin expressing expression vectors where the ß-actin promoter drives the expression of AID/APOBECs in mammalian cells alongside an EGFP reporter gene, linked to the AID/APOBECs transcript through an internal ribosome entry site (IRES) (Arakawa *et al*., 2001 and Saraconi *et al*.,2014).
The human APOBEC3A and APOBEC3B coding sequences obtained by amplifying(primer 6-7 and 4-5 respectively) from plasmids kindly provided by Rueben

Harris and were cloning into pAID-puromycin expressing plasmid digested either with *NheI/EcorI or NheI/BamHI*

Similarly, human APOBEC2, APOBEC3C, APOBEC3F, APOBEC3G and APOBEC3H are obtained by amplifying them with primer pairs 14/15,16/17,20/21,18/19,8/9 and 12/13 respectively and cloned into pAID-puromycin expressing plasmid digested with *NheI/BglII*.

For the plasmids constructed above where fragments were cloned into was performed using the Rapid DNA Dephos & Ligation kit (Roche).

BE-ADAR: pCMV-BE1 plasmid was a gift from David Liu (Addgene plasmid # 73019), was obtained replacing the APOBEC 1 coding sequence with the DNA fragment encoding the deaminase domain if ADAR2 fragment was PCR amplified using the primers 5` ttttgagctcatggccttgcacttggatca 3` and 5` ttttcccgggacagggcgtgagtgagaact 3` and cloned into BE 1 using the restriction site *SacI* and *XmaI*. Additional constructs of BE1 were made where the APOBEC 1 coding sequenced was excised and replace with a catalytically inactive ADAR2. The mCherry-eGFP system was derived from pEGFP-N1-mCherry-apoB-egfp described in Severi., 2015, where the apoB was excised and replaced with a synthesized oligonucleotide sequence specifically designed as a target ADAR2 (Supplemental table 1, 44/45) using *HindIII* and *EcoRI* restriction sites. This was later ligated using the *T4 DNA* ligase (Roche).

Guide (g)RNAs targeting mCherry – eGFP reporter system sequence (Supplementary Table.1, 53-61) were cloned into a px330 in which the cas9 was removed by KpnI and NotI, through Addgene (Plasmid #42230) using the accompanying gRNA cloning protocol.

ADAR2 mutant library preparation**:** The first and second generation ADAR2 mutant libraries were generated by PCR using *Taq* polymerase on 100 ng of template DNA(ADAR2) using degenerate primers (supplementary table 1) in the following mix 250 mM dNTPs and 10 mM MgCl2 in Taq buffer and polymerase. PCR product was digested using *Nco-I* and *Xba-I* and was cloned into pTRC99A vector. These was transformed in *E.coli* DH5-α competent cells on LB agar plates containing 100 μg/ml of

Ampicillin, overnight at 37C. The bacterial colonies were collected by scraping the plates and the DNA was extracted using the euroclone gold DNA-miniprep kit**.**

6. Bibliography

'A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine' (1987) *Cell*. Cell Press, 50(6), pp. 831–840. doi: 10.1016/0092-8674(87)90510-1.

Abend, M., Frombeck, S. and Van Beuningen, D. (1999) 'Indications for an active process underlying spontaneous and radiation-induced micronucleation in L929 cells', *International Journal of Radiation Biology*. Taylor & Francis, 75(12), pp. 1567–1578. doi: 10.1080/095530099139179.

Aït Yahya-Graison, E. *et al.* (2007) 'Classification of human chromosome 21 gene-expression variations in Down syndrome: impact on disease phenotypes.', *American journal of human genetics*. Elsevier, 81(3), pp. 475–91. doi: 10.1086/520000.

Alexandrov, L. B. *et al.* (2013) 'Signatures of mutational processes in human cancer', *Nature*, 500(7463), pp. 415–421. doi: 10.1038/nature12477.

Arakawa, H., HauschiLd, J. and Buerstedde, J. M. (2002) 'Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion', *Science*, 295(5558), pp. 1301–1306. doi: 10.1126/science.1067308.

Bakhoum, S. F. and Compton, D. a (2012) 'Science in medicine Chromosomal instability and cancer : a complex relationship with therapeutic potential', *The Journal of Clinical Investigation*. doi: 10.1172/JCI59954.1138.

Barber, T. D. *et al.* (2008) 'Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(9), pp. 3443–8. doi: 10.1073/pnas.0712384105.

Barrangou, R. *et al.* (2007) 'CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes', *Science*, 315(5819), pp. 1709–1712. doi: 10.1126/science.1138140.

Bartek, J. and Lukas, J. (2007) 'DNA damage checkpoints: from initiation to recovery or adaptation', *Current Opinion in Cell Biology*, 19(2), pp. 238–245. doi: 10.1016/j.ceb.2007.02.009.

Bass, B. L. (2002) 'RNA Editing by Adenosine Deaminases That Act on RNA', *Annual Review of Biochemistry*, 71(1), pp. 817–846. doi: 10.1146/annurev.biochem.71.110601.135501.

Bayani, J. *et al.* (2007) 'Genomic mechanisms and measurement of structural and numerical instability in cancer cells', *Seminars in Cancer Biology*, 17(1), pp. 5–18. doi: 10.1016/j.semcancer.2006.10.006.

Bayani, J. and Squire, J. A. (2001) 'Advances in the detection of chromosomal aberrations using spectral karyotyping.', *Clinical genetics*, 59(2), pp. 65–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11260203 (Accessed: 11 October 2017).

Bedard, L. L. and Massey, T. E. (2006) 'Aflatoxin B1-induced DNA damage and its repair', *Cancer Letters*, 241(2), pp. 174–183. doi: 10.1016/j.canlet.2005.11.018.

Beheshti, B. *et al.* (no date) 'Evidence of chromosomal instability in prostate cancer determined by spectral karyotyping (SKY) and interphase fish analysis.', *Neoplasia (New York, N.Y.)*, 3(1), pp. 62–9. doi: 10.1038/sj/neo/7900125.

Beroukhim, R. *et al.* (2010) 'The landscape of somatic copy-number alteration across human cancers', *Nature*. Nature Publishing Group, 463(7283), pp. 899–905. doi: 10.1038/nature08822.

Bhatia, A. and Kumar, Y. (2013) 'Cancer cell micronucleus: an update on clinical and diagnostic applications', *APMIS*, 121(7), pp. 569–581. doi: 10.1111/apm.12033.

Bibikova, M. *et al.* (2002) 'Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases.', *Genetics*, 161(3), pp. 1169–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12136019 (Accessed: 14 October 2017).

Bibikova, M. *et al.* (2003) 'Enhancing Gene Targeting with Designed Zinc Finger Nucleases', *Science*, 300(5620), pp. 764–764. doi: 10.1126/science.1079512.

Boboila, C. *et al.* (2010) 'Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase 4.', *The Journal of experimental medicine*. The Rockefeller University Press, 207(2), pp. 417–27. doi: 10.1084/jem.20092449.

Bonassi, S. *et al.* (2011) 'Micronuclei frequency in peripheral blood lymphocytes and cancer risk: evidence from human studies', *Mutagenesis*. Oxford University Press, 26(1), pp. 93–100. doi: 10.1093/mutage/geq075.

Boveri, T. (2007) 'Concerning the Origin of Malignant Tumours by Theodor Boveri. Translated and annotated by Henry Harris', *Journal of Cell Science*, 121(Supplement 1), p. 1 LP-84. Available at: http://jcs.biologists.org/content/121/Supplement_1/1.abstract.

Branzei, D. and Foiani, M. (2005) 'The DNA damage response during DNA replication',

Current Opinion in Cell Biology, 17(6), pp. 568–575. doi: 10.1016/j.ceb.2005.09.003.

BRANZEI, D. and FOIANI, M. (2007) 'Interplay of replication checkpoints and repair proteins at stalled replication forks', *DNA Repair*, 6(7), pp. 994–1003. doi: 10.1016/j.dnarep.2007.02.018.

Briggs, A. W. *et al.* (2012) 'Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers', *Nucleic Acids Research*, 40(15), pp. e117–e117. doi: 10.1093/nar/gks624.

Bristow, R. G. and Hill, R. P. (2008) 'Hypoxia and metabolism: Hypoxia, DNA repair and genetic instability', *Nature Reviews Cancer*, 8(3), pp. 180–192. doi: 10.1038/nrc2344.

Brouns, S. J. J. *et al.* (2008) 'Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes', *Science*, 321(5891), pp. 960–964. doi: 10.1126/science.1159689.

Bunting, S. F. and Nussenzweig, A. (2013) 'End-joining, translocations and cancer', *Nature Reviews Cancer*, 13(7), pp. 443–454. doi: 10.1038/nrc3537.

Burns, M. B. *et al.* (2013) 'APOBEC3B is an enzymatic source of mutation in breast cancer', *Nature*, 494(7437), pp. 366–370. doi: 10.1038/nature11881.

Burns, M. B., Temiz, N. A. and Harris, R. S. (2013) 'Evidence for APOBEC3B mutagenesis in multiple human cancers.', *Nature genetics*. NIH Public Access, 45(9), pp. 977–83. doi: 10.1038/ng.2701.

Burrell, R. A. *et al.* (2013) 'Replication stress links structural and numerical cancer chromosomal instability', *Nature*. Nature Research, 494(7438), pp. 492–496. doi: 10.1038/nature11935.

Cadet, J., Sage, E. and Douki, T. (2005) 'Ultraviolet radiation-mediated damage to cellular DNA', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 571(1–2), pp. 3–17. doi: 10.1016/j.mrfmmm.2004.09.012.

Cahill, D. P. *et al.* (1998) 'Mutations of mitotic checkpoint genes in human cancers', *Nature*, 392(6673), pp. 300–303. doi: 10.1038/32688.

Camps, J. *et al.* (2005) 'Comprehensive measurement of chromosomal instability in cancer cells: combination of fluorescence in situ hybridization and cytokinesis-block micronucleus assay', *The FASEB Journal*, 19(7), pp. 828–30. doi: 10.1096/fj.04-2276fje.

Cancer, T. *et al.* (2014) 'Comprehensive molecular characterization of gastric

adenocarcinoma', *Nature*. Nature Publishing Group. doi: 10.1038/nature13480.

Cancer Genome Atlas Research Network *et al.* (2017) 'Integrated genomic characterization of oesophageal carcinoma.', *Nature*. doi: 10.1038/nature20805.

Cermak, T. *et al.* (2011) 'Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting', *Nucleic Acids Research*, 39(12), pp. e82–e82. doi: 10.1093/nar/gkr218.

Chen, C. X. *et al.* (2000) 'A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains.', *RNA (New York, N.Y.)*, 6(5), pp. 755–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10836796 (Accessed: 12 October 2017).

Chester, A. *et al.* (2000) 'RNA editing: cytidine to uridine conversion in apolipoprotein B mRNA.', *Biochimica et biophysica acta*, 1494(1–2), pp. 1–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11072063 (Accessed: 8 October 2017).

Chiang, C. *et al.* (2012) 'Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration', *Nature Genetics*. Nature Research, 44(4), pp. 390–397. doi: 10.1038/ng.2202.

Chiu, Y.-L. and Greene, W. C. (2008) 'The APOBEC3 Cytidine Deaminases: An Innate Defensive Network Opposing Exogenous Retroviruses and Endogenous Retroelements', *Annual Review of Immunology*. Annual Reviews , 26(1), pp. 317–353. doi: 10.1146/annurev.immunol.26.021607.090350.

Christian, M. *et al.* (2010) 'Targeting DNA Double-Strand Breaks with TAL Effector Nucleases', *Genetics*, 186(2), pp. 757–761. doi: 10.1534/genetics.110.120717.

Cimini, D. *et al.* (2001) 'Merotelic kinetochore orientation is a major mechanism of aneuploidy in mitotic mammalian tissue cells.', *The Journal of cell biology*. The Rockefeller University Press, 153(3), pp. 517–27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11331303 (Accessed: 5 October 2017).

Cong, L. *et al.* (2013) 'Multiplex Genome Engineering Using CRISPR/Cas Systems', *Science*, 339(6121), pp. 819–823. doi: 10.1126/science.1231143.

Conticello, S. G. (2008) 'The AID/APOBEC family of nucleic acid mutators.', *Genome biology*. BioMed Central, 9(6), p. 229. doi: 10.1186/gb-2008-9-6-229.

Crasta, K. *et al.* (2012) 'DNA breaks and chromosome pulverization from errors in mitosis.', *Nature*. Nature Publishing Group. doi: 10.1038/nature10802.

Crick, F. (1974) 'The double helix: a personal view', *Nature*. Nature Publishing Group, 248(5451), pp. 766–769. doi: 10.1038/248766a0.

Daley, J. M. and Wilson, T. E. (2005) 'Rejoining of DNA Double-Strand Breaks as a Function of Overhang Length', *Molecular and Cellular Biology*, 25(3), pp. 896–906. doi: 10.1128/MCB.25.3.896-906.2005.

Daniel, J. A. and Nussenzweig, A. (2013) 'The AID-Induced DNA Damage Response in Chromatin', *Molecular Cell*, 50(3), pp. 309–321. doi: 10.1016/j.molcel.2013.04.017.

Davoli, T. and de Lange, T. (2012) 'Telomere-Driven Tetraploidization Occurs in Human Cells Undergoing Crisis and Promotes Transformation of Mouse Cells', *Cancer Cell*, 21(6), pp. 765–776. doi: 10.1016/j.ccr.2012.03.044.

Decordier, I. *et al.* (2002) 'Elimination of micronucleated cells by apoptosis after treatment with inhibitors of microtubules', *Mutagenesis*. Oxford University Press, 17(4), pp. 337–344. doi: 10.1093/mutage/17.4.337.

Deming, P. B. *et al.* (2001) 'The human decatenation checkpoint', *Proceedings of the National Academy of Sciences*, 98(21), pp. 12044–12049. doi: 10.1073/pnas.221430898.

Deming, P. B. *et al.* (2002) 'ATR Enforces the Topoisomerase II-dependent G $_2$ Checkpoint through Inhibition of Plk1 Kinase', *Journal of Biological Chemistry*, 277(39), pp. 36832–36838. doi: 10.1074/jbc.M206109200.

Deng, D. *et al.* (2012) 'Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors', *Science*, 335(6069), pp. 720–723. doi: 10.1126/science.1215670.

Dewey, W. C., Ling, C. C. and Meyn, R. E. (1995) 'Radiation-induced apoptosis: Relevance to radiotherapy', *International Journal of Radiation Oncology*Biology*Physics*, 33(4), pp. 781–796. doi: 10.1016/0360-3016(95)00214-8.

Diffley, J. F. *et al.* (2000) 'Coping with and recovering from hydroxyurea-induced replication fork arrest in budding yeast.', *Cold Spring Harbor symposia on quantitative biology*, 65, pp. 333–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12760047 (Accessed: 30 October 2017).

Doudna, J. A. and Charpentier, E. (2014) 'The new frontier of genome engineering with CRISPR-Cas9', *Science*, 346(6213), pp. 1258096–1258096. doi:

10.1126/science.1258096.

Durrbaum, M. and Storchova, Z. (2015) 'Chromosomal Instability in Cancer Cells', *Chromosomal Instability in Cancer Cells*. doi: 10.1007/978-3-319-20291-4.

Fabre, F. *et al.* (2002) 'Alternate pathways involving Sgs1/Top3, Mus81/ Mms4, and Srs2 prevent formation of toxic recombination intermediates from single-stranded gaps created by DNA replication', *Proceedings of the National Academy of Sciences*, 99(26), pp. 16887–16892. doi: 10.1073/pnas.252652399.

Falck, J., Coates, J. and Jackson, S. P. (2005) 'Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage', *Nature*, 434(7033), pp. 605–611. doi: 10.1038/nature03442.

Farajollahi, S. and Maas, S. (2010) 'Molecular diversity through RNA editing: a balancing act', *Trends in Genetics*, 26(5), pp. 221–230. doi: 10.1016/j.tig.2010.02.001.

Franchitto, A., Oshima, J. and Pichierri, P. (2003) 'The G2-phase decatenation checkpoint is defective in Werner syndrome cells.', *Cancer research*, 63(12), pp. 3289–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12810661 (Accessed: 10 October 2017).

Friedberg, E. C. (2003) 'DNA damage and repair', *Nature*. Nature Publishing Group, 421(6921), pp. 436–440. doi: 10.1038/nature01408.

Friedberg, E. C. (2008) 'A brief history of the DNA repair field', *Cell Research*. Nature Publishing Group, 18(1), pp. 3–7. doi: 10.1038/cr.2007.113.

Gaj, T., Gersbach, C. A. and Barbas, C. F. (2013) 'ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering', *Trends in Biotechnology*, 31(7), pp. 397–405. doi: 10.1016/j.tibtech.2013.04.004.

Ganem, N. J., Godinho, S. A. and Pellman, D. (2009) 'A mechanism linking extra centrosomes to chromosomal instability.', *Nature*. Nature Publishing Group. doi: 10.1038/nature08136.

Gaudelli, N. M. *et al.* (2017) 'Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage', *Nature*. Nature Publishing Group, pp. 1–27. doi: 10.1038/nature24644.

Geigl, J. B. *et al.* (2008) 'Defining "chromosomal instability"', *Trends in Genetics*, 24(2), pp. 64–69. doi: 10.1016/j.tig.2007.11.006.

Gelsomino, F. *et al.* (2016) 'The evolving role of microsatellite instability in colorectal cancer: A review', *Cancer Treatment Reviews*. Elsevier Ltd, 51, pp. 19–26. doi: 10.1016/j.ctrv.2016.10.005.

Gemoll, T. *et al.* (2013) 'Chromosomal aneuploidy affects the global proteome equilibrium of colorectal cancer cells.', *Analytical cellular pathology (Amsterdam)*. Hindawi, 36(5–6), pp. 149–61. doi: 10.3233/ACP-140088.

Ghadimi, B. M. and Ried, T. (eds) (2015) *Chromosomal Instability in Cancer Cells*. Cham: Springer International Publishing (Recent Results in Cancer Research). doi: 10.1007/978-3-319-20291-4.

Ghadimi, B. M. and Ried, T. (no date a) *Chromosomal instability in cancer cells*.

Ghadimi, B. M. and Ried, T. (no date b) *Chromosomal instability in cancer cells*.

Giam, M. and Rancati, G. (2015) 'Aneuploidy and chromosomal instability in cancer: a jackpot to chaos.', *Cell division*. ??? doi: 10.1186/s13008-015-0009-7.

Giannattasio, M. *et al.* (2004) 'Physical and functional interactions between nucleotide excision repair and DNA damage checkpoint', *The EMBO Journal*, 23(2), pp. 429–438. doi: 10.1038/sj.emboj.7600051.

Gisselsson, D. *et al.* (2001) 'Abnormal Nuclear Shape in Solid Tumors Reflects Mitotic Instability', *The American Journal of Pathology*, 158(1), pp. 199–206. doi: 10.1016/S0002-9440(10)63958-2.

Gisselsson, D. (2008) 'Classification of chromosome segregation errors in cancer', *Chromosoma*, 117(6), pp. 511–519. doi: 10.1007/s00412-008-0169-1.

Gollin, S. M. (2005) 'Mechanisms leading to chromosomal instability', *Seminars in Cancer Biology*, 15(1), pp. 33–42. doi: 10.1016/j.semcancer.2004.09.004.

Gondek, L. P. *et al.* (2007) 'Single nucleotide polymorphism arrays complement metaphase cytogenetics in detection of new chromosomal lesions in MDS', *Leukemia*. Nature Publishing Group, 21(9), pp. 2058–2061. doi: 10.1038/sj.leu.2404745.

Gordon, D. J., Resio, B. and Pellman, D. (2012) 'Causes and consequences of aneuploidy in cancer', *Nature Reviews Genetics*. Nature Publishing Group, 13(3), pp. 189–203. doi: 10.1038/nrg3123.

Gottesfeld, J. M. *et al.* (1997) 'Regulation of gene expression by small molecules', *Nature*, 387(6629), pp. 202–205. doi: 10.1038/387202a0.

Granetto, C. *et al.* (1996) 'p53 accumulates in micronuclei after treatment with a DNA breaking chemical, methylnitrosourea, and with the spindle poison, vinblastine', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 352(1–2), pp. 61–64. doi: 10.1016/0027-5107(95)00235-9.

Haaf, T. *et al.* (1999) 'Sequestration of mammalian Rad51-recombination protein into micronuclei.', *The Journal of cell biology*. Rockefeller University Press, 144(1), pp. 11–20. doi: 10.1083/JCB.144.1.11.

Hanahan, D. and Weinberg, R. A. (2011) 'Review Hallmarks of Cancer : The Next Generation', *Cell*. Elsevier Inc., 144(5), pp. 646–674. doi: 10.1016/j.cell.2011.02.013.

Hanks, S. *et al.* (2004) 'Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in BUB1B', *Nature Genetics*. Nature Publishing Group, 36(11), pp. 1159–1161. doi: 10.1038/ng1449.

Harris, R. S. *et al.* (2003) 'DNA deamination mediates innate immunity to retroviral infection', *Cell*, 113(6), pp. 803–809. doi: 10.1016/S0092-8674(03)00423-9.

Harris, R. S. and Liddament, M. T. (2004) 'Retroviral restriction by APOBEC proteins', *Nature Reviews Immunology*. Nature Publishing Group, 4(11), pp. 868–877. doi: 10.1038/nri1489.

Harris, R. S., Petersen-Mahrt, S. K. and Neuberger, M. S. (2002) 'RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators.', *Molecular cell*, 10(5), pp. 1247–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12453430 (Accessed: 8 October 2017).

Hatch, E. M. *et al.* (2013) 'Catastrophic nuclear envelope collapse in cancer cell micronuclei.', *Cell*. NIH Public Access, 154(1), pp. 47–60. doi: 10.1016/j.cell.2013.06.007.

Heinrichs, S. and Look, A. T. (2007) 'Identification of structural aberrations in cancer by SNP array analysis', *Genome Biology*, 8(7), p. 219. doi: 10.1186/gb-2007-8-7-219.

Helleday, T., Eshtad, S. and Nik-Zainal, S. (2014) 'Mechanisms underlying mutational signatures in human cancers', *Nature Reviews Genetics*. Nature Research, 15(9), pp. 585–598. doi: 10.1038/nrg3729.

Henderson, S. *et al.* (2014) 'APOBEC-Mediated Cytosine Deamination Links PIK3CA Helical Domain Mutations to Human Papillomavirus-Driven Tumor Development', *Cell*

Reports, 7(6), pp. 1833–1841. doi: 10.1016/j.celrep.2014.05.012.

Hendriks, G. *et al.* (2010) 'Transcription-Dependent Cytosine Deamination Is a Novel Mechanism in Ultraviolet Light-Induced Mutagenesis', *Current Biology*, 20(2), pp. 170–175. doi: 10.1016/j.cub.2009.11.061.

Herbert, A. and Rich, A. (2001) 'The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1', *Proceedings of the National Academy of Sciences*, 98(21), pp. 12132–12137. doi: 10.1073/pnas.211419898.

Hirano, T. (2006) 'At the heart of the chromosome: SMC proteins in action', *Nature Reviews Molecular Cell Biology*, 7(5), pp. 311–322. doi: 10.1038/nrm1909.

Hoeijmakers, J. H. J. (2001) 'Genome maintenance mechanisms for preventing cancer.', *Nature*, 411(6835), pp. 366–374. doi: 10.1038/35077232.

Holland, A. J. and Cleveland, D. W. (2012) 'Chromoanagenesis and cancer: mechanisms and consequences of localized, complex chromosomal rearrangements.', *Nature medicine*. doi: 10.1038/nm.2988.

Hough, R. F. and Bass, B. L. (1997) 'Analysis of Xenopus dsRNA adenosine deaminase cDNAs reveals similarities to DNA methyltransferases.', *RNA (New York, N.Y.)*, 3(4), pp. 356–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9085843 (Accessed: 12 October 2017).

Huang, Y. *et al.* (2011) 'The fate of micronucleated cells post X-irradiation detected by live cell imaging', *DNA Repair*, 10(6), pp. 629–638. doi: 10.1016/j.dnarep.2011.04.010.

Huang, Y. *et al.* (2012) 'Lagging chromosomes entrapped in micronuclei are not "lost" by cells', *Cell Research*, 22(5), pp. 932–935. doi: 10.1038/cr.2012.26.

Imre Berger, ‡,§ *et al.* (1998) 'Spectroscopic Characterization of a DNA-Binding Domain, Zα, from the Editing Enzyme, dsRNA Adenosine Deaminase: Evidence for Left-Handed Z-DNA in the Zα−DNA Complex†'. American Chemical Society . doi: 10.1021/BI9813126.

Ishkanian, A. S. *et al.* (2004) 'A tiling resolution DNA microarray with complete coverage of the human genome', *Nature Genetics*, 36(3), pp. 299–303. doi: 10.1038/ng1307.

Iwaizumi, M. *et al.* (no date) 'Human Sgo1 downregulation leads to chromosomal instability in colorectal cancer'. doi: 10.1136/gut.2008.149468.

Jackson, A. L., Chen, R. and Loeb, L. A. (1998) 'Induction of microsatellite instability by oxidative DNA damage', *Proceedings of the National Academy of Sciences*, 95(21), pp. 12468–12473. doi: 10.1073/pnas.95.21.12468.

Jackson, S. P. and Bartek, J. (2009) 'The DNA-damage response in human biology and disease', *Nature*, 461(7267), pp. 1071–1078. doi: 10.1038/nature08467.

Janssen, A. *et al.* (2011) 'Chromosome Segregation Errors as a Cause of DNA Damage and Structural Chromosome Aberrations', *Science*, 333(6051), pp. 1895–1898. doi: 10.1126/science.1210214.

Jao, L.-E., Wente, S. R. and Chen, W. (2013) 'Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system', *Proceedings of the National Academy of Sciences*, 110(34), pp. 13904–13909. doi: 10.1073/pnas.1308335110.

Jarmuz, A. *et al.* (2002) 'An Anthropoid-Specific Locus of Orphan C to U RNA-Editing Enzymes on Chromosome 22', *Genomics*, 79(3), pp. 285–296. doi: 10.1006/geno.2002.6718.

Jefford, C. E. and Irminger-Finger, I. (2006) 'Mechanisms of chromosome instability in cancers', *Critical Reviews in Oncology/Hematology*, 59(1), pp. 1–14. doi: 10.1016/j.critrevonc.2006.02.005.

Jinek, M. *et al.* (2012) 'A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity', *Science*, 337(6096), pp. 816–821. doi: 10.1126/science.1225829.

Jiricny, J. (2006) 'The multifaceted mismatch-repair system', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 7(5), pp. 335–346. doi: 10.1038/nrm1907.

John Savage (2000) 'atlasgeneticsoncology'. Available at: http://atlasgeneticsoncology.org/Deep/MicronucleiID20016.html (Accessed: 19 October 2017).

Jones, L. E. *et al.* (2009) 'Differential effects of reactive nitrogen species on DNA base excision repair initiated by the alkyladenine DNA glycosylase', *Carcinogenesis*, 30(12), pp. 2123–2129. doi: 10.1093/carcin/bgp256.

Kanu, N. *et al.* (2016) 'DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer.', *Genome biology*. BioMed Central, 17(1), p. 185. doi: 10.1186/s13059-016-1042-9.

Keegan, L. P., Gallo, A. and O'Connell, M. A. (2001) 'The many roles of an RNA editor.', *Nature Reviews Genetics*, 2(11), pp. 869–878. doi: 10.1038/35098584.

Kim, Y. *et al.* (2013) 'A library of TAL effector nucleases spanning the human genome', *Nature Biotechnology*, 31(3), pp. 251–258. doi: 10.1038/nbt.2517.

Kim, Y. G., Cha, J. and Chandrasegaran, S. (1996) 'Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 93(3), pp. 1156–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8577732 (Accessed: 14 October 2017).

Kirsch-Volders, M. *et al.* (2011) 'The in vitro MN assay in 2011: origin and fate, biological significance, protocols, high throughput methodologies and toxicological relevance', *Archives of Toxicology*. Springer-Verlag, 85(8), pp. 873–899. doi: 10.1007/s00204-011-0691-4.

Klein, I. A. *et al.* (2011) 'Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes.', *Cell*. Elsevier, 147(1), pp. 95–106. doi: 10.1016/j.cell.2011.07.048.

Koito, A. and Ikeda, T. (2012) 'Apolipoprotein B mRNA-editing, catalytic polypeptide cytidine deaminases and retroviral restriction', *Wiley Interdisciplinary Reviews: RNA*. John Wiley & Sons, Inc., 3(4), pp. 529–541. doi: 10.1002/wrna.1117.

Komor, A. C. *et al.* (2016) 'Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage', *Nature*. Nature Publishing Group, 533(7603), pp. 420–424. doi: 10.1038/nature17946.

Komor, A. C. *et al.* (2017) 'Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity', *Science Advances*, 3(8), p. eaao4774. doi: 10.1126/sciadv.aao4774.

Krokan, H. E. and Bjørås, M. (2013) 'Base excision repair.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 5(4), p. a012583. doi: 10.1101/cshperspect.a012583.

Kuo, L. J. and Yang, L.-X. (no date) 'Gamma-H2AX - a novel biomarker for DNA double-strand breaks.', *In vivo (Athens, Greece)*, 22(3), pp. 305–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18610740 (Accessed: 19 October 2017).

De La Chapelle, A. and Hampel, H. (2010) 'Clinical relevance of microsatellite instability in colorectal cancer', *Journal of Clinical Oncology*, 28(20), pp. 3380–3387. doi: 10.1200/JCO.2009.27.0652.

Lada, A. G. *et al.* (2011) 'Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast.', *Biochemistry. Biokhimiia*, 76(1), pp. 131–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21568845 (Accessed: 8 October 2017).

Lammens, K. *et al.* (2011) 'The Mre11:Rad50 structure shows an ATP-dependent molecular clamp in DNA double-strand break repair.', *Cell*. Elsevier, 145(1), pp. 54–66. doi: 10.1016/j.cell.2011.02.038.

Law, E. K. *et al.* (2016) 'The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer'. Available at: http://advances.sciencemag.org/content/advances/2/10/e1601737.full.pdf (Accessed: 8 October 2017).

LEE, H.-C. *et al.* (2005) 'Mitochondrial Genome Instability and mtDNA Depletion in Human Cancers', *Annals of the New York Academy of Sciences*. Blackwell Publishing Ltd, 1042(1), pp. 109–122. doi: 10.1196/annals.1338.011.

Lee, H.-S. *et al.* (2013) 'A new assay for measuring chromosome instability (CIN) and identification of drugs that elevate CIN in cancer cells', *BMC Cancer*. BioMed Central, 13(1), p. 252. doi: 10.1186/1471-2407-13-252.

LEHMANN, A. *et al.* (2007) 'Translesion synthesis: Y-family polymerases and the polymerase switch', *DNA Repair*, 6(7), pp. 891–899. doi: 10.1016/j.dnarep.2007.02.003.

Leibowitz, M. L., Zhang, C.-Z. and Pellman, D. (2015) 'Chromothripsis: A New Mechanism for Rapid Karyotype Evolution.', *Annual review of genetics*. doi: 10.1146/annurev-genet-120213-092228.

Lengauer, C., Kinzler, K. W. and Vogelstein, B. (1998) 'Genetic instabilities in human cancers.', *Nature*, 396(6712), pp. 643–649. doi: 10.1038/25292.

Lettier, G. *et al.* (2006) 'The Role of DNA Double-Strand Breaks in Spontaneous Homologous Recombination in S. cerevisiae', *PLoS Genetics*, 2(11), p. e194. doi: 10.1371/journal.pgen.0020194.

Li, G.-M. (2008) 'Mechanisms and functions of DNA mismatch repair', *Cell Research*, 18(1), pp. 85–98. doi: 10.1038/cr.2007.115.

Lindahl, T. (1993) 'Instability and decay of the primary structure of DNA', *Nature*. Nature Publishing Group, 362(6422), pp. 709–715. doi: 10.1038/362709a0.

Lindahl, T. and Barnes, D. E. (2000) 'Repair of endogenous DNA damage.', *Cold Spring Harbor symposia on quantitative biology*, 65, pp. 127–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12760027 (Accessed: 7 October 2017).

Liu, M. and Schatz, D. G. (2009) 'Balancing AID and DNA repair during somatic hypermutation', *Trends in Immunology*, 30(4), pp. 173–181. doi: 10.1016/j.it.2009.01.007.

Loeb, L. A., Springgate, C. F. and Battula, N. (1974) 'Errors in DNA replication as a basis of malignant changes.', *Cancer research*, 34(9), pp. 2311–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4136142 (Accessed: 8 October 2017).

Macbeth, M. R. *et al.* (2005) 'Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing.', *Science (New York, N.Y.)*. NIH Public Access, 309(5740), pp. 1534–9. doi: 10.1126/science.1113150.

Maciejowski, J. *et al.* (2015) 'Chromothripsis and Kataegis Induced by Telomere Crisis', *Cell*. Elsevier Ltd. doi: 10.1016/j.cell.2015.11.054.

Mak, A. N.-S. *et al.* (2012) 'The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target', *Science*, 335(6069), pp. 716–719. doi: 10.1126/science.1216211.

Mali, P. *et al.* (2013) 'RNA-guided human genome engineering via Cas9.', *Science (New York, N.Y.)*. NIH Public Access, 339(6121), pp. 823–6. doi: 10.1126/science.1232033.

Mangeat, B. *et al.* (2003) 'Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts', *Nature*, 424(6944), pp. 99–103. doi: 10.1038/nature01709.

Markossian, S. *et al.* (2016) 'Quantitative assessment of chromosome instability induced through chemical disruption of mitotic progression', *Cell Cycle*. Taylor & Francis, 15(13), pp. 1706–1714. doi: 10.1080/15384101.2016.1175796.

Matsuoka, S. *et al.* (2007) 'ATM and ATR Substrate Analysis Reveals Extensive Protein Networks Responsive to DNA Damage', *Science*, 316(5828), pp. 1160–1166. doi: 10.1126/science.1140321.

Matthews, M. M. *et al.* (2016) 'Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity', *Nature Structural & Molecular*

Biology, 23(5), pp. 426–433. doi: 10.1038/nsmb.3203.

McClintock, B. (1941) 'The Stability of Broken Ends of Chromosomes in Zea Mays.', *Genetics*. Genetics Society of America, 26(2), pp. 234–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17247004 (Accessed: 5 October 2017).

McGranahan, N. *et al.* (2012) 'Cancer chromosomal instability: therapeutic and diagnostic challenges.', *EMBO reports*. Nature Publishing Group. doi: 10.1038/embor.2012.61.

Meaburn, K. J., Parris, C. N. and Bridger, J. M. (2005) 'The manipulation of chromosomes by mankind: the uses of microcell-mediated chromosome transfer', *Chromosoma*. Springer-Verlag, 114(4), pp. 263–274. doi: 10.1007/s00412-005-0014-8.

Meeker, A. K. *et al.* (2002) 'Telomere Length Assessment in Human Archival Tissues', *The American Journal of Pathology*, 160(4), pp. 1259–1268. doi: 10.1016/S0002-9440(10)62553-9.

Meng, F.-L. *et al.* (2014) 'Convergent Transcription at Intragenic Super-Enhancers Targets AID-Initiated Genomic Instability', *Cell*, 159(7), pp. 1538–1548. doi: 10.1016/j.cell.2014.11.014.

Mercer, A. C. *et al.* (2012) 'Chimeric TALE recombinases with programmable DNA sequence specificity', *Nucleic Acids Research*, 40(21), pp. 11163–11172. doi: 10.1093/nar/gks875.

Merlo, L. M. F. *et al.* (2006) 'Cancer as an evolutionary and ecological process', *Nature Reviews Cancer*, 6(12), pp. 924–935. doi: 10.1038/nrc2013.

Miller, J. C. *et al.* (2011) 'A TALE nuclease architecture for efficient genome editing', *Nature Biotechnology*, 29(2), pp. 143–148. doi: 10.1038/nbt.1755.

Minocherhomji, S. *et al.* (2015) 'Replication stress activates DNA repair synthesis in mitosis', *Nature*. Nature Research, 528(7581), pp. 286–290. doi: 10.1038/nature16139.

Mitelman, F., Mertens, F. and Johansson, B. (1997) 'A breakpoint map of recurrent chromosomal rearrangements in human neoplasia', *Nature Genetics*, 15(4), pp. 417–474. doi: 10.1038/ng0497supp-417.

Morganella, S. *et al.* (2016) 'The topography of mutational processes in breast cancer genomes', *Nature Communications*. Nature Publishing Group, 7, p. 11383. doi: 10.1038/ncomms11383.

Müller, A. and Fishel, R. (2002) 'Mismatch repair and the hereditary non-polyposis colorectal cancer syndrome (HNPCC).', *Cancer investigation*, 20(1), pp. 102–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11852992 (Accessed: 10 October 2017).

Muramatsu, M. *et al.* (2000) 'Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme.', *Cell*, 102(5), pp. 553–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11007474 (Accessed: 8 October 2017).

Nakano, M. *et al.* (2008) 'Inactivation of a Human Kinetochore by Specific Targeting of Chromatin Modifiers', *Developmental Cell*. Elsevier Inc. doi: 10.1016/j.devcel.2008.02.001.

Natarajan, A. T. and Palitti, F. (2008) 'DNA repair and chromosomal alterations', *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 657(1), pp. 3–7. doi: 10.1016/j.mrgentox.2008.08.017.

Navaratnam, N. *et al.* (1993) 'The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase', *Journal of Biological Chemistry*, 268(28), pp. 20709–20712.

Negrini, S., Gorgoulis, V. G. and Halazonetis, T. D. (2010) 'Genomic instability — an evolving hallmark of cancer', *Nature Reviews Molecular Cell Biology*, 11(3), pp. 220–228. doi: 10.1038/nrm2858.

Nik-Zainal, S. *et al.* (2016) 'Landscape of somatic mutations in 560 breast cancer whole-genome sequences', *Nature*. Nature Research, 534(7605), pp. 47–54. doi: 10.1038/nature17676.

Nishida, K. *et al.* (2016) 'Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems', *Science*, 353(6305), p. aaf8729-aaf8729. doi: 10.1126/science.aaf8729.

Nishikura, K. (2010) 'Functions and Regulation of RNA Editing by ADAR Deaminases', *Annual Review of Biochemistry*, 79(1), pp. 321–349. doi: 10.1146/annurev-biochem-060208-105251.

Nishizaki, T. *et al.* (1997) 'Genetic alterations in lobular breast cancer by comparative genomic hybridization.', *International journal of cancer*, 74(5), pp. 513–7. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/9355973 (Accessed: 5 October 2017).

Di Noia, J. and Neuberger, M. S. (2002) 'Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase', *Nature*, pp. 43–48. doi: 10.1038/nature00981.

Di Noia, J. M. and Neuberger, M. S. (2007) 'Molecular Mechanisms of Antibody Somatic Hypermutation', *Annual Review of Biochemistry*, 76(1), pp. 1–22. doi: 10.1146/annurev.biochem.76.061705.090740.

Nordentoft, I. *et al.* (2014) 'Mutational Context and Diverse Clonal Development in Early and Late Bladder Cancer', *Cell Reports*, 7(5), pp. 1649–1663. doi: 10.1016/j.celrep.2014.04.038.

Nowak, M. A. *et al.* (2002) 'The role of chromosomal instability in tumor initiation', *Proceedings of the National Academy of Sciences*, 99(25), pp. 16226–16231. doi: 10.1073/pnas.202617399.

Nowell, P. C. (1976) 'The clonal evolution of tumor cell populations.', *Science (New York, N.Y.)*, 194(4260), pp. 23–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/959840 (Accessed: 8 October 2017).

Olivier, M., Hollstein, M. and Hainaut, P. (2010) 'TP53 mutations in human cancers: origins, consequences, and clinical use.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 2(1), p. a001008. doi: 10.1101/cshperspect.a001008.

Oromendia, A. B. and Amon, A. (2014) 'Aneuploidy: implications for protein homeostasis and disease', *Disease Models & Mechanisms*, 7(1), pp. 15–20. doi: 10.1242/dmm.013391.

Le Page, F. *et al.* (2005) 'Transcription-coupled repair of 8-oxoguanine: requirement for XPG, TFIIH, and CSB and implications for Cockayne syndrome.', *Cell*, 123(4), p. 711. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16329169 (Accessed: 10 October 2017).

Pauklin, S. *et al.* (2009) 'Estrogen directly activates AID transcription and function', *The Journal of Experimental Medicine*, 206(1), pp. 99–111. doi: 10.1084/jem.20080521.

Peltomaki, P. (2001) 'Deficient DNA mismatch repair: a common etiologic factor for colon cancer', *Human Molecular Genetics*. Oxford University Press, 10(7), pp. 735–740.

doi: 10.1093/hmg/10.7.735.

Petersen-Mahrt, S. K., Harris, R. S. and Neuberger, M. S. (2002) 'AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification', *Nature*. Nature Publishing Group, 418(6893), pp. 99–104. doi: 10.1038/nature00862.

Petersen, S. *et al.* (2001) 'AID is required to initiate Nbs1/γ-H2AX focus formation and mutations at sites of class switching', *Nature*, 414(6864), pp. 660–665. doi: 10.1038/414660a.

Pfeifer, G. P., You, Y.-H. and Besaratinia, A. (2005) 'Mutations induced by ultraviolet light', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 571(1–2), pp. 19–31. doi: 10.1016/j.mrfmmm.2004.06.057.

Pikor, L. *et al.* (2013) 'The detection and implication of genome instability in cancer', *Cancer and Metastasis Reviews*, 32(3–4), pp. 341–352. doi: 10.1007/s10555-013-9429-5.

Pinkel, D. *et al.* (1998) 'High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.', *Nature Genetics*, 20(2), pp. 207–211. doi: 10.1038/2524.

Pollack, J. R. *et al.* (2002) 'Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors', *Proceedings of the National Academy of Sciences*, 99(20), pp. 12963–12968. doi: 10.1073/pnas.162471999.

Popp, H. D. and Bohlander, S. K. (2010) 'Genetic Instability in Inherited and Sporadic Leukemias', *GENES, CHROMOSOMES & CANCER*, 49, pp. 1071–1081. doi: 10.1002/gcc.20823.

Qian, J. *et al.* (2014) 'B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity.', *Cell*. NIH Public Access, 159(7), pp. 1524–37. doi: 10.1016/j.cell.2014.11.013.

Ramiro, A. R. *et al.* (2004) 'AID Is Required for c-myc/IgH Chromosome Translocations In Vivo', *Cell*, 118(4), pp. 431–438. doi: 10.1016/j.cell.2004.08.006.

Ramiro, A. R. *et al.* (2006) 'Role of genomic instability and p53 in AID-induced c-myc–Igh translocations', *Nature*. Nature Publishing Group, 440(7080), pp. 105–109. doi: 10.1038/nature04495.

Ramiro, A. R., Nussenzweig, M. C. and Nussenzweig, A. (2006) 'Switching on

chromosomal translocations', *Cancer Research*, pp. 7837–7839. doi: 10.1158/0008-5472.CAN-06-0863.

Rao, C. V. and Yamada, H. Y. (2013) 'Genomic Instability and Colon Carcinogenesis: From the Perspective of Genes', *Frontiers in Oncology*, 3(May), pp. 1–13. doi: 10.3389/fonc.2013.00130.

Rassool, F. V. (2003) 'DNA double strand breaks (DSB) and non-homologous end joining (NHEJ) pathways in human leukemia.', *Cancer letters*, 193(1), pp. 1–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12691817 (Accessed: 10 October 2017).

Reyon, D. *et al.* (2012) 'FLASH assembly of TALENs for high-throughput genome editing', *Nature Biotechnology*, 30(5), pp. 460–465. doi: 10.1038/nbt.2170.

'RNA editing in brain controls a determinant of ion flow in glutamate-gated channels' (1991) *Cell*. Cell Press, 67(1), pp. 11–19. doi: 10.1016/0092-8674(91)90568-J.

Roberts, S. A. *et al.* (2012) 'Clustered Mutations in Yeast and in Human Cancers Can Arise from Damaged Long Single-Strand DNA Regions', *Molecular Cell*, 46(4), pp. 424–435. doi: 10.1016/j.molcel.2012.03.030.

Roger, A. S. *et al.* (2006) *DNA Repair and Mutagenesis, Second Edition*. American Society of Microbiology. doi: 10.1128/9781555816704.

Rogozin, I. B. *et al.* (2007) 'Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase', *Nature Immunology*, 8(6), pp. 647–656. doi: 10.1038/ni1463.

Roschke, A. V. *et al.* (2003) 'Karyotypic Complexity of the NCI-60 Drug-Screening Panel', *Cancer Research*, 63(24), pp. 8634–8647.

Rosenberg, B. R. *et al.* (2011) 'Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs', *Nature Structural & Molecular Biology*, 18(2), pp. 230–236. doi: 10.1038/nsmb.1975.

Roth, D. B. and Wilson, J. H. (1986) 'Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction.', *Molecular and cellular biology*. American Society for Microbiology (ASM), 6(12), pp. 4295–304. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3025650 (Accessed: 10 October 2017).

Russo, M. T. *et al.* (2004) 'Accumulation of the Oxidative Base Lesion 8-Hydroxyguanine in DNA of Tumor-Prone Mice Defective in Both the ***Myh*** and ***Ogg1***

DNA Glycosylases', *Cancer Research*, 64(13), pp. 4411–4414. doi: 10.1158/0008-5472.CAN-04-0355.

Sambrook, J. and Russell, D. W. (David W. (2001) *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory Press.

Sancar, A. *et al.* (2004) 'Molecular Mechanisms of Mammalian DNA Repair and the DNA Damage Checkpoints', *Annual Review of Biochemistry*, 73(1), pp. 39–85. doi: 10.1146/annurev.biochem.73.011303.073723.

Sansregret, L. and Swanton, C. (2017) 'The role of aneuploidy in cancer evolution', *Cold Spring Harbor Perspectives in Medicine*, 7(1), pp. 1–18. doi: 10.1101/cshperspect.a028373.

Sato, Y. *et al.* (2010) 'Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy', *Journal of Biological Chemistry*, 285(10), pp. 7111–7118. doi: 10.1074/jbc.M109.052977.

Savage, J. R. (1988) 'A comment on the quantitative relationship between micronuclei and chromosomal aberrations.', *Mutation research*, 207(1), pp. 33–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3336377 (Accessed: 5 October 2017).

Savva, Y. A., Rieder, L. E. and Reenan, R. A. (2012) 'The ADAR protein family', *Genome Biology*. BioMed Central, 13(12), p. 252. doi: 10.1186/gb-2012-13-12-252.

Schmid-Burgk, J. L. *et al.* (2012) 'A ligation-independent cloning technique for high-throughput assembly of transcription activator–like effector genes', *Nature Biotechnology*, 31(1), pp. 76–81. doi: 10.1038/nbt.2460.

Schoch, C. *et al.* (2005) 'Genomic gains and losses influence expression levels of genes located within the affected regions: a study on acute myeloid leukemias with trisomy 8, 11, or 13, monosomy 7, or deletion 5q', *Leukemia*, 19(7), pp. 1224–1228. doi: 10.1038/sj.leu.2403810.

Schwartz, T. *et al.* (1999) 'Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA.', *Science (New York, N.Y.)*, 284(5421), pp. 1841–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10364558 (Accessed: 12 October 2017).

Severi, F., Chicca, A. and Conticello, S. G. (2011a) 'Analysis of Reptilian APOBEC1 Suggests that RNA Editing May Not Be Its Ancestral Function', *Molecular Biology and*

Evolution, 28(3), pp. 1125–1129. doi: 10.1093/molbev/msq338.

Severi, F., Chicca, A. and Conticello, S. G. (2011b) 'Analysis of Reptilian APOBEC1 Suggests that RNA Editing May Not Be Its Ancestral Function', *Molecular Biology and Evolution*, 28(3), pp. 1125–1129. doi: 10.1093/molbev/msq338.

Sheehy, A. M. *et al.* (2002) 'Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein', *Nature*, 418(6898), pp. 646–650. doi: 10.1038/nature00939.

Shimizu, N., Misaka, N. and Utani, K. (2007) 'Nonselective DNA damage induced by a replication inhibitor results in the selective elimination of extrachromosomal double minutes from human cancer cells', *Genes, Chromosomes and Cancer*. Wiley Subscription Services, Inc., A Wiley Company, 46(10), pp. 865–874. doi: 10.1002/gcc.20473.

Shuck, S. C., Short, E. A. and Turchi, J. J. (2008) 'Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology', *Cell Research*, 18(1), pp. 64–72. doi: 10.1038/cr.2008.2.

Sjögren, C. and Nasmyth, K. (2001) 'Sister chromatid cohesion is required for postreplicative double-strand break repair in Saccharomyces cerevisiae.', *Current biology : CB*, 11(12), pp. 991–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11448778 (Accessed: 10 October 2017).

Smith, H. C. *et al.* (2012) 'Functions and regulation of the APOBEC family of proteins', *Seminars in Cell & Developmental Biology*, 23(3), pp. 258–268. doi: 10.1016/j.semcdb.2011.10.004.

Sotillo, R. *et al.* (2007) 'Mad2 Overexpression Promotes Aneuploidy and Tumorigenesis in Mice', *Cancer Cell*, 11(1), pp. 9–23. doi: 10.1016/j.ccr.2006.10.019.

Soutoglou, E. *et al.* (2007) 'Positional stability of single double-strand breaks in mammalian cells.', *Nature cell biology*. NIH Public Access, 9(6), pp. 675–82. doi: 10.1038/ncb1591.

Stenglein, M. D. *et al.* (2010) 'APOBEC3 proteins mediate the clearance of foreign DNA from human cells', *Nature Structural & Molecular Biology*. Nature Research, 17(2), pp. 222–229. doi: 10.1038/nsmb.1744.

Stephens, O. M., Yi-Brunozzi, H. Y. and Beal, P. A. (2000) 'Analysis of the RNA-editing

reaction of ADAR2 with structural and fluorescent analogues of the GluR-B R/G editing site.', *Biochemistry*, 39(40), pp. 12243–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11015203 (Accessed: 12 October 2017).

Stephens, P. J. *et al.* (2011) 'Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development', *Cell*, 144(1), pp. 27–40. doi: 10.1016/j.cell.2010.11.055.

Strobel, S. A. and Dervan, P. B. (1990) 'Site-specific cleavage of a yeast chromosome by oligonucleotide-directed triple-helix formation.', *Science (New York, N.Y.)*, 249(4964), pp. 73–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2195655 (Accessed: 14 October 2017).

Ström, L. *et al.* (2004) 'Postreplicative Recruitment of Cohesin to Double-Strand Breaks Is Required for DNA Repair', *Molecular Cell*, 16(6), pp. 1003–1015. doi: 10.1016/j.molcel.2004.11.026.

Stuart, K. D. *et al.* (2005) 'Complex management: RNA editing in trypanosomes.', *Trends in biochemical sciences*. Elsevier, 30(2), pp. 97–105. doi: 10.1016/j.tibs.2004.12.006.

Swanton, C. *et al.* (2015) 'APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity', *Cancer Discovery*, 5(7), pp. 704–712. doi: 10.1158/2159-8290.CD-15-0344.

Swift, L. H. and Golsteyn, R. M. (2014) 'Genotoxic anti-cancer agents and their relationship to DNA damage, mitosis, and checkpoint adaptation in proliferating cancer cells.', *International journal of molecular sciences*. Multidisciplinary Digital Publishing Institute (MDPI), 15(3), pp. 3403–31. doi: 10.3390/ijms15033403.

Takata, M. *et al.* (1998) 'Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells.', *The EMBO journal*. European Molecular Biology Organization, 17(18), pp. 5497–508. doi: 10.1093/emboj/17.18.5497.

Taylor, B. S. *et al.* (2010) 'Integrative Genomic Profiling of Human Prostate Cancer', *Cancer Cell*, 18(1), pp. 11–22. doi: 10.1016/j.ccr.2010.05.026.

Terradas, M. *et al.* (2009) 'DNA lesions sequestered in micronuclei induce a local defective-damage response', *DNA Repair*, 8(10), pp. 1225–1234. doi:

10.1016/j.dnarep.2009.07.004.

Terradas, M. *et al.* (2012) 'Is DNA Damage Response Ready for Action Anywhere?', *International Journal of Molecular Sciences*. Molecular Diversity Preservation International, 13(12), pp. 11569–11583. doi: 10.3390/ijms130911569.

Terradas, M., Mart??n, M. and Genesc??, A. (2016) 'Impaired nuclear functions in micronuclei results in genome instability and chromothripsis', *Archives of Toxicology*. Springer Berlin Heidelberg. doi: 10.1007/s00204-016-1818-4.

Terzoudi, G. I. *et al.* (2015) 'Stress induced by premature chromatin condensation triggers chromosome shattering and chromothripsis at DNA sites still replicating in micronuclei or multinucleate cells when primary nuclei enter mitosis', *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*. Elsevier Ltd. doi: 10.1016/j.mrgentox.2015.07.014.

Thomas, J. M. and Beal, P. A. (2017) 'How do ADARs bind RNA? New protein-RNA structures illuminate substrate recognition by the RNA editing ADARs', *BioEssays*, 39(4), p. 1600187. doi: 10.1002/bies.201600187.

Thompson, S. L. and Compton, D. A. (2011) 'Chromosome missegregation in human cells arises through specific types of kinetochore-microtubule attachment errors.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(44), pp. 17974–8. doi: 10.1073/pnas.1109720108.

Tomasetti, C. and Vogelstein, B. (2015) 'Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions.', *Science (New York, N.Y.)*. NIH Public Access, 347(6217), pp. 78–81. doi: 10.1126/science.1260825.

Torres, E. (2015) 'Yeast as Models of Mitotic Fidelity', in, pp. 143–164. doi: 10.1007/978-3-319-20291-4_7.

Torres, E. M. *et al.* (2007) 'Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast', *Science*, 317(5840), pp. 916–924. doi: 10.1126/science.1142210.

Tsafrir, D. *et al.* (2006) 'Relationship of Gene Expression and Chromosomal Abnormalities in Colorectal Cancer', *Cancer Research*, 66(4), pp. 2129–2137. doi: 10.1158/0008-5472.CAN-05-2569.

Tsuruoka, N. *et al.* (2013) 'ADAR1 protein induces adenosine-targeted DNA mutations

in senescent Bcl6 gene-deficient cells.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 288(2), pp. 826–36. doi: 10.1074/jbc.M112.365718.

Tubbs, A. and Nussenzweig, A. (2017) 'Endogenous DNA Damage as a Source of Genomic Instability in Cancer', *Cell*, 168(4), pp. 644–656. doi: 10.1016/j.cell.2017.01.002.

Uhlmann, F. and Nasmyth, K. (1998) 'Cohesion between sister chromatids must be established during DNA replication.', *Current biology : CB*, 8(20), pp. 1095–101. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9778527 (Accessed: 10 October 2017).

Utani, K., Okamoto, A. and Shimizu, N. (2011) 'Generation of Micronuclei during Interphase by Coupling between Cytoplasmic Membrane Blebbing and Nuclear Budding', *PLoS ONE*. Edited by J. M. Bridger. Public Library of Science, 6(11), p. e27233. doi: 10.1371/journal.pone.0027233.

Venkatesan, S., Natarajan, A. T. and Hande, M. P. (2015) 'Chromosomal instability-mechanisms and consequences', *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*. Elsevier Ltd, 793, pp. 176–184. doi: 10.1016/j.mrgentox.2015.08.008.

Vermeulen, K., Van Bockstaele, D. R. and Berneman, Z. N. (2003) 'The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer.', *Cell proliferation*, 36(3), pp. 131–49. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12814430 (Accessed: 10 October 2017).

Vieira, V. C. and Soares, M. A. (2013) 'The Role of Cytidine Deaminases on Innate Immune Responses against Human Viral Infections', *BioMed Research International*, 2013, pp. 1–18. doi: 10.1155/2013/683095.

Vilardell, M. *et al.* (2011) 'Meta-analysis of heterogeneous Down Syndrome data reveals consistent genome-wide dosage effects related to neurological processes', *BMC Genomics*, 12(1), p. 229. doi: 10.1186/1471-2164-12-229.

Walker, J. A. *et al.* (1996) 'Chromosome content and ultrastructure of radiation-induced micronuclei.', *Mutagenesis*, 11(5), pp. 419–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8921502 (Accessed: 6 October 2017).

Wang, H. *et al.* (2003) 'Biochemical evidence for Ku-independent backup pathways of NHEJ.', *Nucleic acids research*. Oxford University Press, 31(18), pp. 5377–88. doi: 10.1093/NAR/GKG728.

Wang, H. *et al.* (2013) 'One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering', *Cell*, 153(4), pp. 910–918. doi: 10.1016/j.cell.2013.04.025.

Wedekind, J. E. *et al.* (no date) 'Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business'. doi: 10.1016/S0168-9525(03)00054-4.

Whitwell, J. *et al.* (2015) 'Relationships between p53 status, apoptosis and induction of micronuclei in different human and mouse cell lines in vitro: Implications for improving existing assays', *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 789–790, pp. 7–27. doi: 10.1016/j.mrgentox.2015.05.011.

Williams, G. J., Lees-Miller, S. P. and Tainer, J. A. (2010) 'Mre11-Rad50-Nbs1 conformations and the control of sensing, signaling, and effector responses at DNA double-strand breaks.', *DNA repair*. NIH Public Access, 9(12), pp. 1299–306. doi: 10.1016/j.dnarep.2010.10.001.

Yasui, M. *et al.* (2010) 'Live cell imaging of micronucleus formation and development', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 692(1–2), pp. 12–18. doi: 10.1016/j.mrfmmm.2010.07.009.

Yi-Brunozzi, H. Y., Stephens, O. M. and Beal, P. A. (2001) 'Conformational changes that occur during an RNA-editing adenosine deamination reaction.', *The Journal of biological chemistry*, 276(41), pp. 37827–33. doi: 10.1074/jbc.M106299200.

You, Z. *et al.* (2005) 'ATM Activation and Its Recruitment to Damaged DNA Require Binding to the C Terminus of Nbs1', *Molecular and Cellular Biology*, 25(13), pp. 5363–5379. doi: 10.1128/MCB.25.13.5363-5379.2005.

Zeitlin, S. G. *et al.* (2005) 'Xenopus CENP-A assembly into chromatin requires base excision repair proteins', *DNA Repair*, 4(7), pp. 760–772. doi: 10.1016/j.dnarep.2005.02.007.

Zhang, C.-Z. *et al.* (2015) 'Chromothripsis from DNA damage in micronuclei', *Nature*. doi: 10.1038/nature14493.

Zhang, F. *et al.* (2011) 'Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription', *Nature Biotechnology*, 29(2), pp. 149–153. doi: 10.1038/nbt.1775.

Zhang, N. *et al.* (2008) 'Overexpression of Separase induces aneuploidy and mammary tumorigenesis.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(35), pp. 13033–8. doi: 10.1073/pnas.0801610105.

Zhao, X. *et al.* (2004) 'An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays.', *Cancer research*, 64(9), pp. 3060–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15126342 (Accessed: 11 October 2017).

Zharkov, D. O. (2008) 'Base excision DNA repair', *Cellular and Molecular Life Sciences*, 65(10), pp. 1544–1565. doi: 10.1007/s00018-008-7543-2.

Zheng, Y., Lorenzo, C. and Beal, P. A. (2017) 'DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA', *Nucleic Acids Research*. Oxford University Press, 45(6), p. gkx050. doi: 10.1093/nar/gkx050.

Zong, C. *et al.* (2012) 'Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell', *Science*, 338(6114), pp. 1622–1626. doi: 10.1126/science.1229164.

'A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine' (1987) *Cell*. Cell Press, 50(6), pp. 831–840. doi: 10.1016/0092-8674(87)90510-1.

Abend, M., Frombeck, S. and Van Beuningen, D. (1999) 'Indications for an active process underlying spontaneous and radiation-induced micronucleation in L929 cells', *International Journal of Radiation Biology*. Taylor & Francis, 75(12), pp. 1567–1578. doi: 10.1080/095530099139179.

Aït Yahya-Graison, E. *et al.* (2007) 'Classification of human chromosome 21 gene-expression variations in Down syndrome: impact on disease phenotypes.', *American journal of human genetics*. Elsevier, 81(3), pp. 475–91. doi: 10.1086/520000.

Alexandrov, L. B. *et al.* (2013) 'Signatures of mutational processes in human cancer', *Nature*, 500(7463), pp. 415–421. doi: 10.1038/nature12477.

Arakawa, H., HauschiLd, J. and Buerstedde, J. M. (2002) 'Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion', *Science*, 295(5558), pp. 1301–1306. doi: 10.1126/science.1067308.

Bakhoum, S. F. and Compton, D. a (2012) 'Science in medicine Chromosomal instability and cancer : a complex relationship with therapeutic potential', *The Journal of Clinical Investigation*. doi: 10.1172/JCI59954.1138.

Barber, T. D. *et al.* (2008) 'Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(9), pp. 3443–8. doi: 10.1073/pnas.0712384105.

Barrangou, R. *et al.* (2007) 'CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes', *Science*, 315(5819), pp. 1709–1712. doi: 10.1126/science.1138140.

Bartek, J. and Lukas, J. (2007) 'DNA damage checkpoints: from initiation to recovery or adaptation', *Current Opinion in Cell Biology*, 19(2), pp. 238–245. doi: 10.1016/j.ceb.2007.02.009.

Bass, B. L. (2002) 'RNA Editing by Adenosine Deaminases That Act on RNA', *Annual Review of Biochemistry*, 71(1), pp. 817–846. doi: 10.1146/annurev.biochem.71.110601.135501.

Bayani, J. *et al.* (2007) 'Genomic mechanisms and measurement of structural and numerical instability in cancer cells', *Seminars in Cancer Biology*, 17(1), pp. 5–18. doi: 10.1016/j.semcancer.2006.10.006.

Bayani, J. and Squire, J. A. (2001) 'Advances in the detection of chromosomal aberrations using spectral karyotyping.', *Clinical genetics*, 59(2), pp. 65–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11260203 (Accessed: 11 October 2017).

Bedard, L. L. and Massey, T. E. (2006) 'Aflatoxin B1-induced DNA damage and its repair', *Cancer Letters*, 241(2), pp. 174–183. doi: 10.1016/j.canlet.2005.11.018.

Beheshti, B. *et al.* (no date) 'Evidence of chromosomal instability in prostate cancer determined by spectral karyotyping (SKY) and interphase fish analysis.', *Neoplasia (New York, N.Y.)*, 3(1), pp. 62–9. doi: 10.1038/sj/neo/7900125.

Beroukhim, R. *et al.* (2010) 'The landscape of somatic copy-number alteration across human cancers', *Nature*. Nature Publishing Group, 463(7283), pp. 899–905. doi:

10.1038/nature08822.

Bhatia, A. and Kumar, Y. (2013) 'Cancer cell micronucleus: an update on clinical and diagnostic applications', *APMIS*, 121(7), pp. 569–581. doi: 10.1111/apm.12033.

Bibikova, M. *et al.* (2002) 'Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases.', *Genetics*, 161(3), pp. 1169–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12136019 (Accessed: 14 October 2017).

Bibikova, M. *et al.* (2003) 'Enhancing Gene Targeting with Designed Zinc Finger Nucleases', *Science*, 300(5620), pp. 764–764. doi: 10.1126/science.1079512.

Boboila, C. *et al.* (2010) 'Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase 4.', *The Journal of experimental medicine*. The Rockefeller University Press, 207(2), pp. 417–27. doi: 10.1084/jem.20092449.

Bonassi, S. *et al.* (2011) 'Micronuclei frequency in peripheral blood lymphocytes and cancer risk: evidence from human studies', *Mutagenesis*. Oxford University Press, 26(1), pp. 93–100. doi: 10.1093/mutage/geq075.

Boveri, T. (2007) 'Concerning the Origin of Malignant Tumours by Theodor Boveri. Translated and annotated by Henry Harris', *Journal of Cell Science*, 121(Supplement 1), p. 1 LP-84. Available at: http://jcs.biologists.org/content/121/Supplement_1/1.abstract.

Branzei, D. and Foiani, M. (2005) 'The DNA damage response during DNA replication', *Current Opinion in Cell Biology*, 17(6), pp. 568–575. doi: 10.1016/j.ceb.2005.09.003.

BRANZEI, D. and FOIANI, M. (2007) 'Interplay of replication checkpoints and repair proteins at stalled replication forks', *DNA Repair*, 6(7), pp. 994–1003. doi: 10.1016/j.dnarep.2007.02.018.

Briggs, A. W. *et al.* (2012) 'Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers', *Nucleic Acids Research*, 40(15), pp. e117–e117. doi: 10.1093/nar/gks624.

Bristow, R. G. and Hill, R. P. (2008) 'Hypoxia and metabolism: Hypoxia, DNA repair and genetic instability', *Nature Reviews Cancer*, 8(3), pp. 180–192. doi: 10.1038/nrc2344.

Brouns, S. J. J. *et al.* (2008) 'Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes', *Science*, 321(5891), pp. 960–964. doi: 10.1126/science.1159689.

Bunting, S. F. and Nussenzweig, A. (2013) 'End-joining, translocations and cancer',

Nature Reviews Cancer, 13(7), pp. 443–454. doi: 10.1038/nrc3537.

Burns, M. B. *et al.* (2013) 'APOBEC3B is an enzymatic source of mutation in breast cancer', *Nature*, 494(7437), pp. 366–370. doi: 10.1038/nature11881.

Burns, M. B., Temiz, N. A. and Harris, R. S. (2013) 'Evidence for APOBEC3B mutagenesis in multiple human cancers.', *Nature genetics*. NIH Public Access, 45(9), pp. 977–83. doi: 10.1038/ng.2701.

Burrell, R. A. *et al.* (2013) 'Replication stress links structural and numerical cancer chromosomal instability', *Nature*. Nature Research, 494(7438), pp. 492–496. doi: 10.1038/nature11935.

Cadet, J., Sage, E. and Douki, T. (2005) 'Ultraviolet radiation-mediated damage to cellular DNA', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 571(1–2), pp. 3–17. doi: 10.1016/j.mrfmmm.2004.09.012.

Cahill, D. P. *et al.* (1998) 'Mutations of mitotic checkpoint genes in human cancers', *Nature*, 392(6673), pp. 300–303. doi: 10.1038/32688.

Camps, J. *et al.* (2005) 'Comprehensive measurement of chromosomal instability in cancer cells: combination of fluorescence in situ hybridization and cytokinesis-block micronucleus assay', *The FASEB Journal*, 19(7), pp. 828–30. doi: 10.1096/fj.04-2276fje.

Cancer, T. *et al.* (2014) 'Comprehensive molecular characterization of gastric adenocarcinoma', *Nature*. Nature Publishing Group. doi: 10.1038/nature13480.

Cancer Genome Atlas Research Network *et al.* (2017) 'Integrated genomic characterization of oesophageal carcinoma.', *Nature*. doi: 10.1038/nature20805.

Cermak, T. *et al.* (2011) 'Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting', *Nucleic Acids Research*, 39(12), pp. e82–e82. doi: 10.1093/nar/gkr218.

Chen, C. X. *et al.* (2000) 'A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains.', *RNA (New York, N.Y.)*, 6(5), pp. 755–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10836796 (Accessed: 12 October 2017).

Chester, A. *et al.* (2000) 'RNA editing: cytidine to uridine conversion in apolipoprotein B mRNA.', *Biochimica et biophysica acta*, 1494(1–2), pp. 1–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11072063 (Accessed: 8 October 2017).

Chiang, C. *et al.* (2012) 'Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration', *Nature Genetics*. Nature Research, 44(4), pp. 390–397. doi: 10.1038/ng.2202.

Chiu, Y.-L. and Greene, W. C. (2008) 'The APOBEC3 Cytidine Deaminases: An Innate Defensive Network Opposing Exogenous Retroviruses and Endogenous Retroelements', *Annual Review of Immunology*. Annual Reviews , 26(1), pp. 317–353. doi: 10.1146/annurev.immunol.26.021607.090350.

Christian, M. *et al.* (2010) 'Targeting DNA Double-Strand Breaks with TAL Effector Nucleases', *Genetics*, 186(2), pp. 757–761. doi: 10.1534/genetics.110.120717.

Cimini, D. *et al.* (2001) 'Merotelic kinetochore orientation is a major mechanism of aneuploidy in mitotic mammalian tissue cells.', *The Journal of cell biology*. The Rockefeller University Press, 153(3), pp. 517–27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11331303 (Accessed: 5 October 2017).

Cong, L. *et al.* (2013) 'Multiplex Genome Engineering Using CRISPR/Cas Systems', *Science*, 339(6121), pp. 819–823. doi: 10.1126/science.1231143.

Conticello, S. G. (2008) 'The AID/APOBEC family of nucleic acid mutators.', *Genome biology*. BioMed Central, 9(6), p. 229. doi: 10.1186/gb-2008-9-6-229.

Crasta, K. *et al.* (2012) 'DNA breaks and chromosome pulverization from errors in mitosis.', *Nature*. Nature Publishing Group. doi: 10.1038/nature10802.

Crick, F. (1974) 'The double helix: a personal view', *Nature*. Nature Publishing Group, 248(5451), pp. 766–769. doi: 10.1038/248766a0.

Daley, J. M. and Wilson, T. E. (2005) 'Rejoining of DNA Double-Strand Breaks as a Function of Overhang Length', *Molecular and Cellular Biology*, 25(3), pp. 896–906. doi: 10.1128/MCB.25.3.896-906.2005.

Daniel, J. A. and Nussenzweig, A. (2013) 'The AID-Induced DNA Damage Response in Chromatin', *Molecular Cell*, 50(3), pp. 309–321. doi: 10.1016/j.molcel.2013.04.017.

Davoli, T. and de Lange, T. (2012) 'Telomere-Driven Tetraploidization Occurs in Human Cells Undergoing Crisis and Promotes Transformation of Mouse Cells', *Cancer Cell*, 21(6), pp. 765–776. doi: 10.1016/j.ccr.2012.03.044.

Decordier, I. *et al.* (2002) 'Elimination of micronucleated cells by apoptosis after

treatment with inhibitors of microtubules', *Mutagenesis*. Oxford University Press, 17(4), pp. 337–344. doi: 10.1093/mutage/17.4.337.

Deming, P. B. *et al.* (2001) 'The human decatenation checkpoint', *Proceedings of the National Academy of Sciences*, 98(21), pp. 12044–12049. doi: 10.1073/pnas.221430898.

Deming, P. B. *et al.* (2002) 'ATR Enforces the Topoisomerase II-dependent G 2 Checkpoint through Inhibition of Plk1 Kinase', *Journal of Biological Chemistry*, 277(39), pp. 36832–36838. doi: 10.1074/jbc.M206109200.

Deng, D. *et al.* (2012) 'Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors', *Science*, 335(6069), pp. 720–723. doi: 10.1126/science.1215670.

Dewey, W. C., Ling, C. C. and Meyn, R. E. (1995) 'Radiation-induced apoptosis: Relevance to radiotherapy', *International Journal of Radiation Oncology*Biology*Physics*, 33(4), pp. 781–796. doi: 10.1016/0360-3016(95)00214-8.

Diffley, J. F. *et al.* (2000) 'Coping with and recovering from hydroxyurea-induced replication fork arrest in budding yeast.', *Cold Spring Harbor symposia on quantitative biology*, 65, pp. 333–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12760047 (Accessed: 30 October 2017).

Doudna, J. A. and Charpentier, E. (2014) 'The new frontier of genome engineering with CRISPR-Cas9', *Science*, 346(6213), pp. 1258096–1258096. doi: 10.1126/science.1258096.

Durrbaum, M. and Storchova, Z. (2015) 'Chromosomal Instability in Cancer Cells', *Chromosomal Instability in Cancer Cells*. doi: 10.1007/978-3-319-20291-4.

Fabre, F. *et al.* (2002) 'Alternate pathways involving Sgs1/Top3, Mus81/ Mms4, and Srs2 prevent formation of toxic recombination intermediates from single-stranded gaps created by DNA replication', *Proceedings of the National Academy of Sciences*, 99(26), pp. 16887–16892. doi: 10.1073/pnas.252652399.

Falck, J., Coates, J. and Jackson, S. P. (2005) 'Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage', *Nature*, 434(7033), pp. 605–611. doi: 10.1038/nature03442.

Farajollahi, S. and Maas, S. (2010) 'Molecular diversity through RNA editing: a balancing act', *Trends in Genetics*, 26(5), pp. 221–230. doi: 10.1016/j.tig.2010.02.001.

Franchitto, A., Oshima, J. and Pichierri, P. (2003) 'The G2-phase decatenation

checkpoint is defective in Werner syndrome cells.', *Cancer research*, 63(12), pp. 3289–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12810661 (Accessed: 10 October 2017).

Friedberg, E. C. (2003) 'DNA damage and repair', *Nature*. Nature Publishing Group, 421(6921), pp. 436–440. doi: 10.1038/nature01408.

Friedberg, E. C. (2008) 'A brief history of the DNA repair field', *Cell Research*. Nature Publishing Group, 18(1), pp. 3–7. doi: 10.1038/cr.2007.113.

Gaj, T., Gersbach, C. A. and Barbas, C. F. (2013) 'ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering', *Trends in Biotechnology*, 31(7), pp. 397–405. doi: 10.1016/j.tibtech.2013.04.004.

Ganem, N. J., Godinho, S. A. and Pellman, D. (2009) 'A mechanism linking extra centrosomes to chromosomal instability.', *Nature*. Nature Publishing Group. doi: 10.1038/nature08136.

Gaudelli, N. M. *et al.* (2017) 'Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage', *Nature*. Nature Publishing Group, pp. 1–27. doi: 10.1038/nature24644.

Geigl, J. B. *et al.* (2008) 'Defining "chromosomal instability"', *Trends in Genetics*, 24(2), pp. 64–69. doi: 10.1016/j.tig.2007.11.006.

Gelsomino, F. *et al.* (2016) 'The evolving role of microsatellite instability in colorectal cancer: A review', *Cancer Treatment Reviews*. Elsevier Ltd, 51, pp. 19–26. doi: 10.1016/j.ctrv.2016.10.005.

Gemoll, T. *et al.* (2013) 'Chromosomal aneuploidy affects the global proteome equilibrium of colorectal cancer cells.', *Analytical cellular pathology (Amsterdam)*. Hindawi, 36(5–6), pp. 149–61. doi: 10.3233/ACP-140088.

Ghadimi, B. M. and Ried, T. (eds) (2015) *Chromosomal Instability in Cancer Cells*. Cham: Springer International Publishing (Recent Results in Cancer Research). doi: 10.1007/978-3-319-20291-4.

Ghadimi, B. M. and Ried, T. (no date a) *Chromosomal instability in cancer cells*.

Ghadimi, B. M. and Ried, T. (no date b) *Chromosomal instability in cancer cells*.

Giam, M. and Rancati, G. (2015) 'Aneuploidy and chromosomal instability in cancer: a jackpot to chaos.', *Cell division*. ??? doi: 10.1186/s13008-015-0009-7.

Giannattasio, M. *et al.* (2004) 'Physical and functional interactions between nucleotide excision repair and DNA damage checkpoint', *The EMBO Journal*, 23(2), pp. 429–438. doi: 10.1038/sj.emboj.7600051.

Gisselsson, D. *et al.* (2001) 'Abnormal Nuclear Shape in Solid Tumors Reflects Mitotic Instability', *The American Journal of Pathology*, 158(1), pp. 199–206. doi: 10.1016/S0002-9440(10)63958-2.

Gisselsson, D. (2008) 'Classification of chromosome segregation errors in cancer', *Chromosoma*, 117(6), pp. 511–519. doi: 10.1007/s00412-008-0169-1.

Gollin, S. M. (2005) 'Mechanisms leading to chromosomal instability', *Seminars in Cancer Biology*, 15(1), pp. 33–42. doi: 10.1016/j.semcancer.2004.09.004.

Gondek, L. P. *et al.* (2007) 'Single nucleotide polymorphism arrays complement metaphase cytogenetics in detection of new chromosomal lesions in MDS', *Leukemia*. Nature Publishing Group, 21(9), pp. 2058–2061. doi: 10.1038/sj.leu.2404745.

Gordon, D. J., Resio, B. and Pellman, D. (2012) 'Causes and consequences of aneuploidy in cancer', *Nature Reviews Genetics*. Nature Publishing Group, 13(3), pp. 189–203. doi: 10.1038/nrg3123.

Gottesfeld, J. M. *et al.* (1997) 'Regulation of gene expression by small molecules', *Nature*, 387(6629), pp. 202–205. doi: 10.1038/387202a0.

Granetto, C. *et al.* (1996) 'p53 accumulates in micronuclei after treatment with a DNA breaking chemical, methylnitrosourea, and with the spindle poison, vinblastine', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 352(1–2), pp. 61–64. doi: 10.1016/0027-5107(95)00235-9.

Haaf, T. *et al.* (1999) 'Sequestration of mammalian Rad51-recombination protein into micronuclei.', *The Journal of cell biology*. Rockefeller University Press, 144(1), pp. 11–20. doi: 10.1083/JCB.144.1.11.

Hanahan, D. and Weinberg, R. A. (2011) 'Review Hallmarks of Cancer : The Next Generation', *Cell*. Elsevier Inc., 144(5), pp. 646–674. doi: 10.1016/j.cell.2011.02.013.

Hanks, S. *et al.* (2004) 'Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in BUB1B', *Nature Genetics*. Nature Publishing Group, 36(11), pp. 1159–1161. doi: 10.1038/ng1449.

Harris, R. S. *et al.* (2003) 'DNA deamination mediates innate immunity to retroviral

infection', *Cell*, 113(6), pp. 803–809. doi: 10.1016/S0092-8674(03)00423-9.

Harris, R. S. and Liddament, M. T. (2004) 'Retroviral restriction by APOBEC proteins', *Nature Reviews Immunology*. Nature Publishing Group, 4(11), pp. 868–877. doi: 10.1038/nri1489.

Harris, R. S., Petersen-Mahrt, S. K. and Neuberger, M. S. (2002) 'RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators.', *Molecular cell*, 10(5), pp. 1247–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12453430 (Accessed: 8 October 2017).

Hatch, E. M. *et al.* (2013) 'Catastrophic nuclear envelope collapse in cancer cell micronuclei.', *Cell*. NIH Public Access, 154(1), pp. 47–60. doi: 10.1016/j.cell.2013.06.007.

Heinrichs, S. and Look, A. T. (2007) 'Identification of structural aberrations in cancer by SNP array analysis', *Genome Biology*, 8(7), p. 219. doi: 10.1186/gb-2007-8-7-219.

Helleday, T., Eshtad, S. and Nik-Zainal, S. (2014) 'Mechanisms underlying mutational signatures in human cancers', *Nature Reviews Genetics*. Nature Research, 15(9), pp. 585–598. doi: 10.1038/nrg3729.

Henderson, S. *et al.* (2014) 'APOBEC-Mediated Cytosine Deamination Links PIK3CA Helical Domain Mutations to Human Papillomavirus-Driven Tumor Development', *Cell Reports*, 7(6), pp. 1833–1841. doi: 10.1016/j.celrep.2014.05.012.

Hendriks, G. *et al.* (2010) 'Transcription-Dependent Cytosine Deamination Is a Novel Mechanism in Ultraviolet Light-Induced Mutagenesis', *Current Biology*, 20(2), pp. 170–175. doi: 10.1016/j.cub.2009.11.061.

Herbert, A. and Rich, A. (2001) 'The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1', *Proceedings of the National Academy of Sciences*, 98(21), pp. 12132–12137. doi: 10.1073/pnas.211419898.

Hirano, T. (2006) 'At the heart of the chromosome: SMC proteins in action', *Nature Reviews Molecular Cell Biology*, 7(5), pp. 311–322. doi: 10.1038/nrm1909.

Hoeijmakers, J. H. J. (2001) 'Genome maintenance mechanisms for preventing cancer.', *Nature*, 411(6835), pp. 366–374. doi: 10.1038/35077232.

Holland, A. J. and Cleveland, D. W. (2012) 'Chromoanagenesis and cancer: mechanisms and consequences of localized, complex chromosomal rearrangements.', *Nature*

medicine. doi: 10.1038/nm.2988.

Hough, R. F. and Bass, B. L. (1997) 'Analysis of Xenopus dsRNA adenosine deaminase cDNAs reveals similarities to DNA methyltransferases.', *RNA (New York, N.Y.)*, 3(4), pp. 356–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9085843 (Accessed: 12 October 2017).

Huang, Y. *et al.* (2011) 'The fate of micronucleated cells post X-irradiation detected by live cell imaging', *DNA Repair*, 10(6), pp. 629–638. doi: 10.1016/j.dnarep.2011.04.010.

Huang, Y. *et al.* (2012) 'Lagging chromosomes entrapped in micronuclei are not "lost" by cells', *Cell Research*, 22(5), pp. 932–935. doi: 10.1038/cr.2012.26.

Imre Berger, ‡,§ *et al.* (1998) 'Spectroscopic Characterization of a DNA-Binding Domain, Zα, from the Editing Enzyme, dsRNA Adenosine Deaminase: Evidence for Left-Handed Z-DNA in the Zα−DNA Complex†'. American Chemical Society . doi: 10.1021/BI9813126.

Ishkanian, A. S. *et al.* (2004) 'A tiling resolution DNA microarray with complete coverage of the human genome', *Nature Genetics*, 36(3), pp. 299–303. doi: 10.1038/ng1307.

Iwaizumi, M. *et al.* (no date) 'Human Sgo1 downregulation leads to chromosomal instability in colorectal cancer'. doi: 10.1136/gut.2008.149468.

Jackson, A. L., Chen, R. and Loeb, L. A. (1998) 'Induction of microsatellite instability by oxidative DNA damage', *Proceedings of the National Academy of Sciences*, 95(21), pp. 12468–12473. doi: 10.1073/pnas.95.21.12468.

Jackson, S. P. and Bartek, J. (2009) 'The DNA-damage response in human biology and disease', *Nature*, 461(7267), pp. 1071–1078. doi: 10.1038/nature08467.

Janssen, A. *et al.* (2011) 'Chromosome Segregation Errors as a Cause of DNA Damage and Structural Chromosome Aberrations', *Science*, 333(6051), pp. 1895–1898. doi: 10.1126/science.1210214.

Jao, L.-E., Wente, S. R. and Chen, W. (2013) 'Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system', *Proceedings of the National Academy of Sciences*, 110(34), pp. 13904–13909. doi: 10.1073/pnas.1308335110.

Jarmuz, A. *et al.* (2002) 'An Anthropoid-Specific Locus of Orphan C to U RNA-Editing Enzymes on Chromosome 22', *Genomics*, 79(3), pp. 285–296. doi:

10.1006/geno.2002.6718.

Jefford, C. E. and Irminger-Finger, I. (2006) 'Mechanisms of chromosome instability in cancers', *Critical Reviews in Oncology/Hematology*, 59(1), pp. 1–14. doi: 10.1016/j.critrevonc.2006.02.005.

Jinek, M. *et al.* (2012) 'A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity', *Science*, 337(6096), pp. 816–821. doi: 10.1126/science.1225829.

Jiricny, J. (2006) 'The multifaceted mismatch-repair system', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 7(5), pp. 335–346. doi: 10.1038/nrm1907.

John Savage (2000) 'atlasgeneticsoncology'. Available at: http://atlasgeneticsoncology.org/Deep/MicronucleiID20016.html (Accessed: 19 October 2017).

Jones, L. E. *et al.* (2009) 'Differential effects of reactive nitrogen species on DNA base excision repair initiated by the alkyladenine DNA glycosylase', *Carcinogenesis*, 30(12), pp. 2123–2129. doi: 10.1093/carcin/bgp256.

Kanu, N. *et al.* (2016) 'DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer.', *Genome biology*. BioMed Central, 17(1), p. 185. doi: 10.1186/s13059-016-1042-9.

Keegan, L. P., Gallo, A. and O'Connell, M. A. (2001) 'The many roles of an RNA editor.', *Nature Reviews Genetics*, 2(11), pp. 869–878. doi: 10.1038/35098584.

Kim, Y. *et al.* (2013) 'A library of TAL effector nucleases spanning the human genome', *Nature Biotechnology*, 31(3), pp. 251–258. doi: 10.1038/nbt.2517.

Kim, Y. G., Cha, J. and Chandrasegaran, S. (1996) 'Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 93(3), pp. 1156–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8577732 (Accessed: 14 October 2017).

Kirsch-Volders, M. *et al.* (2011) 'The in vitro MN assay in 2011: origin and fate, biological significance, protocols, high throughput methodologies and toxicological relevance', *Archives of Toxicology*. Springer-Verlag, 85(8), pp. 873–899. doi: 10.1007/s00204-011-0691-4.

Klein, I. A. *et al.* (2011) 'Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes.', *Cell*. Elsevier, 147(1), pp. 95–106. doi: 10.1016/j.cell.2011.07.048.

Koito, A. and Ikeda, T. (2012) 'Apolipoprotein B mRNA-editing, catalytic polypeptide cytidine deaminases and retroviral restriction', *Wiley Interdisciplinary Reviews: RNA*. John Wiley & Sons, Inc., 3(4), pp. 529–541. doi: 10.1002/wrna.1117.

Komor, A. C. *et al.* (2016) 'Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage', *Nature*. Nature Publishing Group, 533(7603), pp. 420–424. doi: 10.1038/nature17946.

Komor, A. C. *et al.* (2017) 'Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity', *Science Advances*, 3(8), p. eaao4774. doi: 10.1126/sciadv.aao4774.

Krokan, H. E. and Bjørås, M. (2013) 'Base excision repair.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 5(4), p. a012583. doi: 10.1101/cshperspect.a012583.

Kuo, L. J. and Yang, L.-X. (no date) 'Gamma-H2AX - a novel biomarker for DNA double-strand breaks.', *In vivo (Athens, Greece)*, 22(3), pp. 305–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18610740 (Accessed: 19 October 2017).

De La Chapelle, A. and Hampel, H. (2010) 'Clinical relevance of microsatellite instability in colorectal cancer', *Journal of Clinical Oncology*, 28(20), pp. 3380–3387. doi: 10.1200/JCO.2009.27.0652.

Lada, A. G. *et al.* (2011) 'Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast.', *Biochemistry. Biokhimiia*, 76(1), pp. 131–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21568845 (Accessed: 8 October 2017).

Lammens, K. *et al.* (2011) 'The Mre11:Rad50 structure shows an ATP-dependent molecular clamp in DNA double-strand break repair.', *Cell*. Elsevier, 145(1), pp. 54–66. doi: 10.1016/j.cell.2011.02.038.

Law, E. K. *et al.* (2016) 'The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer'. Available at: http://advances.sciencemag.org/content/advances/2/10/e1601737.full.pdf (Accessed: 8 October 2017).

LEE, H.-C. *et al.* (2005) 'Mitochondrial Genome Instability and mtDNA Depletion in Human Cancers', *Annals of the New York Academy of Sciences*. Blackwell Publishing Ltd, 1042(1), pp. 109–122. doi: 10.1196/annals.1338.011.

Lee, H.-S. *et al.* (2013) 'A new assay for measuring chromosome instability (CIN) and identification of drugs that elevate CIN in cancer cells', *BMC Cancer*. BioMed Central, 13(1), p. 252. doi: 10.1186/1471-2407-13-252.

LEHMANN, A. *et al.* (2007) 'Translesion synthesis: Y-family polymerases and the polymerase switch', *DNA Repair*, 6(7), pp. 891–899. doi: 10.1016/j.dnarep.2007.02.003.

Leibowitz, M. L., Zhang, C.-Z. and Pellman, D. (2015) 'Chromothripsis: A New Mechanism for Rapid Karyotype Evolution.', *Annual review of genetics*. doi: 10.1146/annurev-genet-120213-092228.

Lengauer, C., Kinzler, K. W. and Vogelstein, B. (1998) 'Genetic instabilities in human cancers.', *Nature*, 396(6712), pp. 643–649. doi: 10.1038/25292.

Lettier, G. *et al.* (2006) 'The Role of DNA Double-Strand Breaks in Spontaneous Homologous Recombination in S. cerevisiae', *PLoS Genetics*, 2(11), p. e194. doi: 10.1371/journal.pgen.0020194.

Li, G.-M. (2008) 'Mechanisms and functions of DNA mismatch repair', *Cell Research*, 18(1), pp. 85–98. doi: 10.1038/cr.2007.115.

Lindahl, T. (1993) 'Instability and decay of the primary structure of DNA', *Nature*. Nature Publishing Group, 362(6422), pp. 709–715. doi: 10.1038/362709a0.

Lindahl, T. and Barnes, D. E. (2000) 'Repair of endogenous DNA damage.', *Cold Spring Harbor symposia on quantitative biology*, 65, pp. 127–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12760027 (Accessed: 7 October 2017).

Liu, M. and Schatz, D. G. (2009) 'Balancing AID and DNA repair during somatic hypermutation', *Trends in Immunology*, 30(4), pp. 173–181. doi: 10.1016/j.it.2009.01.007.

Loeb, L. A., Springgate, C. F. and Battula, N. (1974) 'Errors in DNA replication as a basis of malignant changes.', *Cancer research*, 34(9), pp. 2311–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4136142 (Accessed: 8 October 2017).

Macbeth, M. R. *et al.* (2005) 'Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing.', *Science (New York, N.Y.)*. NIH Public Access, 309(5740), pp.

1534–9. doi: 10.1126/science.1113150.

Maciejowski, J. *et al.* (2015) 'Chromothripsis and Kataegis Induced by Telomere Crisis', *Cell*. Elsevier Ltd. doi: 10.1016/j.cell.2015.11.054.

Mak, A. N.-S. *et al.* (2012) 'The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target', *Science*, 335(6069), pp. 716–719. doi: 10.1126/science.1216211.

Mali, P. *et al.* (2013) 'RNA-guided human genome engineering via Cas9.', *Science (New York, N.Y.)*. NIH Public Access, 339(6121), pp. 823–6. doi: 10.1126/science.1232033.

Mangeat, B. *et al.* (2003) 'Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts', *Nature*, 424(6944), pp. 99–103. doi: 10.1038/nature01709.

Markossian, S. *et al.* (2016) 'Quantitative assessment of chromosome instability induced through chemical disruption of mitotic progression', *Cell Cycle*. Taylor & Francis, 15(13), pp. 1706–1714. doi: 10.1080/15384101.2016.1175796.

Matsuoka, S. *et al.* (2007) 'ATM and ATR Substrate Analysis Reveals Extensive Protein Networks Responsive to DNA Damage', *Science*, 316(5828), pp. 1160–1166. doi: 10.1126/science.1140321.

Matthews, M. M. *et al.* (2016) 'Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity', *Nature Structural & Molecular Biology*, 23(5), pp. 426–433. doi: 10.1038/nsmb.3203.

McClintock, B. (1941) 'The Stability of Broken Ends of Chromosomes in Zea Mays.', *Genetics*. Genetics Society of America, 26(2), pp. 234–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17247004 (Accessed: 5 October 2017).

McGranahan, N. *et al.* (2012) 'Cancer chromosomal instability: therapeutic and diagnostic challenges.', *EMBO reports*. Nature Publishing Group. doi: 10.1038/embor.2012.61.

Meaburn, K. J., Parris, C. N. and Bridger, J. M. (2005) 'The manipulation of chromosomes by mankind: the uses of microcell-mediated chromosome transfer', *Chromosoma*. Springer-Verlag, 114(4), pp. 263–274. doi: 10.1007/s00412-005-0014-8.

Meeker, A. K. *et al.* (2002) 'Telomere Length Assessment in Human Archival Tissues', *The American Journal of Pathology*, 160(4), pp. 1259–1268. doi: 10.1016/S0002-9440(10)62553-9.

Meng, F.-L. *et al.* (2014) 'Convergent Transcription at Intragenic Super-Enhancers Targets AID-Initiated Genomic Instability', *Cell*, 159(7), pp. 1538–1548. doi: 10.1016/j.cell.2014.11.014.

Mercer, A. C. *et al.* (2012) 'Chimeric TALE recombinases with programmable DNA sequence specificity', *Nucleic Acids Research*, 40(21), pp. 11163–11172. doi: 10.1093/nar/gks875.

Merlo, L. M. F. *et al.* (2006) 'Cancer as an evolutionary and ecological process', *Nature Reviews Cancer*, 6(12), pp. 924–935. doi: 10.1038/nrc2013.

Miller, J. C. *et al.* (2011) 'A TALE nuclease architecture for efficient genome editing', *Nature Biotechnology*, 29(2), pp. 143–148. doi: 10.1038/nbt.1755.

Minocherhomji, S. *et al.* (2015) 'Replication stress activates DNA repair synthesis in mitosis', *Nature*. Nature Research, 528(7581), pp. 286–290. doi: 10.1038/nature16139.

Mitelman, F., Mertens, F. and Johansson, B. (1997) 'A breakpoint map of recurrent chromosomal rearrangements in human neoplasia', *Nature Genetics*, 15(4), pp. 417–474. doi: 10.1038/ng0497supp-417.

Morganella, S. *et al.* (2016) 'The topography of mutational processes in breast cancer genomes', *Nature Communications*. Nature Publishing Group, 7, p. 11383. doi: 10.1038/ncomms11383.

Müller, A. and Fishel, R. (2002) 'Mismatch repair and the hereditary non-polyposis colorectal cancer syndrome (HNPCC).', *Cancer investigation*, 20(1), pp. 102–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11852992 (Accessed: 10 October 2017).

Muramatsu, M. *et al.* (2000) 'Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme.', *Cell*, 102(5), pp. 553–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11007474 (Accessed: 8 October 2017).

Nakano, M. *et al.* (2008) 'Inactivation of a Human Kinetochore by Specific Targeting of Chromatin Modifiers', *Developmental Cell*. Elsevier Inc. doi: 10.1016/j.devcel.2008.02.001.

Natarajan, A. T. and Palitti, F. (2008) 'DNA repair and chromosomal alterations', *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 657(1), pp. 3–7.

doi: 10.1016/j.mrgentox.2008.08.017.

Navaratnam, N. *et al.* (1993) 'The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase', *Journal of Biological Chemistry*, 268(28), pp. 20709–20712.

Negrini, S., Gorgoulis, V. G. and Halazonetis, T. D. (2010) 'Genomic instability — an evolving hallmark of cancer', *Nature Reviews Molecular Cell Biology*, 11(3), pp. 220–228. doi: 10.1038/nrm2858.

Nik-Zainal, S. *et al.* (2016) 'Landscape of somatic mutations in 560 breast cancer whole-genome sequences', *Nature*. Nature Research, 534(7605), pp. 47–54. doi: 10.1038/nature17676.

Nishida, K. *et al.* (2016) 'Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems', *Science*, 353(6305), p. aaf8729-aaf8729. doi: 10.1126/science.aaf8729.

Nishikura, K. (2010) 'Functions and Regulation of RNA Editing by ADAR Deaminases', *Annual Review of Biochemistry*, 79(1), pp. 321–349. doi: 10.1146/annurev-biochem-060208-105251.

Nishizaki, T. *et al.* (1997) 'Genetic alterations in lobular breast cancer by comparative genomic hybridization.', *International journal of cancer*, 74(5), pp. 513–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9355973 (Accessed: 5 October 2017).

Di Noia, J. and Neuberger, M. S. (2002) 'Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase', *Nature*, pp. 43–48. doi: 10.1038/nature00981.

Di Noia, J. M. and Neuberger, M. S. (2007) 'Molecular Mechanisms of Antibody Somatic Hypermutation', *Annual Review of Biochemistry*, 76(1), pp. 1–22. doi: 10.1146/annurev.biochem.76.061705.090740.

Nordentoft, I. *et al.* (2014) 'Mutational Context and Diverse Clonal Development in Early and Late Bladder Cancer', *Cell Reports*, 7(5), pp. 1649–1663. doi: 10.1016/j.celrep.2014.04.038.

Nowak, M. A. *et al.* (2002) 'The role of chromosomal instability in tumor initiation', *Proceedings of the National Academy of Sciences*, 99(25), pp. 16226–16231. doi: 10.1073/pnas.202617399.

Nowell, P. C. (1976) 'The clonal evolution of tumor cell populations.', *Science (New York, N.Y.)*, 194(4260), pp. 23–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/959840 (Accessed: 8 October 2017).

Olivier, M., Hollstein, M. and Hainaut, P. (2010) 'TP53 mutations in human cancers: origins, consequences, and clinical use.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 2(1), p. a001008. doi: 10.1101/cshperspect.a001008.

Oromendia, A. B. and Amon, A. (2014) 'Aneuploidy: implications for protein homeostasis and disease', *Disease Models & Mechanisms*, 7(1), pp. 15–20. doi: 10.1242/dmm.013391.

Le Page, F. *et al.* (2005) 'Transcription-coupled repair of 8-oxoguanine: requirement for XPG, TFIIH, and CSB and implications for Cockayne syndrome.', *Cell*, 123(4), p. 711. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16329169 (Accessed: 10 October 2017).

Pauklin, S. *et al.* (2009) 'Estrogen directly activates AID transcription and function', *The Journal of Experimental Medicine*, 206(1), pp. 99–111. doi: 10.1084/jem.20080521.

Peltomaki, P. (2001) 'Deficient DNA mismatch repair: a common etiologic factor for colon cancer', *Human Molecular Genetics*. Oxford University Press, 10(7), pp. 735–740. doi: 10.1093/hmg/10.7.735.

Petersen-Mahrt, S. K., Harris, R. S. and Neuberger, M. S. (2002) 'AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification', *Nature*. Nature Publishing Group, 418(6893), pp. 99–104. doi: 10.1038/nature00862.

Petersen, S. *et al.* (2001) 'AID is required to initiate Nbs1/γ-H2AX focus formation and mutations at sites of class switching', *Nature*, 414(6864), pp. 660–665. doi: 10.1038/414660a.

Pfeifer, G. P., You, Y.-H. and Besaratinia, A. (2005) 'Mutations induced by ultraviolet light', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 571(1–2), pp. 19–31. doi: 10.1016/j.mrfmmm.2004.06.057.

Pikor, L. *et al.* (2013) 'The detection and implication of genome instability in cancer', *Cancer and Metastasis Reviews*, 32(3–4), pp. 341–352. doi: 10.1007/s10555-013-9429-5.

Pinkel, D. *et al.* (1998) 'High resolution analysis of DNA copy number variation using

comparative genomic hybridization to microarrays.', *Nature Genetics*, 20(2), pp. 207–211. doi: 10.1038/2524.

Pollack, J. R. *et al.* (2002) 'Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors', *Proceedings of the National Academy of Sciences*, 99(20), pp. 12963–12968. doi: 10.1073/pnas.162471999.

Popp, H. D. and Bohlander, S. K. (2010) 'Genetic Instability in Inherited and Sporadic Leukemias', *GENES, CHROMOSOMES & CANCER*, 49, pp. 1071–1081. doi: 10.1002/gcc.20823.

Qian, J. *et al.* (2014) 'B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity.', *Cell*. NIH Public Access, 159(7), pp. 1524–37. doi: 10.1016/j.cell.2014.11.013.

Ramiro, A. R. *et al.* (2004) 'AID Is Required for c-myc/IgH Chromosome Translocations In Vivo', *Cell*, 118(4), pp. 431–438. doi: 10.1016/j.cell.2004.08.006.

Ramiro, A. R. *et al.* (2006) 'Role of genomic instability and p53 in AID-induced c-myc–Igh translocations', *Nature*. Nature Publishing Group, 440(7080), pp. 105–109. doi: 10.1038/nature04495.

Ramiro, A. R., Nussenzweig, M. C. and Nussenzweig, A. (2006) 'Switching on chromosomal translocations', *Cancer Research*, pp. 7837–7839. doi: 10.1158/0008-5472.CAN-06-0863.

Rao, C. V. and Yamada, H. Y. (2013) 'Genomic Instability and Colon Carcinogenesis: From the Perspective of Genes', *Frontiers in Oncology*, 3(May), pp. 1–13. doi: 10.3389/fonc.2013.00130.

Rassool, F. V. (2003) 'DNA double strand breaks (DSB) and non-homologous end joining (NHEJ) pathways in human leukemia.', *Cancer letters*, 193(1), pp. 1–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12691817 (Accessed: 10 October 2017).

Reyon, D. *et al.* (2012) 'FLASH assembly of TALENs for high-throughput genome editing', *Nature Biotechnology*, 30(5), pp. 460–465. doi: 10.1038/nbt.2170.

'RNA editing in brain controls a determinant of ion flow in glutamate-gated channels' (1991) *Cell*. Cell Press, 67(1), pp. 11–19. doi: 10.1016/0092-8674(91)90568-J.

Roberts, S. A. *et al.* (2012) 'Clustered Mutations in Yeast and in Human Cancers Can

Arise from Damaged Long Single-Strand DNA Regions', *Molecular Cell*, 46(4), pp. 424–435. doi: 10.1016/j.molcel.2012.03.030.

Roger, A. S. *et al.* (2006) *DNA Repair and Mutagenesis, Second Edition*. American Society of Microbiology. doi: 10.1128/9781555816704.

Rogozin, I. B. *et al.* (2007) 'Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase', *Nature Immunology*, 8(6), pp. 647–656. doi: 10.1038/ni1463.

Roschke, A. V. *et al.* (2003) 'Karyotypic Complexity of the NCI-60 Drug-Screening Panel', *Cancer Research*, 63(24), pp. 8634–8647.

Rosenberg, B. R. *et al.* (2011) 'Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs', *Nature Structural & Molecular Biology*, 18(2), pp. 230–236. doi: 10.1038/nsmb.1975.

Roth, D. B. and Wilson, J. H. (1986) 'Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction.', *Molecular and cellular biology*. American Society for Microbiology (ASM), 6(12), pp. 4295–304. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3025650 (Accessed: 10 October 2017).

Russo, M. T. *et al.* (2004) 'Accumulation of the Oxidative Base Lesion 8-Hydroxyguanine in DNA of Tumor-Prone Mice Defective in Both the ***Myh*** and ***Ogg1*** DNA Glycosylases', *Cancer Research*, 64(13), pp. 4411–4414. doi: 10.1158/0008-5472.CAN-04-0355.

Sambrook, J. and Russell, D. W. (David W. (2001) *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory Press.

Sancar, A. *et al.* (2004) 'Molecular Mechanisms of Mammalian DNA Repair and the DNA Damage Checkpoints', *Annual Review of Biochemistry*, 73(1), pp. 39–85. doi: 10.1146/annurev.biochem.73.011303.073723.

Sansregret, L. and Swanton, C. (2017) 'The role of aneuploidy in cancer evolution', *Cold Spring Harbor Perspectives in Medicine*, 7(1), pp. 1–18. doi: 10.1101/cshperspect.a028373.

Sato, Y. *et al.* (2010) 'Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy', *Journal of Biological Chemistry*, 285(10), pp. 7111–7118. doi: 10.1074/jbc.M109.052977.

Savage, J. R. (1988) ‘A comment on the quantitative relationship between micronuclei and chromosomal aberrations.’, *Mutation research*, 207(1), pp. 33–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3336377 (Accessed: 5 October 2017).

Savva, Y. A., Rieder, L. E. and Reenan, R. A. (2012) ‘The ADAR protein family’, *Genome Biology*. BioMed Central, 13(12), p. 252. doi: 10.1186/gb-2012-13-12-252.

Schmid-Burgk, J. L. *et al.* (2012) ‘A ligation-independent cloning technique for high-throughput assembly of transcription activator–like effector genes’, *Nature Biotechnology*, 31(1), pp. 76–81. doi: 10.1038/nbt.2460.

Schoch, C. *et al.* (2005) ‘Genomic gains and losses influence expression levels of genes located within the affected regions: a study on acute myeloid leukemias with trisomy 8, 11, or 13, monosomy 7, or deletion 5q’, *Leukemia*, 19(7), pp. 1224–1228. doi: 10.1038/sj.leu.2403810.

Schwartz, T. *et al.* (1999) ‘Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA.’, *Science (New York, N.Y.)*, 284(5421), pp. 1841–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10364558 (Accessed: 12 October 2017).

Severi, F., Chicca, A. and Conticello, S. G. (2011a) ‘Analysis of Reptilian APOBEC1 Suggests that RNA Editing May Not Be Its Ancestral Function’, *Molecular Biology and Evolution*, 28(3), pp. 1125–1129. doi: 10.1093/molbev/msq338.

Severi, F., Chicca, A. and Conticello, S. G. (2011b) ‘Analysis of Reptilian APOBEC1 Suggests that RNA Editing May Not Be Its Ancestral Function’, *Molecular Biology and Evolution*, 28(3), pp. 1125–1129. doi: 10.1093/molbev/msq338.

Sheehy, A. M. *et al.* (2002) ‘Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein’, *Nature*, 418(6898), pp. 646–650. doi: 10.1038/nature00939.

Shimizu, N., Misaka, N. and Utani, K. (2007) ‘Nonselective DNA damage induced by a replication inhibitor results in the selective elimination of extrachromosomal double minutes from human cancer cells’, *Genes, Chromosomes and Cancer*. Wiley Subscription Services, Inc., A Wiley Company, 46(10), pp. 865–874. doi: 10.1002/gcc.20473.

Shuck, S. C., Short, E. A. and Turchi, J. J. (2008) ‘Eukaryotic nucleotide excision repair:

from understanding mechanisms to influencing biology', *Cell Research*, 18(1), pp. 64–72. doi: 10.1038/cr.2008.2.

Sjögren, C. and Nasmyth, K. (2001) 'Sister chromatid cohesion is required for postreplicative double-strand break repair in Saccharomyces cerevisiae.', *Current biology : CB*, 11(12), pp. 991–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11448778 (Accessed: 10 October 2017).

Smith, H. C. *et al.* (2012) 'Functions and regulation of the APOBEC family of proteins', *Seminars in Cell & Developmental Biology*, 23(3), pp. 258–268. doi: 10.1016/j.semcdb.2011.10.004.

Sotillo, R. *et al.* (2007) 'Mad2 Overexpression Promotes Aneuploidy and Tumorigenesis in Mice', *Cancer Cell*, 11(1), pp. 9–23. doi: 10.1016/j.ccr.2006.10.019.

Soutoglou, E. *et al.* (2007) 'Positional stability of single double-strand breaks in mammalian cells.', *Nature cell biology*. NIH Public Access, 9(6), pp. 675–82. doi: 10.1038/ncb1591.

Stenglein, M. D. *et al.* (2010) 'APOBEC3 proteins mediate the clearance of foreign DNA from human cells', *Nature Structural & Molecular Biology*. Nature Research, 17(2), pp. 222–229. doi: 10.1038/nsmb.1744.

Stephens, O. M., Yi-Brunozzi, H. Y. and Beal, P. A. (2000) 'Analysis of the RNA-editing reaction of ADAR2 with structural and fluorescent analogues of the GluR-B R/G editing site.', *Biochemistry*, 39(40), pp. 12243–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11015203 (Accessed: 12 October 2017).

Stephens, P. J. *et al.* (2011) 'Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development', *Cell*, 144(1), pp. 27–40. doi: 10.1016/j.cell.2010.11.055.

Strobel, S. A. and Dervan, P. B. (1990) 'Site-specific cleavage of a yeast chromosome by oligonucleotide-directed triple-helix formation.', *Science (New York, N.Y.)*, 249(4964), pp. 73–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2195655 (Accessed: 14 October 2017).

Ström, L. *et al.* (2004) 'Postreplicative Recruitment of Cohesin to Double-Strand Breaks Is Required for DNA Repair', *Molecular Cell*, 16(6), pp. 1003–1015. doi: 10.1016/j.molcel.2004.11.026.

Stuart, K. D. *et al.* (2005) 'Complex management: RNA editing in trypanosomes.', *Trends in biochemical sciences*. Elsevier, 30(2), pp. 97–105. doi: 10.1016/j.tibs.2004.12.006.

Swanton, C. *et al.* (2015) 'APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity', *Cancer Discovery*, 5(7), pp. 704–712. doi: 10.1158/2159-8290.CD-15-0344.

Swift, L. H. and Golsteyn, R. M. (2014) 'Genotoxic anti-cancer agents and their relationship to DNA damage, mitosis, and checkpoint adaptation in proliferating cancer cells.', *International journal of molecular sciences*. Multidisciplinary Digital Publishing Institute (MDPI), 15(3), pp. 3403–31. doi: 10.3390/ijms15033403.

Takata, M. *et al.* (1998) 'Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells.', *The EMBO journal*. European Molecular Biology Organization, 17(18), pp. 5497–508. doi: 10.1093/emboj/17.18.5497.

Taylor, B. S. *et al.* (2010) 'Integrative Genomic Profiling of Human Prostate Cancer', *Cancer Cell*, 18(1), pp. 11–22. doi: 10.1016/j.ccr.2010.05.026.

Terradas, M. *et al.* (2009) 'DNA lesions sequestered in micronuclei induce a local defective-damage response', *DNA Repair*, 8(10), pp. 1225–1234. doi: 10.1016/j.dnarep.2009.07.004.

Terradas, M. *et al.* (2012) 'Is DNA Damage Response Ready for Action Anywhere?', *International Journal of Molecular Sciences*. Molecular Diversity Preservation International, 13(12), pp. 11569–11583. doi: 10.3390/ijms130911569.

Terradas, M., Mart??n, M. and Genesc??, A. (2016) 'Impaired nuclear functions in micronuclei results in genome instability and chromothripsis', *Archives of Toxicology*. Springer Berlin Heidelberg. doi: 10.1007/s00204-016-1818-4.

Terzoudi, G. I. *et al.* (2015) 'Stress induced by premature chromatin condensation triggers chromosome shattering and chromothripsis at DNA sites still replicating in micronuclei or multinucleate cells when primary nuclei enter mitosis', *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*. Elsevier Ltd. doi: 10.1016/j.mrgentox.2015.07.014.

Thomas, J. M. and Beal, P. A. (2017) 'How do ADARs bind RNA? New protein-RNA

structures illuminate substrate recognition by the RNA editing ADARs', *BioEssays*, 39(4), p. 1600187. doi: 10.1002/bies.201600187.

Thompson, S. L. and Compton, D. A. (2011) 'Chromosome missegregation in human cells arises through specific types of kinetochore-microtubule attachment errors.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(44), pp. 17974–8. doi: 10.1073/pnas.1109720108.

Tomasetti, C. and Vogelstein, B. (2015) 'Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions.', *Science (New York, N.Y.)*. NIH Public Access, 347(6217), pp. 78–81. doi: 10.1126/science.1260825.

Torres, E. (2015) 'Yeast as Models of Mitotic Fidelity', in, pp. 143–164. doi: 10.1007/978-3-319-20291-4_7.

Torres, E. M. *et al.* (2007) 'Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast', *Science*, 317(5840), pp. 916–924. doi: 10.1126/science.1142210.

Tsafrir, D. *et al.* (2006) 'Relationship of Gene Expression and Chromosomal Abnormalities in Colorectal Cancer', *Cancer Research*, 66(4), pp. 2129–2137. doi: 10.1158/0008-5472.CAN-05-2569.

Tsuruoka, N. *et al.* (2013) 'ADAR1 protein induces adenosine-targeted DNA mutations in senescent Bcl6 gene-deficient cells.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 288(2), pp. 826–36. doi: 10.1074/jbc.M112.365718.

Tubbs, A. and Nussenzweig, A. (2017) 'Endogenous DNA Damage as a Source of Genomic Instability in Cancer', *Cell*, 168(4), pp. 644–656. doi: 10.1016/j.cell.2017.01.002.

Uhlmann, F. and Nasmyth, K. (1998) 'Cohesion between sister chromatids must be established during DNA replication.', *Current biology : CB*, 8(20), pp. 1095–101. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9778527 (Accessed: 10 October 2017).

Utani, K., Okamoto, A. and Shimizu, N. (2011) 'Generation of Micronuclei during Interphase by Coupling between Cytoplasmic Membrane Blebbing and Nuclear Budding', *PLoS ONE*. Edited by J. M. Bridger. Public Library of Science, 6(11), p.

e27233. doi: 10.1371/journal.pone.0027233.

Venkatesan, S., Natarajan, A. T. and Hande, M. P. (2015) 'Chromosomal instability-mechanisms and consequences', *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*. Elsevier Ltd, 793, pp. 176–184. doi: 10.1016/j.mrgentox.2015.08.008.

Vermeulen, K., Van Bockstaele, D. R. and Berneman, Z. N. (2003) 'The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer.', *Cell proliferation*, 36(3), pp. 131–49. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12814430 (Accessed: 10 October 2017).

Vieira, V. C. and Soares, M. A. (2013) 'The Role of Cytidine Deaminases on Innate Immune Responses against Human Viral Infections', *BioMed Research International*, 2013, pp. 1–18. doi: 10.1155/2013/683095.

Vilardell, M. *et al.* (2011) 'Meta-analysis of heterogeneous Down Syndrome data reveals consistent genome-wide dosage effects related to neurological processes', *BMC Genomics*, 12(1), p. 229. doi: 10.1186/1471-2164-12-229.

Walker, J. A. *et al.* (1996) 'Chromosome content and ultrastructure of radiation-induced micronuclei.', *Mutagenesis*, 11(5), pp. 419–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8921502 (Accessed: 6 October 2017).

Wang, H. *et al.* (2003) 'Biochemical evidence for Ku-independent backup pathways of NHEJ.', *Nucleic acids research*. Oxford University Press, 31(18), pp. 5377–88. doi: 10.1093/NAR/GKG728.

Wang, H. *et al.* (2013) 'One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering', *Cell*, 153(4), pp. 910–918. doi: 10.1016/j.cell.2013.04.025.

Wedekind, J. E. *et al.* (no date) 'Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business'. doi: 10.1016/S0168-9525(03)00054-4.

Whitwell, J. *et al.* (2015) 'Relationships between p53 status, apoptosis and induction of micronuclei in different human and mouse cell lines in vitro: Implications for improving existing assays', *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 789–790, pp. 7–27. doi: 10.1016/j.mrgentox.2015.05.011.

Williams, G. J., Lees-Miller, S. P. and Tainer, J. A. (2010) 'Mre11-Rad50-Nbs1 conformations and the control of sensing, signaling, and effector responses at DNA double-strand breaks.', *DNA repair*. NIH Public Access, 9(12), pp. 1299–306. doi: 10.1016/j.dnarep.2010.10.001.

Yasui, M. *et al.* (2010) 'Live cell imaging of micronucleus formation and development', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 692(1–2), pp. 12–18. doi: 10.1016/j.mrfmmm.2010.07.009.

Yi-Brunozzi, H. Y., Stephens, O. M. and Beal, P. A. (2001) 'Conformational changes that occur during an RNA-editing adenosine deamination reaction.', *The Journal of biological chemistry*, 276(41), pp. 37827–33. doi: 10.1074/jbc.M106299200.

You, Z. *et al.* (2005) 'ATM Activation and Its Recruitment to Damaged DNA Require Binding to the C Terminus of Nbs1', *Molecular and Cellular Biology*, 25(13), pp. 5363–5379. doi: 10.1128/MCB.25.13.5363-5379.2005.

Zeitlin, S. G. *et al.* (2005) 'Xenopus CENP-A assembly into chromatin requires base excision repair proteins', *DNA Repair*, 4(7), pp. 760–772. doi: 10.1016/j.dnarep.2005.02.007.

Zhang, C.-Z. *et al.* (2015) 'Chromothripsis from DNA damage in micronuclei', *Nature*. doi: 10.1038/nature14493.

Zhang, F. *et al.* (2011) 'Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription', *Nature Biotechnology*, 29(2), pp. 149–153. doi: 10.1038/nbt.1775.

Zhang, N. *et al.* (2008) 'Overexpression of Separase induces aneuploidy and mammary tumorigenesis.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(35), pp. 13033–8. doi: 10.1073/pnas.0801610105.

Zhao, X. *et al.* (2004) 'An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays.', *Cancer research*, 64(9), pp. 3060–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15126342 (Accessed: 11 October 2017).

Zharkov, D. O. (2008) 'Base excision DNA repair', *Cellular and Molecular Life Sciences*, 65(10), pp. 1544–1565. doi: 10.1007/s00018-008-7543-2.

Zheng, Y., Lorenzo, C. and Beal, P. A. (2017) 'DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA', *Nucleic Acids Research*. Oxford University Press, 45(6), p. gkx050. doi: 10.1093/nar/gkx050.

Zong, C. *et al.* (2012) 'Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell', *Science*, 338(6114), pp. 1622–1626. doi: 10.1126/science.1229164.

7. A novel fluorescent reporter for quantification of APOBEC-Cas9 base editing in living cells (manuscript submitted)

A novel fluorescent reporter for quantification of APOBEC-Cas9 base editing in living cells

Amber St. Martin[1, †], Daniel Salamango[1, †], Nadine Shaban[1], William L. Brown[1], Francesco Donati[2], Uday Munagala[2], Silvestro G. Conticello[2], Reuben S. Harris[1,3,*]

[1] Department of Biochemistry, Molecular Biology and Biophysics, Masonic Cancer Center, Center for Genome Engineering, Institute for Molecular Virology, University of Minnesota, MN 55455, USA

[2] Istituto Toscano Tumori, Firenze 50134, Italy

[3] Howard Hughes Medical Institute, University of Minnesota, Minneapolis, MN 55455, USA

[*] To whom correspondence should be addressed. Tel. +1-612-624-0457; Fax: +1-612-625-2163; Email: rsh@umn.edu

[†] These authors contributed equally to the work as the first authors.

Keywords: APOBEC-Cas9; base editing; CRISPR; DNA base editing; editosome; genome engineering

ABSTRACT

Base editing is an exciting new application for genome engineering technology. C-to-T mutations in genomic DNA have been achieved using ribonucleoprotein complexes comprised of rat APOBEC1 single-stranded DNA deaminase, Cas9 nickase (Cas9n), uracil DNA glycosylase inhibitor (UGI), and guide (g)RNA. Here, we report the first real-time system for quantification of base editing in living human cells as well as next-generation editing constructs that achieve higher editing frequencies. Mutation of an APOBEC-preferred trinucleotide, 5'-TCA-to-TTA, restores mCherry fluorescence in a reporter marked by eGFP, and editing frequencies are quantified through ratios of mCherry-positive to eGFP-positive cells. Using this system as both an episomal and a chromosomal editing reporter, we show that human APOBEC3A and APOBEC3B base editing constructs are more efficient than a rat APOBEC1 construct. We also demonstrate an enrichment of editing events at a heterologous chromosomal locus in reporter-activated, mCherry-positive cells. The combination of a rapid, fluorescence-based base editing reporter system and more efficient, structurally defined DNA editing enzymes expands the versatility of this powerful new technology.

INTRODUCTION

APOBEC enzymes are single-stranded (ss) polynucleotide cytosine deaminases. Human cells encode nine active family members with AID functioning in antibody DNA diversification, APOBEC1 in mRNA editing, and APOBEC3A-H in DNA virus and

transposon restriction (1-4). APOBEC1 is also an efficient DNA mutator (5,6), and the rat enzyme was combined recently with Cas9 and guide (g)RNA to create ribonucleoprotein complexes capable of editing single cytosine nucleobases and making site-specific C-to-T mutations in genomic DNA (7). A construct comprised of rat APOBEC1, Cas9 nickase (Cas9n), and uracil DNA glycosylase inhibitor (UGI) has been shown to yield base editing frequencies ranging from 5-50% (BE3) (7-9). This editing complex has already been adopted by many labs and harnessed for biotechnology applications (10-14). Two orthologs, human AID and lamprey PmCDA1, have also been combined with Cas9n but with lower overall base editing efficiencies, likely due to lower intrinsic enzyme activities (15-18). PmCDA1 has also been used in plant genome engineering (17).

A significant impediment to optimizing base editing technologies and deployment in limitless cell types is a lack of an efficient, real-time, rapid, and quantitative editing assay (ideally one that is also transferrable across species and, at least initially, independent of DNA sequencing to assess efficiencies). Here, we report a fluorescence-based base editing reporter system for quantification of real-time editing in living mammalian cells. The system is called AMBER for APOBEC-mediated base editing reporter. AMBER is a bicistronic construct that expresses eGFP constitutively as a marker, and mutation of an APOBEC-preferred trinucleotide, 5'-TCA-to-TTA, reverts an essential amino acid and restores mCherry fluorescence. The ratio of mCherry-positive to eGFP-positive cells thereby enables rapid quantification of DNA editing frequencies by fluorescence microscopy or flow cytometry. The AMBER system was validated episomally in transient transfection experiments and chromosomally following stable

integration by lentivirus-mediated transduction. The AMBER system was used to develop highly efficient base editing constructs based upon APOBEC3A and APOBEC3B (catalytic domain), which like Cas9 are structurally defined enzymes. In addition, the AMBER system was used to enrich for cells with editing events at heterologous chromosomal sites. The success of these two applications demonstrates the power and utility of AMBER as a rapid, fluorescence-based base editing reporter system.

MATERIALS AND METHODS

Cell lines and culture conditions

293T cells were maintained in DMEM (Hyclone) supplemented with 10% FBS (Gibco) and 0.5% penicillin/streptomycin (50 units). HeLa were maintained in RPMI (Hyclone) supplemented with 10% FBS (Gibco) and 0.5% penicillin/streptomycin (50 units). 293T and HeLa cells were transfected with TransIT LTI (Mirus) according to the manufacture's protocol. Single time point episomal editing experiments were harvested 72 hrs post-transfection, and chromosomal editing experiments were harvested 96 hrs post-transfection.

APOBEC-mediated base editing reporter construct

The AMBER system was derived from HIV-1 NL4-3 by excising the *gag-pol*, *vif*, and *vpr* open reading frames using *Swa*I and *Sal*I restriction sites and blunt end ligation. *vpr* and the first ~1,200 bp of *env* were removed using *Sac*I and *Psi*I restriction sites and

blunt end ligation to retain the Rev response element (RRE). A gBlock, synthesized by integrated DNA technologies (IDT) to introduce a CMV promoter with a 3' *Age*I restriction site, was cloned into the *nef* open reading frame using *Bam*HI and *Kpn*I restriction sites. mCherry was PCR amplified using Phusion high-fidelity DNA polymerase (NEB) from a pcDNA3.1 expression plasmid with primers that introduce a 3' T2A self-cleaving peptide sequence (primers in **Table 1**) and cloned into a CloneJET PCR cloning vector (Thermo Fisher). eGFP was PCR amplified from a pcDNA3.1 expression plasmid with primers introducing scrambled nucleotide sequences at the 5' and 3' ends of the gene that retained the wild-type protein sequence (primers in **Table 1**). This was done to eliminate recombination during reverse-transcription of the viral reporter because the 5' and 3' ~20 nt of mCherry and eGFP are identical. The eGFP PCR amplicon was cloned into the mCherry-T2A cloning vector using *Xho*I and *Kpn*I restriction sites. Finally, the single mCherry-T2A-eGFP cassette was cloned into the modified NL4-3 vector using *Age*I and *Kpn*I restriction sites. mCherry L59S mutant was created using site-directed mutagenesis with Phusion high-fidelity DNA polymerase (NEB) (primers in **Table 1**) .

Base editing constructs

The rat APOBEC1-Cas9n-UGI-NLS construct (BE3) was provided by David Liu, Harvard University (7). A3A and A3Bctd cDNA sequences, each disrupted by an L1 intron to prevent toxicity in *E. coli* (19), were amplified using primers in **Table 1** and used to replace rat APOBEC1 in BE3 using a *Not*I site in the MCS and a *Xma*I site in the XTEN linker. Guide (g)RNAs targeting mCherry or non-specific (NS) sequence as a

control (**Table 1**) were cloned into MLM3636, obtained from J. Keith Joung, Harvard University, through Addgene (Plasmid #43860), using the accompanying Joung Lab gRNA cloning protocol.

Episomal base editing experiments

Semi-confluent 293T and HeLa cells in a 6-well plate format were transfected with 200 ng gRNA, 400 ng AMBER, and 600 ng of each base-editor [10 min, RT with 6 µl of TransIT LT1 (Mirus) and 200 µl of serum-free DMEM (Hyclone)]. Cells were harvested at indicated time points for editing quantification by flow cytometry.

Chromosomal base editing experiments

A semi-confluent 10 cm plate of 293T cells was transfected with 8 µg of an HIV-1 Gag-Pol packaging plasmid, 1.5 µg of a VSV-G expression plasmid, and 3 µg of the AMBER lentiviral reporter plasmid. Virus was harvested 48 hrs post-transfection, frozen at minus 80°C for 8 hrs, thawed, and used to transduce target cells (MOI = 1). 48 hrs post-transduction, 600 ng APOBEC-Cas9n-UGI editor and 250 ng of targeting or NS-gRNA were transfected into a semi-confluent 6-well plate of AMBER-transduced cells. Cells were harvested 96 hrs post-transfection and editing was quantified by flow-cytometry.

In a subset of experiments, mCherry-positive cells were recovered by FACS, converted to genomic DNA (Qiagen Gentra Puregene), and subjected to high-fidelity PCR using Phusion (NEB) to amplify mCherry target sequences (Primers in **Table 1**). PCR products were gel-purified (GeneJET Gel Extraction Kit, Thermo Scientific) and cloned into a sequencing plasmid (CloneJET PCR Cloning Kit, Thermo Fisher). Sanger

sequencing was done in 96-well format (Genewiz) using primers recommended with the CloneJET PCR Cloning Kit (**Table 1**).

To carry out *FANCF* editing enrichment experiments, semi-confluent HEK293T transduced with AMBER were co-transfected with 600 ng of A3Bctd-Cas9n-UGI and 200 ng of gRNA targeting both mCherry and *FANCF* in a 6-well format. 72 hrs post-transfection, cells were harvested and FACS was used to collect cells expressing mCherry. gDNA was harvested and a 452 bp fragment of *FANCF* was PCR amplified using nested primers shown in **Table 1**. A *Pst*1-HF (New England Biolabs) digest was done, and products were fractionated on an agarose gel to quantify editing efficiencies.

Immunoblots

$1x10^6$ cells were lysed directly into 2.5x Laemmli sample buffer, separated by a 4-20% gradient SDS-PAGE gel, and transferred to PVDF-FL membranes (Millipore). Membranes were blocked in 5% milk in PBS and incubated with primary antibody diluted in 5% milk in PBS supplemented with 0.1% Tween20. Secondary antibodies were diluted in 5% milk in PBS supplemented with 0.1% Tween20 and 0.01% SDS. Membranes were imaged with a Licor Odyssey instrument. Primary antibodies used in these experiments were rabbit anti-Cas9 (Abcam ab204448) and mouse anti-HSP90 (BD Transduction Laboratories 610418). Secondary antibodies used were goat anti-rabbit IRdye 800CW (Licor 827-08365) and goat anti-mouse Alexa Fluor 680 (Molecular Probes A-21057).

RESULTS

Construction and initial validation of an APOBEC-mediated base editing reporter system

An HIV-1 proviral backbone was chosen for the APOBEC-mediated base editing reporter (AMBER) system for maximal versatility, enabling use as a transient, multi-copy plasmid-based episomal editing reporter or as a stable, single-copy chromosomal DNA editing reporter. The AMBER system is a dual mCherry-T2A-eGFP fluorescence reporter driven by a CMV promoter to enable expression and quantification of real-time editing in living mammalian cells (schematic in **Figure 1A**). The AMBER system is based on creating a tight "off-to-on" gain of function fluorescence reporter in which an APOBEC editing hotspot, TCA, was introduced at a codon essential for mCherry function. A single C-to-U editing event at this codon results in reversion of TCA-to-TTA and restoration of fluorescence activity. Several T-to-C mutations were tested in mCherry and eGFP codons and most failed to completely ablate fluorescence, were not located an appropriate distance from a guide (g)RNA anchoring motif (PAM), and/or did not become substrates for editing (data not shown; APOBEC editosome schematic in **Figure 1B**).

One site in mCherry proved robust with no background fluorescence, and strong mCherry-positive signal upon transient co-expression of the rat APOBEC1 editosome BE3 and an appropriate mCherry-directed gRNA (fluorescence microscopy images in **Figure 1C** and quantification in **Figure 1D**). APOBEC-mediated editing of the mutant TCA codon (Ser59) back to the wild-type TTA codon (Leu59) restores mCherry fluorescence. As expected, base editing requires targeting of the APOBEC-Cas9 complex to the TCA hotspot, because a non-specific (NS) gRNA does not restore fluorescence

activity. Structural rationalization of this tight off-to-on system is shown in **Figure 1E**. Wild-type Leu59 has several stabilizing hydrophobic contacts with essential residues within the mCherry β-barrel (PDB 2H5Q) (20). This system is portable and capable of providing real-time read-outs of editing activity in a variety of different human cell lines (*e.g.*, 293T in **Figure 1C-D**, HeLa in **Figure 1F**, and U2OS and 3T3 data not shown).

Application of the AMBER system to create highly efficient next-generation base editing constructs based on human APOBEC3A and APOBEC3B

Optimization of base editing technologies will require editosomes with the highest possible efficiencies and structural information to guide rational improvements such as single nucleobase specificity [APOBEC1 and PmCDA1 have yet to yield structures, and the crystalized form of AID is significantly divergent (21)]. We therefore tested human APOBEC3A (A3A) and APOBEC3B (A3B) for Cas9n-directed base editing. These enzymes are the most efficient ssDNA C-to-U deaminases in human cells (22-25), and high-resolution crystal structures of both apo- and ssDNA-bound forms have been determined (26-29). A3A-ssDNA and A3B C-terminal domain (A3Bctd)-ssDNA structures share a unique U-shaped bound ssDNA conformation and provide an atomic explanation for the intrinsic 5'-TC specificity of these enzymes (28,29). As testament to the utility of this structural information, it informed a single amino acid change in a loop region adjacent to the active site of A3A that altered its intrinsic specificity from 5'-TC to 5'-CC (28). Additional enzyme customization is anticipated to enable tailoring these enzymes to all possible di- and trinucleotide contexts (5'-NC and 5'-NCN, respectively).

A3A-Cas9n-UGI and A3Bctd-Cas9n-UGI constructs were assembled and tested

in parallel with BE3 to directly compare editing efficiencies. These constructs were co-transfected into 293T cells with AMBER and a gRNA to direct editosomes to mCherry codon 59 or a NS-gRNA as a negative control. In a single time point experiment, the rat APOBEC1 editosome yielded 47% mCherry-positive cells, and both A3A and A3Bctd achieved 70% mCherry-positive cells (representative fluorescence images in **Figure 2A** and quantification in **Figure 2B**). Similarly, higher editing efficiencies were observed in time course studies in 293T cells, with both A3A and A3Bctd editosomes achieving nearly 40% mCherry fluorescence by 24 hrs and maximal fluorescence (70%) by 72 hrs before declining (as expected for transient transfection with non-replicating plasmids; **Figure 2C**). Anti-Cas9 immunoblots indicated that at least some of the improved editing efficiencies might be due to higher expression levels of the A3A- and A3Bctd-Cas9-UGI editosomes (**Figure 2D**).

Improved chromosomal DNA editing efficiencies using A3A and A3Bctd editosomes

To further compare the efficiencies of these editosomes, AMBER was pre-delivered to 293T and HeLa cells by lentiviral transduction (MOI 1.0). After 48 hrs incubation, the resulting mCherry-negative/eGFP-positive pools were co-transfected with the base editing constructs and a mCherry codon 59-directed gRNA or NS-gRNA as a negative control. As above, the A3A and A3Bctd editosomes performed better than the rat APOBEC1 editosome (**Figure 3A**). However, the single copy nature of the AMBER system in the context of the chromosome caused a 10-fold reduction in the overall efficiency of each editosome. This result is to be expected because reversion of a single copy chromosomal reporter, which is chromatinized to varying degrees depending on

integration position, will occur less frequently than editing of one of many episomal copies in a transient co-transfection experiment.

To further investigate the utility of the AMBER system, DNA sequencing was used to ask whether editing events catalyzed by APOBEC editisomes are specific to the intended 5'TCA motif or distributed more broadly within the ssDNA loop created by the gRNA base pairing to the target region. Chromosomal editing events were enriched by FACS for mCherry-positive cells, and single high-fidelity PCR amplicons were captured by cloning into a plasmid vector for Sanger sequencing (**Figure 3B**). As expected, almost all clones had a base-editing event at codon 59, TCA-to-TTA, which is necessary for restoration of mCherry fluorescence. Interestingly, many of the sequences also had C-to-T mutations in flanking regions displaced by annealing of the gRNA, but not in surrounding DNA regions that are presumably double-stranded DNA and protected from the single-strand specific DNA deaminase activity of the APOBEC enzymes. Furthermore, most of the C-to-T mutations occurred on the 5'-side of the target TCA trinucleotide, distal to the gRNA PAM. Codon 57 has the cytosine base in the wobble position (5'-GGC) and codon 58 is a proline (5'-CCC), which reconstruction experiments showed is dispensable for mCherry fluorescence (5'-TTT, Phe58 data not shown). Thus, in addition to enabling quantification of editing efficiencies in episomes and chromosomes, the AMBER system unexpectedly reports both on-target and off-target editing events. This helps explain why other sites in mCherry and eGFP were much less amendable to being developed into a reporter system.

Application of the AMBER system to enrich for editing events in a disease-relevant gene

We next asked if the AMBER system could be used to enrich for chromosomal DNA editing events at an unlinked genetic locus with disease relevance. AMBER-transduced eGFP-positive 293T cells were transfected with an A3A-, A3Bctd-, or rat APOBEC1-Cas9n-UGI base editing construct and gRNAs for *mCherry* codon 59 and *FANCF* codon 5. After 96 hrs incubation, mCherry-positive (AMBER-edited) cells were purified by FACS and editing events at *FANCF* were assessed using a PCR and restriction enzyme-based assay (**Figure 4A**). Wild-type *FANCF* DNA amplicons are 452 bp, and restriction by *Pst*I (5'-CTGCAG) results in two fragments, 192 and 260 bp, visible by agarose gel electrophoresis. APOBEC-mediated editing destroys the *Pst*I cleavage site and preserves the full-length fragment. The A3A and A3Bctd reactions yielded >10,000 mCherry-positive cells for this analysis, and unfortunately the rat APOBEC1 editosome yielded too few fluorescent cells for reliable purification (concordant with chromosomal editing data in **Figure 3**). Nevertheless, this restriction assay yielded very clear results with *FANCF* editing events being highly enriched in sorted mCherry-positive cells in comparison to unsorted pools (**Figure 4B**; 290-fold and 5-fold for A3A and A3Bctd editosomes, respectively). These data demonstrate the utility of the AMBER system for isolating subpopulations of cells with heterologous chromosomal editing events.

DISCUSSION

We report the development of a fluorescence-based APOBEC-mediated base editing

reporter (AMBER) system for rapid, efficient, and quantitative read-outs of base editing activity in living mammalian cells. The AMBER system is also, to our knowledge, the first to enable comparisons of the base editing efficiencies of the same isogenic reporter system in two different subcellular contexts - episomal high-copy conditions versus chromosomal single-copy conditions. Standard molecular biology procedures may be used to adapt this system to other mammalian and non-mammalian cell types and thereby reach near-universal status along with now near-ubiquitous CRISPR gene disruption technologies. For instance, as demonstrated here for *FANCF* as a representative disease-relevant chromosomal gene, transduction of AMBER and subsequently transient transfection of an appropriate APOBEC editosome into mammalian cells, along with gRNAs targeting mCherry codon 59 and a genomic site of interest, enables rapid enrichment by FACS of editing competent and properly targeted cells. This enrichment approach could be applied to almost any chromosomal target. Additionally, it is easy to envisage how the live cell AMBER system may be leveraged for further applications such as screening for modifiers (enhancers or inhibitors) of base editing activity.

In addition to utility, base editing efficiencies are an important consideration. Here, we use the AMBER system to validate new editosome complexes comprised of A3A and A3Bctd and demonstrate that these base editing complexes are more efficient than the previously described (7) rat APOBEC1-based editosome BE3 (**Figures 2**, **3**). Immunoblots indicate that at least part of the increased efficiencies may be due to higher expression levels (**Figure 2**). Attempts to transfect more BE3 and achieve similar protein expression levels were not successful, likely due to cellular toxicity caused by excessive amounts of transfected plasmid DNA (data not shown). Nevertheless, regardless of the

molecular explanation for the higher base editing efficiencies demonstrated here, many applications such as site-directed mutation and anti-viral mutagenesis are likely to benefit from using the most efficient available editosome complexes.

Base editing also has enormous potential for medical applications such as reverting disease mutations. Recent surveys estimate that over 25% of human disease mutations may be targetable by APOBEC-mediated base editing technologies (7,8). However, a major limitation of current base editing technologies, including the highly efficient A3A and A3Bctd editosomes reported here, is the significant level of adjacent off-target effects (*i.e.*, mutation of adjacent cytosines within the ssDNA region created by gRNA annealing). This problem was recognized in the original BE3 study (7) and confirmed in subsequent work (8-10,14). Lower frequencies of adjacent off-target events may be achieved by one or more of the following strategies: altering the linker between APOBEC and Cas9n, engineering APOBEC, engineering Cas9n, and/or directly delivering shorter-lived editosome complexes to cells (7-9,16). Engineering may be structure-guided for A3A, A3Bctd, and Cas9 (28-31), and/or instructed by functional screens using the AMBER system as an experimental readout (qualitative or quantitative). Distal off-target effects have also been documented for base editing (10,16), and these should also be considered in future optimization strategies. The AMBER system described here is well suited to quantify both on-target as well as adjacent off-target mutational events, and it may therefore also be a useful tool for developing truly specific editosomes and thus avoiding potentially detrimental off-target effects and advancing base editing technologies toward clinical applications.

ACKNOWLEDGEMENTS

We thank David Liu for providing BE3, Erik Toso and Michael Kyba for FACS, and Hideki Aihara, Maria Montiel-González, and Artur Serebrenik for helpful comments.

FUNDING

This work was supported by NIGMS R01 GM118000, NIAID R37 AI064046, and NCI R21 CA206309. A.S. received salary support from NSF-GRFP 00039202, and D.S. from NIH T90DE022732. R.S.H. is the Margaret Harvey Schering Land Grant Chair for Cancer Research, a Distinguished McKnight University Professor, and an Investigator of the Howard Hughes Medical Institute.

CONFLICT OF INTEREST STATEMENT

R.S.H. is a co-founder, shareholder, and consultant of ApoGen Biotechnologies Inc. The other authors declare no competing financial interests.

REFERENCES

1. Conticello, S.G. (2008) The AID/APOBEC family of nucleic acid mutators. *Genome Biol*, **9**, 229.
2. Harris, R.S. and Dudley, J.P. (2015) APOBECs and virus restriction. *Virology*, **479-480C**, 131-145.
3. Nabel, C.S., Manning, S.A. and Kohli, R.M. (2012) The curious chemical biology of cytosine: deamination, methylation, and oxidation as modulators of genomic

potential. *ACS Chem Biol*, **7**, 20-30.

4. Simon, V., Bloch, N. and Landau, N.R. (2015) Intrinsic host restrictions to HIV-1 and mechanisms of viral escape. *Nat Immunol*, **16**, 546-553.
5. Harris, R.S., Petersen-Mahrt, S.K. and Neuberger, M.S. (2002) RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell*, **10**, 1247-1253.
6. Severi, F., Chicca, A. and Conticello, S.G. (2011) Analysis of reptilian APOBEC1 suggests that RNA editing may not be its ancestral function. *Mol Biol Evol*, **28**, 1125-1129.
7. 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.
8. Rees, H.A., Komor, A.C., Yeh, W.H., Caetano-Lopes, J., Warman, M., Edge, A.S.B. and Liu, D.R. (2017) Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat Commun*, **8**, 15790.
9. Kim, Y.B., Komor, A.C., Levy, J.M., Packer, M.S., Zhao, K.T. and Liu, D.R. (2017) Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol*, **35**, 371-376.
10. Kim, D., Lim, K., Kim, S.T., Yoon, S.H., Kim, K., Ryu, S.M. and Kim, J.S. (2017) Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat Biotechnol*, **35**, 475-480.
11. Kim, K., Ryu, S.M., Kim, S.T., Baek, G., Kim, D., Lim, K., Chung, E., Kim, S. and Kim, J.S. (2017) Highly efficient RNA-guided base editing in mouse embryos.

Nat Biotechnol, **35**, 435-437.

12. Li, J., Sun, Y., Du, J., Zhao, Y. and Xia, L. (2017) Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system. *Molecular plant*, **10**, 526-529.
13. Lu, Y. and Zhu, J.K. (2017) Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Molecular plant*, **10**, 523-525.
14. Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J.L., Wang, D. and Gao, C. (2017) Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat Biotechnol*, **35**, 438-440.
15. Hess, G.T., Fresard, L., Han, K., Lee, C.H., Li, A., Cimprich, K.A., Montgomery, S.B. and Bassik, M.C. (2016) Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat Methods*, **13**, 1036-1042.
16. Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y. *et al.* (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*, **353**.
17. Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., Teramura, H., Yamamoto, T., Komatsu, H., Miura, K. *et al.* (2017) Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat Biotechnol*, **35**, 441-443.
18. Kuscu, C. and Adli, M. (2016) CRISPR-Cas9-AID base editor is a powerful gain-of-function screening tool. *Nat Methods*, **13**, 983-984.
19. Hultquist, J.F., Lengyel, J.A., Refsland, E.W., LaRue, R.S., Lackey, L., Brown, W.L.

and Harris, R.S. (2011) Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *J Virol*, **85**, 11220-11234.

20. Shu, X., Shaner, N.C., Yarbrough, C.A., Tsien, R.Y. and Remington, S.J. (2006) Novel chromophores and buried charges control color in mFruits. *Biochemistry*, **45**, 9639-9647.
21. Pham, P., Afif, S.A., Shimoda, M., Maeda, K., Sakaguchi, N., Pedersen, L.C. and Goodman, M.F. (2017) Activation-induced deoxycytidine deaminase: Structural basis for favoring WRC hot motif specificities unique among APOBEC family members. *DNA Repair (Amst)*, **54**, 8-12.
22. Stenglein, M.D., Burns, M.B., Li, M., Lengyel, J. and Harris, R.S. (2010) APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat Struct Mol Biol*, **17**, 222-229.
23. Carpenter, M.A., Li, M., Rathore, A., Lackey, L., Law, E.K., Land, A.M., Leonard, B., Shandilya, S.M., Bohn, M.F., Schiffer, C.A. *et al.* (2012) Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. *J Biol Chem*, **287**, 34801-34808.
24. Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B. *et al.* (2013) APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*, **494**, 366-370.
25. Ito, F., Fu, Y., Kao, S.A., Yang, H. and Chen, X.S. (2017) Family-wide comparative analysis of cytidine and methylcytidine deamination by eleven human APOBEC

proteins. *J Mol Biol*.

26. Bohn, M.F., Shandilya, S.M., Silvas, T.V., Nalivaika, E.A., Kouno, T., Kelch, B.A., Ryder, S.P., Kurt-Yilmaz, N., Somasundaran, M. and Schiffer, C.A. (2015) The ssDNA mutator APOBEC3A is regulated by cooperative dimerization. *Structure*, **23**, 903-911.
27. Shi, K., Carpenter, M.A., Kurahashi, K., Harris, R.S. and Aihara, H. (2015) Crystal structure of the DNA deaminase APOBEC3B catalytic domain. *J Biol Chem*, **290**, 28120-28130.
28. Shi, K., Carpenter, M.A., Banerjee, S., Shaban, N.M., Kurahashi, K., Salamango, D.J., McCann, J.L., Starrett, G.J., Duffy, J.V., Demir, O. *et al.* (2017) Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat Struct Mol Biol*, **24**, 131-139.
29. Kouno, T., Silvas, T.V., Hilbert, B.J., Shandilya, S.M.D., Bohn, M.F., Kelch, B.A., Royer, W.E., Somasundaran, M., Kurt Yilmaz, N., Matsuo, H. *et al.* (2017) Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. *Nat Commun*, **8**, 15024.
30. Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E. and Doudna, J.A. (2016) Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*, **351**, 867-871.
31. Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F. and Nureki, O. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, **156**, 935-949.

Figure Legends

Figure 1. A real-time fluorescent reporter system for APOBEC-Cas9 base editing.

A) Schematic of the APOBEC-mediated base editing reporter (AMBER) system in the context of a lentiviral construct with a CMV promoter that drives expression of a bicistronic message encoding mutant mCherry and wild-type eGFP. Reversion of 5'-TCA (Ser59) to 5'-TTA (Leu59) by APOBEC-mediated editing restores mCherry fluorescence.

B) Schematic of an APOBEC-Cas9/gRNA editosome engaging a DNA target. C-to-U editing occurs in the ssDNA loop displaced by Cas9-mediated gRNA annealing to target DNA. The non-edited strand is broken by the Cas9 nickase, which targets DNA repair mechanisms (not shown) to the nicked strand and facilitates conversion of the uracil lesion into a thymine mutation.

C) Representative images of mCherry-positive cells catalyzed by BE3 and mCherry codon 59-directed gRNA (#59-gRNA) but not with NS-gRNA (NS, non-specific; inset white bar = 30 μm).

D) Quantification of the base editing experiment in panel C (n=3; average +/- SD).

E) A ribbon schematic of the mCherry structure with a zoom-in highlighting essential interactions between Leu59 and β-barrel residues required for fluorescence (Leu59 is labeled Leu54 in the mCherry structure pdb 2H5Q).

F) AMBER activity catalyzed by BE3 in transfected HeLa cells (n=3; average +/- SD).

Figure 2. High-efficiency editing by human A3A and A3Bctd editosomes.

A) Representative fluorescence microscopy images of AMBER activity in 293T cells catalyzed by human A3A, human A3Bctd, or rat APOBEC1 editosomes (mCherry codon 59-directed gRNA versus NS-gRNA; inset white bar = 30 μm).

B) Quantification of the experiment in panel 'A' together with 2 independent parallel experiments (n=3; average +/- SD). Corresponding immunoblots of expressed APOBEC-Cas9n-UGI constructs, and HSP90 as a loading control.

C) Time course of AMBER activity in 293T cells catalyzed by human A3A, human A3Bctd, or rat APOBEC1 editosomes (mCherry codon 59-directed gRNA versus NS-gRNA; n=3; mean +/- SD; error bars smaller than symbols are not shown).

Figure 3. Chromosomal editing by A3A and A3Bctd editosomes.

A) Base-editing of integrated, genomic AMBER by the indicated editosomes in 293T and HeLa cells (n=3, average +/- SD).

B) Sanger sequencing results for genomic AMBER edited by A3A or A3Bctd editosomes. Wild-type nucleotides are marked by black dots, and mutations by changed letters (mostly C-to-T). The number of individual sequencing reads representing each mutated sequence is indicated to the right of each line.

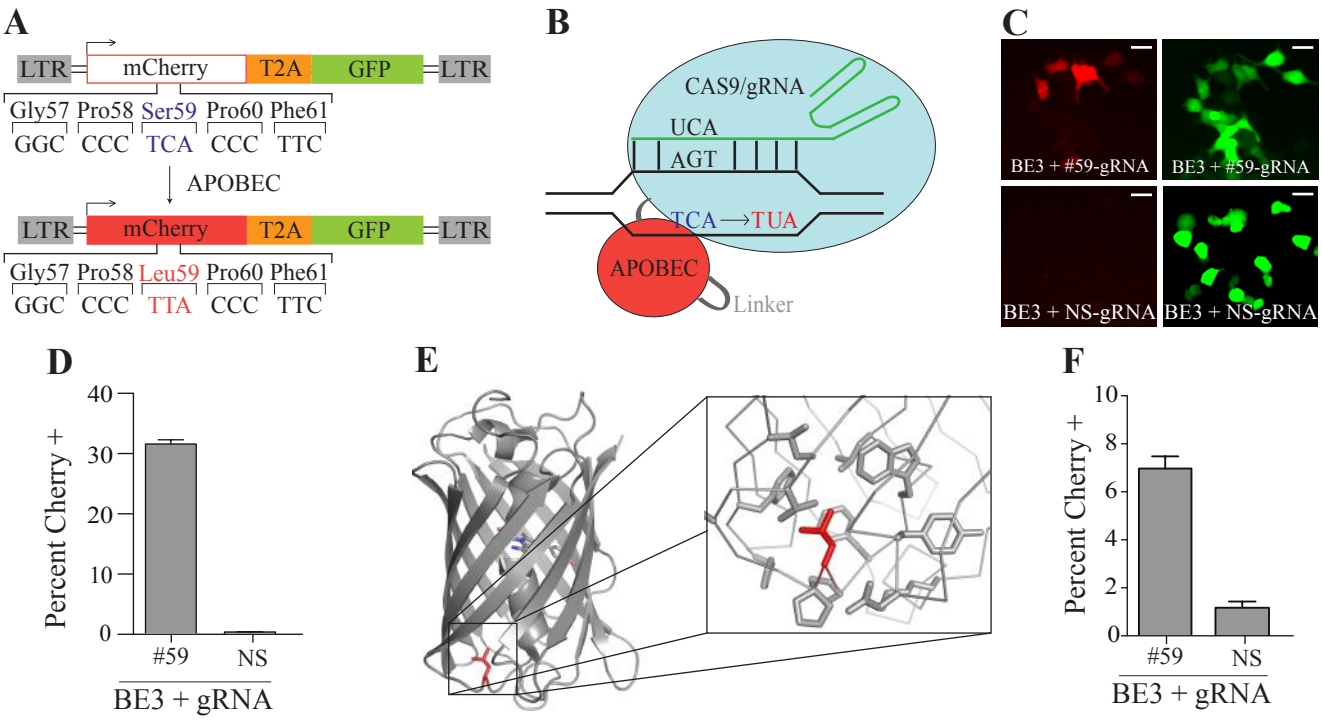
Figure 4. AMBER enriches for base-editing events at unlinked genomic loci.

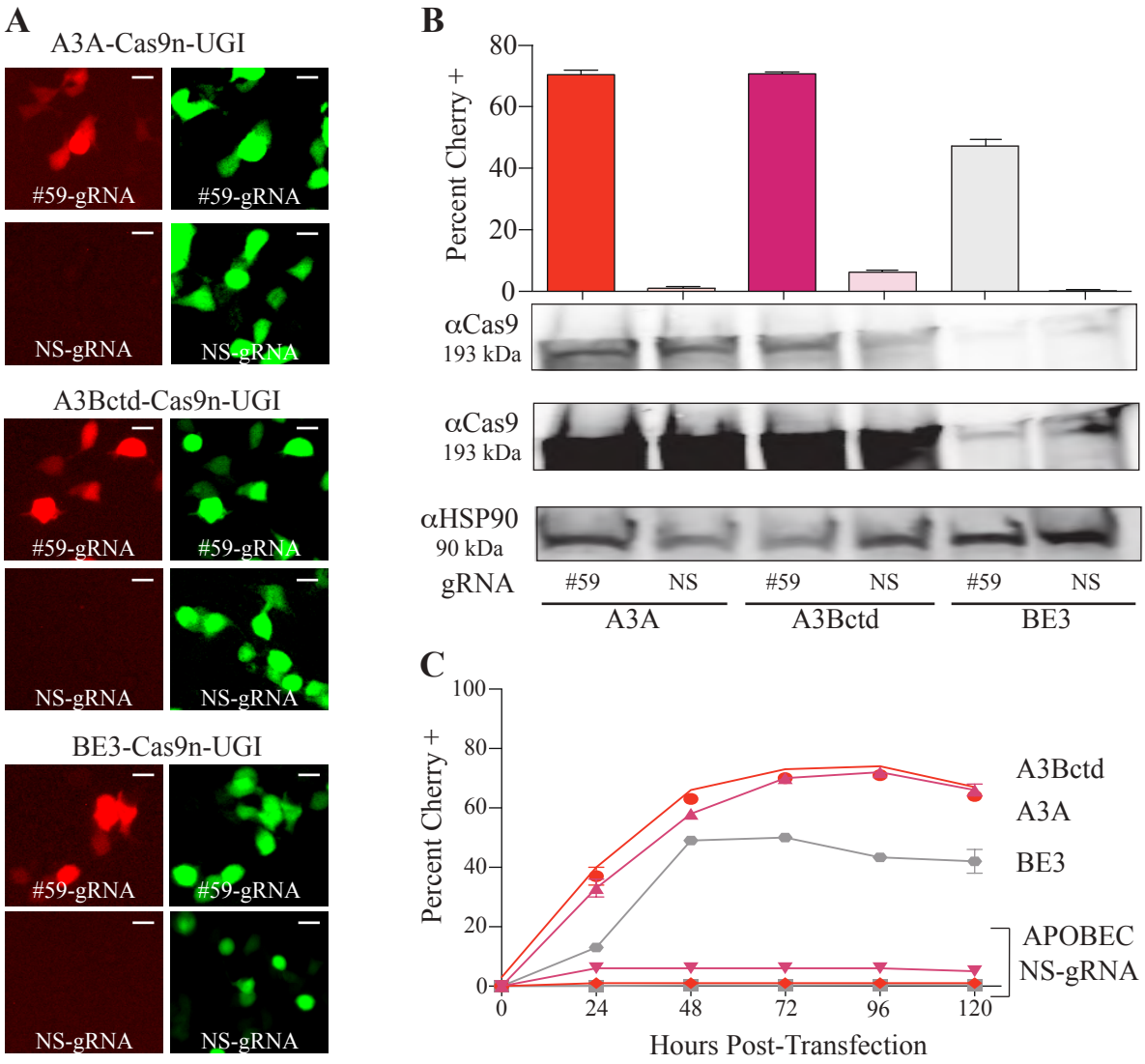
A) Schematic of *FANCF* and the *Pst*I restriction assay used to quantify chromosomal base editing of this locus. Base editing events destroy the *Pst*I cleavage site and block cleavage of the 452 bp amplicon into 260 and 192 bp products.

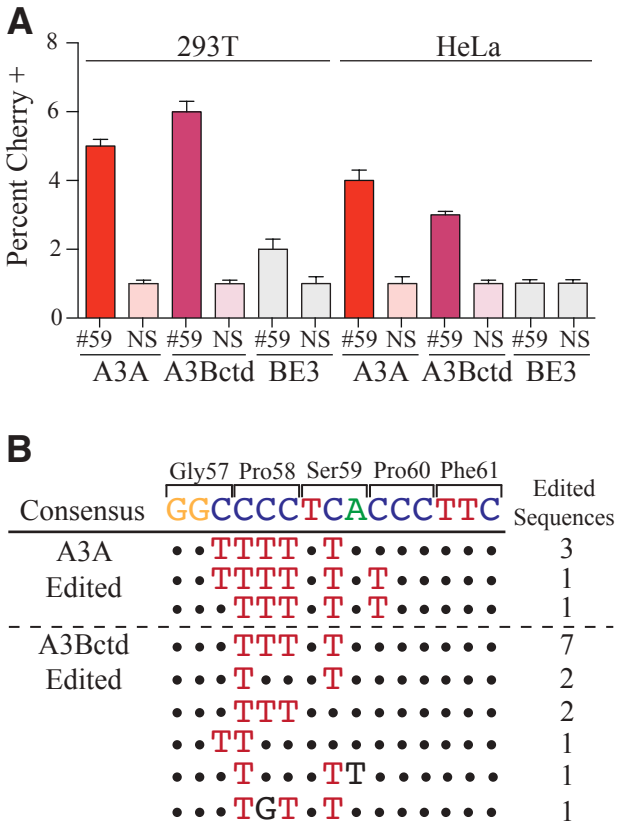
B) Representative agarose gels images showing the results of *FANCF* base editing by A3A and A3Bctd editsomes (top and bottom, respectively). The percentage of base editing was calculated by dividing the percentage of substrate band by the total of substrate and product bands following *Pst*I cleavage.

Table 1. Oligonucleotide sequences.

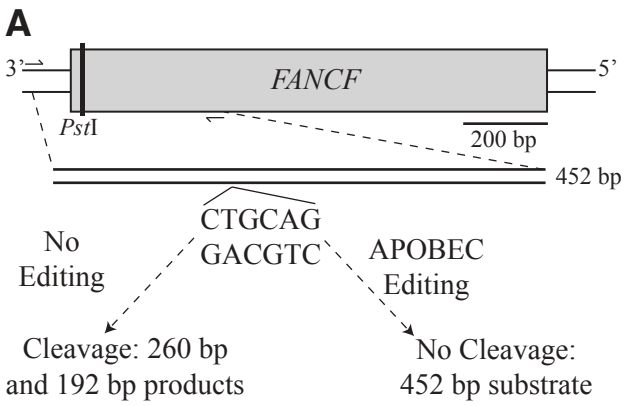
Primer	Sequence (5'-to-3')
A3A Cloning Forward Primer	AGATCCGCGGCCGCGCCGCCACCATGA TGGAAGCCAGCCCAGCATCCGGGC
A3A Cloning Reverse Primer	TGAGGTCCCGGGAGTCTCGCTGCCGC TTCCGTTTCCCTGATTCTGGAGAATG
A3Bctd Cloning Forward Primer	AGATCCGCGGCCGCGCCGCCACCATG GATCCAGACACATTCACTTTCAACT
A3Bctd Cloning Reverse Primer	TGAGGTCCCGGGAGTCTCGCTGCC GCTGTTTCCCTGATTCTGGAGAATGGCC
mCherry L59S SDM Forward Primer	AAGGGTGGCCCCTCACCCTTCGCCTGGG
mCherry L59S SDM Reverse Primer	CCCAGGCGAAGGGTGAGGGGCCACCCTT
Codon #59-directed mCherry gRNA Forward Primer	ACACCTGGCCCCTCACCCTTCGCCTG
Codon #59-directed mCherry gRNA Reverse Primer	AAAACAGGCGAAGGGTGAGGGGCCAG
NS gRNA Forward	ACACCGCACTACCAGAGCTAACTCAG
NS gRNA Reverse	AAAACTGAGTTAGCTCTGGTAGTGCG
T2A Cloning Forward Primer	CTGGCTACCGGTATGGTGAGCAAGGGCGAGG
T2A Cloning Reverse Primer	TTAAAGGTACCAGGGCCGGGATTCTCCTCCACG TCACCGCATGTTAGAAGACTTCCTCTGCCCTCCT TGTACTCGAGATCTGCACCGGGCTTGTACAGCTC GTCCATGCC
eGFP Cloning Forward Primer	GCAGATCTCGAGTACAAGGAGGGCAGAGGAAG TCTTCTAACATGCGGTGACGTGGAGGAGAATCC CGGCCCTCTGGTCAGTAAAGGTGAAGAACTGTTCACCG
eGFP Cloning Reverse Primer	CTTAAAGGTACCTTATTTATATAATTCATCCATA CCGAGAG
mCherry Amplification Forward Primer	ATGGCCATCATCAAGGAGTT
mCherry Amplification Reverse Primer	CTCTGCCCTCCTTGTACTCG
CloneJET Sequencing Forward Primer	CGACTCACTATAGGGAGAGCGGC
CloneJET Sequencing Reverse Primer	AAGAACATCGATTTTCCATGGCAG

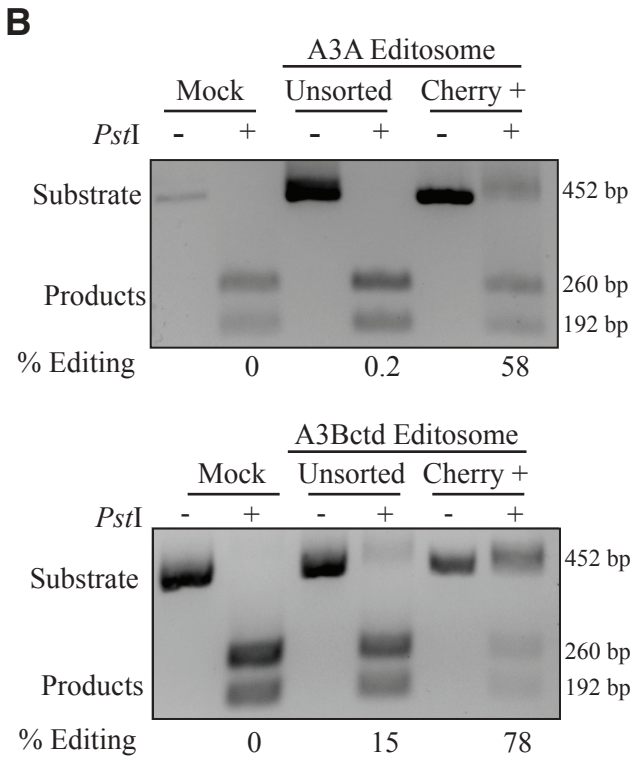

A
LTR
mCherry
T2A
GFP
LTR
Gly57 Pro58 Ser59 Pro60 Phe61
GGC CCC TCA CCC TTC
APOBEC
LTR
mCherry
T2A
GFP
LTR
Gly57 Pro58 Leu59 Pro60 Phe61
GGC CCC TTA CCC TTC
B
CAS9/gRNA
UCA
AGT
TCA→TUA
APOBEC
Linker
C
BE3 + #59-gRNA
BE3 + #59-gRNA
BE3 + NS-gRNA
BE3 + NS-gRNA
D
Percent Cherry +
40
30
20
10
0
#59
NS
BE3 + gRNA
E
F
Percent Cherry +
10
8
6
4
2
0
#59
NS
BE3 + gRNA


A
A3A-Cas9n-UGI
#59-gRNA
#59-gRNA
NS-gRNA
NS-gRNA
A3Bctd-Cas9n-UGI
#59-gRNA
#59-gRNA
NS-gRNA
NS-gRNA
BE3-Cas9n-UGI
#59-gRNA
#59-gRNA
NS-gRNA
NS-gRNA
B
Percent Cherry +
80
60
40
20
0
αCas9
193 kDa
αCas9
193 kDa
αHSP90
90 kDa
gRNA
#59
NS
#59
NS
#59
NS
A3A
A3Bctd
BE3
C
Percent Cherry +
100
80
60
40
20
0
0
24
48
72
96
120
Hours Post-Transfection
A3Bctd
A3A
BE3
APOBEC
NS-gRNA



St. Martin et al., Figure 3


A
3'
FANCF
5'
PstI
200 bp
452 bp
CTGCAG
GACGTC
No
Editing
APOBEC
Editing
Cleavage: 260 bp
and 192 bp products
No Cleavage:
452 bp substrate


B
A3A Editosome
Mock
Unsorted
Cherry +
PstI
- + - + - +
Substrate
Products
452 bp
260 bp
192 bp
% Editing
0
0.2
58
A3Bctd Editosome
Mock
Unsorted
Cherry +
PstI
- + - + - +
Substrate
Products
452 bp
260 bp
192 bp
% Editing
0
15
78

8. Appendices

Oligonucleotides and Primers (Supplementary table 1)

Number	Sequence	Enzyme	Purpose
1	AAAGCTAGCACCATGGACAGCCTCTTGATG	NheI	to clone AID
2	ATCAGATCTCAAAGTCCCAAAGTACGAAATG	BglII	to clone AID
3	AAAgaattcGGACAAACCACAACTAGAATGCAG	EcoR1	to clone into poly A Pbuer
4	AAAgaattcggatccTTAGTTGCCTTGGTTTTGCA	EcoR1-BamH1	to clone into A3BPbuer
5	tttgctagcATGAATCCACAGATCCGTAAC	Nhe1	to clone into A3BPbuer
6	tttgctagcATGGAAGCCAGCCCAGCATC	Nhe1	to clone into A3APbuer
7	AAAgaattcTCCGTTTCCCTGATTCTGGAG	EcoR1	to clone into A3APbuer
8	tttgctagcATGAAGCCTCACTTCAGAAA	Nhe1	to clone into A3GPbuer
9	AAAgaattcagatcTCAGTTTTCCTGATTCTGGA	EcoR1-BglII	to clone into A3GPbuer
10	TTGCGCCGACATCATAACGGT		seq for ptrc99
11	TATCAGACCGCTTCTGCGTT		seq for ptrc100
12	gctagcGAATTCCCACCATGGCTC	NheI	to clone A3H into pbuer
13	agatctTCTAGACTTCTAAAACATCC	BglII	to clone A3H into pbuer
14	gctagcTCTCGAGATGGCCCAGAAGGAA	NheI	to clone APOBEC2 into pbuer
15	agatctCCGCGGTACCGTCGACTaa	BglII	to clone APOBEC2 into pbuer
16	gctagcACCATGAATCCACAGATCAGAAACC	NheI	to clone APOBEC 3C into pbuer
17	agatctTCGACGGAGACCCCTCA	BglII	to clone APOBEC 3C into pbuer
18	gctagcAGACCATGGCTCTGTTAACAGCCGAAA	NheI	to clone APOBEC 3H into pbuer
19	agatctCGCGGTACCGTCGACTGC	BglII	to clone APOBEC 3H into pbuer
20	gctagcCCTACGCAAAGCCCTATGGTGGAAC	NheI	to clone APOBEC 3F into pbuer
21	agatctCAGTATGTCGTCACAGAACCAAGAG	BglII	to clone APOBEC 3F into pbuer
22	caccgTTGAGGCCTTCGTTGGAAAC		sgRNA for visualizing the centromere of HAC HT1080 cells
23	aaacGTTTCCAACGAAGGCCTCAAc		sgRNA for visualizing the centromere of HAC HT1080 cells
24	ACCATGCCCAAGAAGAAGCGCA		to clone into px330 to make it into dcas9
25	ggaattcGTttaCCGCGGTGCCTGAGCCT	EcoRI	to clone into px330 to make it into dcas9
26	gagaagggacaagcacatgg		RT- PCR analysis for exprsn for APOBEC 3 A
27	tggatccatcaagtgtctgg		RT- PCR analysis for exprsnfor APOBEC 3 A
28	gacccttggtccttcgac		RT- PCR analysis for exprsn for APOBEC 3 B

29	gcacagccccaggagaag		RT- PCR analysis for exprsn for APOBEC 3 B
30	agcgcttcagaaaagagtgg		RT- PCR analysis for exprsn for APOBEC 3 C
31	aagtttcgttccgatcgttg		RT- PCR analysis for exprsn for APOBEC 3 C
32	acccaaacgtcagtcgaatc		RT- PCR analysis for exprsn for APOBEC 3 D/DE
33	cacatttctgcgtggttctc		RT- PCR analysis for exprsn for APOBEC 3 D/DE
34	ccgtttggacgcaaagat		RT- PCR analysis for exprsn for APOBEC 3 F
35	ccaggtgatctggaaacactt		RT- PCR analysis for exprsn for APOBEC 3 F
36	ccgaggacccgaaggttac		RT- PCR analysis for exprsn for APOBEC 3 G
37	tccaacagtgctgaaattcg		RT- PCR analysis for exprsn for APOBEC 3 G
38	agctgtggccagaagcac		RT- PCR analysis for exprsn for APOBEC 3 H
39	cggaatgtttcggctgtt		RT- PCR analysis for exprsn for APOBEC 3 H
40	AAAATGTCCGCTGGGCTAAGGG		RT pcr exprsn recognizes splice variants 12345
41	GCAGTAAAAATAATCTTTGAAGGTC		RT pcr exprsn recognizes splice variants 1
42	GCTACGGACCAAAATAGAGnnnGGTsAGGGGnnnATTCCAGTGCGCTCCAATG		To generate ADAR mutant library T375 S490 E488
43	GCTACGGACCAAAATAGAGrgTGGTsAGGGGhmmATTCCAGTGCGCTCCAATG		To generate ADAR mutant library T375 S490 E489
45	CTCTATTTTGGTCCGTAGCT		To generate ADAR mutant library T375 S490 E490
46	CACATAAGTGTTTCTACAGGArscAAATGTATTAATGGTGAATACA		To generate ADAR mutant library T375 S490 E491
47	TGTATTCACCATTAATACATTTgsyTCCTGTAGAAACACTTATCAC		To generate ADAR mutant library T375 S490 E492
48	ataccatggccTTGCACTTGGATCAGACGCCA		ADAR2 forward primer
49	ttttCCCGGgAGTCTCGCTGCCGCTTTTACAGGGCGTGAGTGAGAACT	XmaI	clone ADAR2 in BE1 restoring the entire linker
50	ttttCCCGGgACAGGGCGTGAGTGAGAACT	XmaI	clone ADAR2 in BE1
51	AGCTTacctctgctCTGTAGgcgagatgcGGGGGGaG		Oligo to create the target site for ADAR-BE system
52	AATTCtCCCCCCgcatctcgcCTACAGagcagaggtA		Oligo to create the target site for ADAR-BE system
53	caccgtctgctCTGCAGgcgagatgcGGG		sgRNA for the ADAR2 to target

			mCherry-Target-eGFP
54	AAACCCCgcatctcgcCTGCAGagcagaC		sgRNA for the ADAR2 to target mCherry-Target-eGFP
55	caccgctctgctCTGCAGgcgagatgcGG		sgRNA for the ADAR2 to target mCherry-Target-eGFP
56	AAACCCgcatctcgcCTGCAGagcagagC		sgRNA for the ADAR2 to target mCherry-Target-eGFP
57	caccgcctctgctCTGCAGgcgagatgcG		sgRNA for the ADAR2 to target mCherry-Target-eGFP
58	AAACCgcatctcgcCTGCAGagcagaggC		sgRNA for the ADAR2 to target mCherry-Target-eGFP
59	caccgacctctgctCTGCAGgcgagatgc		sgRNA for the ADAR2 to target mCherry-Target-eGFP
60	AAACgcatctcgcCTGCAGagcagaggtC		sgRNA for the ADAR2 to target mCherry-Target-eGFP
61	caccgTacctctgctCTGCAGgcgagatg		sgRNA for the ADAR2 to target mCherry-Target-eGFP
62	gcattaaatgactgccatgcacagataatatctcggagatccttg		To mutagenize the deamse domain from HAE to HAQ to create a dead mutant of ADAR
63	caaggatctccgagatattatctgtgcatggcagtcatttaatgc		To mutagenize the deamse domain from HAE to HAQ to create a dead mutant of ADAR
64	caccgGGCGCGCGAAATTTGCGTGA		sgRNA for visualizing the centromere of HAC HT1080 cells
65	aaaacTCACGCAAATTTCGCGCGCC		sgRNA for visualizing the centromere of HAC HT1080 cells
66	caccgTGAAGGAGTGCAGTGCTCTCGGTG		sgRNA for visualizing the centromere of HAC HT1080 cells
67	aaaacCACCGAGAGCACTGCACTCCTTCA		sgRNA for visualizing the centromere of HAC HT1080 cells
68	aaggatctccgagatattattgctgcatggcagtcatttaatg		to obatin catalytically deficient ADAR E396A
69	cattaaatgactgccatgcagcaataatatctcggagatcctt		to obatin catalytically deficient ADAR E396A

Supplementary data 2. (Scripts)

Patient sample data were taken from follow supplementary materials.

https://tcga-data.nci.nih.gov/docs/publications/esca_2016/

The following script was used to build the data set for STAD and ESCA:

```
awk'{print$1"\t"$2"\t"$3"\t"$4"\t"$5"\t"$6"\t"$7"\t"$8"\t"$10"\t"$14"\t"$15"\t"$16"\t"$17
"\t"$21"\t"$22"\t"$24"\t"$36"\t"$37"\t"$43"\t"$45"\t"$46}'Starting_dataset.txt>Completa_s
tarting_Dataset_1.txt (for STAD dataset)
```

This was followed by division of the set of samples into two tumor types Stomach adenocarcinoma (STAD) and Esophageal carcinoma (ESCA).

```
awk'{if($6=="ESCA")print$0}'Completa_starting_Dataset_1.txt>Completa_starting_dataset_ESC
A_1.txt (for ESCA)
        awk'{if($6=="STAD")print$0}'Completa_starting_Dataset_1.txt>Completa_starting_data
        set_STAD_1.txt
```

Supplementary data 3. (Scripts)

```
# Open and match dataset from TCGA consortium

 #! /bin/perl -w

use strict;

use File::Find;

use warnings;


die "USAGE: perl Correlation_CIN_AID-APO_ex.pl <divided_sample_file> <Keys_file> <exporession_file>" unless @ARGV==2;


print "\n

    ############################

    ##### MUTATION MATCHIN ######

    ############################\n";

my $Dataset = $ARGV[0];

my $keys = $ARGV[1];

#################################

##### FIRST PART DATASET HASH #####

#################################

print "\nCreating dataset hash ... \n\n";

#### HASH_PARTS #####

my %dataset;
```

```
my %keys;

my %expression;

##### COUNTER PART #####

my $Count_code;

my $hash_element;

#my $count;

my $Count_code_ke;

my $keys_hash_ele;


#To obtain barcode ID and stored as hash.

open (DATA, "<$Dataset") || die "Cannot open file $Dataset $!";

while (my $i = <DATA>){

   chomp $i;

   #print $i;

   if ($i =~ /^((TCGA\S+)\t\S.+\t(\S+)\t(\S+))/){

      $Count_code++;


      $dataset{$2}{$3}{$4} = $1;


   }

}

close (DATA);

#Here, I match the metadata with barcode ID with the barcode present in the metadata of the new genome GRCh.38. This gives me the expression file ID (FPKM OQ).


$hash_element = keys %dataset;

open (KEYS, "<$keys") || die "Cannot open file $keys $!\n";


while (my $e = <KEYS>){

   chomp $e;
```

```
    if ($e =~ /^((TCG\S+)\t(\S+.gz)\t(\S+)\t\S+\t\S+\t\S+\t(\S+)\t(\S.+))/){

        $Count_code_ke++;

        $keys{$3}{$2}{$4}{$5} = $6;
    }
}
close KEYS;
##### VERIFY SECOND HASH #### PASSED
#
$keys_hash_ele = keys %keys;
my %match;

#### COUNT SAMPLE NO RNA-seq ####
my $negative=0;

foreach my $file (keys %keys){
    foreach my $barcode (keys $keys{$file}){
        foreach my $nuc (keys $keys{$file}{$barcode}){
            foreach my $tcel (keys $keys{$file}{$barcode}{$nuc}){
                if (exists $dataset{$barcode}{$nuc}{$tcel}){
                    #print "$file\t$dataset{$barcode}{$nuc}{$tcel}\n";
                    $match{$file}{$barcode} = undef;
                }
                else {
                    $negative++;
                }
            }
        }
```

```
    }
}
print "My count code: $Count_code_ke\tMy hash Keys elements: $keys_hash_ele\tN=$negative have not RNA-seq data in new
GDC\n\n";
#From the above script I have the expressionID which needs to be matched
my $dir = "/Users/Salvatorebioinfo/Desktop/EDITING/GRCH38/gen_expression/gene_ex/";


find(\&do_process, $dir);
sub do_process{
    if ($_ =~ /^TCGA-(\S+)_(\S+\.FPKM-UQ\S+\.gz)/){
#       #print $_,"\n";
#
        foreach my $file (keys %match){
            if ($2 eq $file){
                open my $in, '-|', 'gzip', '-dc', $_;
#
                while (my $line = <$in>){
                    chomp $line;


                    $line =~/^(ENS\S+).\d+\t(\S+)$/;
                    #print "$1\t$2\n";
                    my $gene= $1;
                    my $value = $2;
#
                    if ($line =~ /(#for each APOBEC gene)\t(\S.+)/){
                        $expression{$file}{#foreach APOBEC gene} = log($2+1)/log(2);
                        #print "$file\t$1\t$2\n"



                    }
```

```
            }
          }
        }
      }
}

my $out = "Final_dataset_STAD_1.txt";
open (OUTPUT, ">$out") || die "Cannot open file $out $!\n";
#
print OUTPUT "file\tbarcode\tEC\tESCC\tGEA\tGEA-CIN\tDisease_code\tHistological_Type_Oesophagus\tHistological_Type\tBarretts_oesophagus\tEBV_positive\tMSI_status\tSCNA_High/Low\tGastric_classification\tPathologic_stage\tGrade\tAge_at_initial_pathologic_diagnosis\tMutationRate\tMutation_Rate_Category\tTP53_mutation\tPercent_tumor_nuclei\tPercent_tumor_cells\tTissues\tAPOBEC1\tAICDA\tAPOBEC3A\tAPOBEC3AP1\tAPOBEC3B\tAPOBEC3B-AS1\tAPOBEC3C\tAPOBEC3D\tAPOBEC3F\tAPOBEC3G\tAPOBEC3G\tTP53\tKRAS\tRBM47\n";
#
foreach my $file (keys %match){
   foreach my $barcode (keys $keys{$file}){
      foreach my $nuc (keys $keys{$file}{$barcode}){
         foreach my $tc (keys $keys{$file}{$barcode}{$nuc}){


            if (exists $expression{$file}){

               if (exists $dataset{$barcode}){
                  foreach my $nuc (keys $dataset{$barcode}){
                     foreach my $tcel (keys $dataset{$barcode}{$nuc}){
                                                                                    print OUTPUT "$file\t$dataset{$barcode}{$nuc}{$tcel}\t$keys{$file}{$barcode}{$nuc}{$tc}\t$expression{$file}{APOBEC1}\t$expression{$file}{AICDA}\t$expression{$file}{APOBEC3A}\t$expression{$file}{APOBEC3AP1}\t$expression{$file}{APOBEC3B}\t$expression{$file}{'APOBEC3B-
```

```
AS1'}\t$expression{$file}{APOBEC3C}\t$expression{$file}{APOBEC3D}\t$expression{$file}{APOBEC3F}\t$expression{$file}{APOB
EC3G}\t$expression{$file}{APOBEC3H}\t$expression{$file}{TP53}\t$expression{$file}{KRAS}\t$expression{$file}{RBM47}\n";


                    }
                  }
                }
              }
            }
          }
        }
    }
print "DONE\n\n";
close OUTPUT;
```

Using the above script, I obtained a dataset for ESCA and STAD where, all the patient's tumors were matched against their respective barcodes, tumor type and CIN status to the expression levels of AID/APOBECs.

The following script below was made using the R programing language to obtain boxplots and analyse the correlation data.

```
File1 <- as.data.frame(read.table("/Users/Salvatorebioinfo/Desktop/UDAY/Tumor_vs_Normal.txt", sep = "", header = TRUE, quote = ",", fill = TRUE))


attach(File1)


#names(File1)

Plot <- data.frame(subset(File1))


Plot$pcolor[Plot$Gastric_classification=="CIN"] <- "orange"
Plot$pcolor[Plot$Gastric_classification=="GS"] <- "red"
Plot$pcolor[Plot$Gastric_classification=="MSI"] <- "blue"
Plot$pcolor[Plot$Gastric_classification=="EBV"] <- "green"
Plot$pcolor[Plot$Gastric_classification=="NOROM"] <- "purple"
#Plot$pcolor[Plot$Gastric_classification=="CIN-NO"] <- "purple"
head(Plot)
names(Plot)
Plot_sorted  <- Plot[order(Gastric_classification, SCNA_High.Low),]


rm(File1)
rm(Plot)


library(ggplot2)
#quartz()
qplot(x=Gastric_classification, y=KRAS, data=Plot_sorted , geom=c("boxplot","jitter") , fill=Gastric_classification, main="KRAS_exp in CIN ")

library(scales)
pdf(file.path(paste("/Users/Salvatorebioinfo/Desktop/UDAY/Box_plot2/", "STAD_KRAS.pdf")), width=10, height=10)
qplot(x=Gastric_classification, y=KRAS, data=Plot_sorted , geom=c("boxplot","jitter") , fill=Gastric_classification, main="KRAS_exp in CIN ")
dev.off()

names(Plot_sorted)
```