Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

Master of Sciences Philipp Konstantin Zimmermann

Born in Berlin-Zehlendorf

Oral examination: April 16th, 2015

Genome-wide detection of induced DNA double strand breaks

Referees:

Prof. Dr. Christof von Kalle

Prof. Dr. Stefan Wiemann

1. INTRODUCTION

1.1   The DNA damage response and genomic instability

1.1.1   DNA damaging agents and the DNA damage response

The DNA in the cell encodes the genetic information required for the functioning of all cells and living organisms. Every day, the DNA is under constant pressure to be destabilized by various processes and agents. These lesions can block DNA replication and transcription and are broadly divided according to their origin into either endogenous or exogenous (Figure 1) [1]. Endogenous processes such as DNA replication by DNA polymerases induce low levels of DNA damage. Other endogenous sources of DNA damage include genomic fragile sites, nucleases and reactive oxygen species (ROS) such as superoxide and hydrogen peroxide stemming from metabolic processes in the mitochondria [2]. These latter, highly reactive molecules form chemical bonds with other molecules within seconds of their production. Among their target molecules, the DNA is a susceptible target, in which ROS can cause base and sugar modifications, DNA-protein crosslinks, single-strand and double-strand breaks. In addition to endogenous processes, numerous exogenous sources also threaten the genomic integrity (Figure 1). The most pervasive form of exogenous DNA damaging agent is ultraviolet (UV) rays originating from sunlight that can hit the DNA directly and induce up to 100,000 DNA lesions per exposed cell [2]. These UV light generate various forms of DNA modifications, the most stable being the covalent link between two pyrimidine DNA bases [2]. X-rays represent another type of DNA damaging agent, which is often used for imaging and cancer radiation-therapy. These rays generate radicals in the DNA or in other cellular molecules that attack DNA bases, causing DNA oxidations as well as single-strand and double-strand breaks. Additional examples of exogenous DNA damaging agents include benzo-α-pyrene, polycyclic-aromatic hydrocarbons and cancer chemotherapeutics that interfere with the DNA replication or the DNA damage repair machinery.

Along with the various types of DNA damages induced by endogenous processes and environmental agents, DNA double strand breaks (DSB) are the most dangerous DNA lesion [1]. In contrast to other types of DNA damage, DSB directly threaten genomic integrity, because they disrupt the continuity of the DNA. Failure to efficiently repair such DNA lesions can compromise genome integrity and result in a variety of genomic changes. Amongst others these include point mutations, chromosome number variations, gross chromosomal rearrangements such as translocations, amplifications, insertions, deletions, dimeric oncogenes and copy-number variations, which can initiate neoplastic transformation [1]. In order to maintain genomic integrity and counteract the damage inflicted on the genome, the cell triggers a coordinated network of signaling cascades, which are collectively termed the DNA damage response (DDR) (Figure 1).
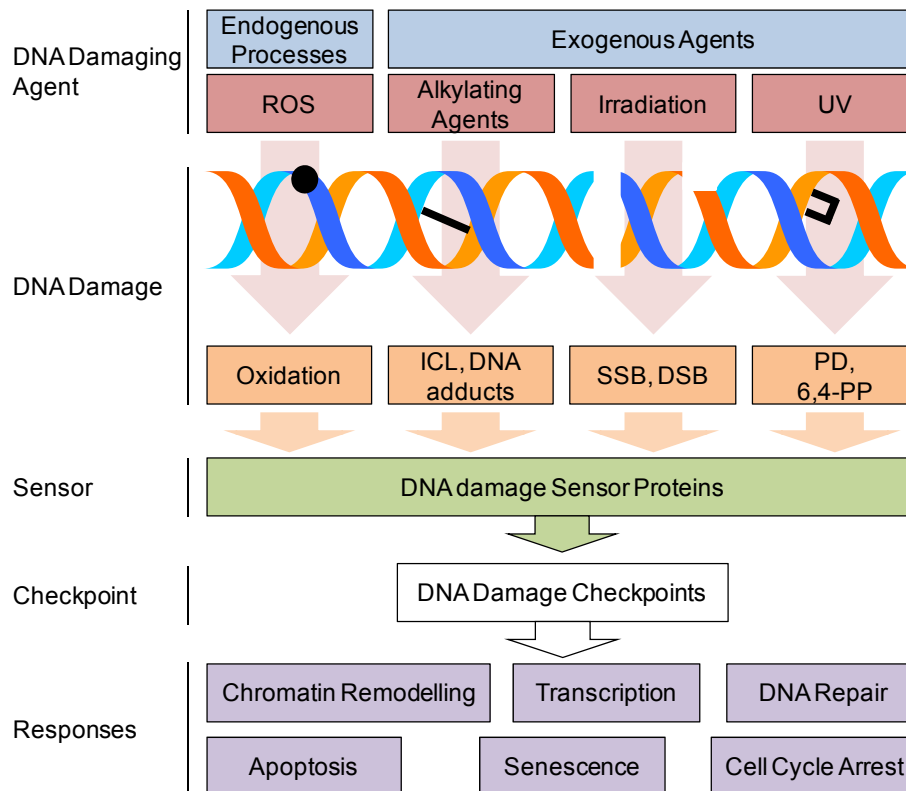
Figure 1: DNA damaging agents and the DNA damage response in mammalian cells. DNA damage can either arise by endogenous processes such as inflammation, reactive oxygen species, and DNA replication or can be induced by irradiation, UV light, or chemicals. The DNA damage inflicted is diverse, ranging from oxidized DNA bases to interstrand cross links (ICL), DNA adducts, single-strand (SSB) and double-strand breaks (DSB), pyrimidine dimers (PD) and 6,4-photoproducts (PP). These damages activate DNA damage sensors that in turn activate various cellular responses which are collectively termed the DNA damage response (DDR). In addition to pathways dedicated to removing DNA modifications by DNA repair, cells have evolved several pathways to activate cell cycle checkpoints, block or induce transcription and remodel local chromatin structures. Moreover, if the DNA damage is too severe, apoptosis is induced. Image modified from [3].

The DDR can be divided into different sequential steps. First, the DNA damage sensor proteins of the DDR sense DNA lesions and bind to the damaged site. Subsequently, these proteins transmit the signal of the DNA damage to effector proteins by modifying these proteins post-translationally. The phosphorylation, acetylation, ubiquitination and sumolyation of the effector proteins leads to an amplification of the damage signal. This signalling blocks cell cycle progression and allows the cell sufficient time to repair the damage. Additionally, the transcription of DDR-related proteins can be activated and the transcription of non-DDR-related proteins as well as replication blocked. Furthermore, the DDR promotes alterations in local chromatin structure by post-translationally modifying the histone components in order to reveal the underlying DNA sequence and to recruit the DNA repair machinery (Figure 1). Histone modifications include phosphorylation, acetylation, methylation, and ubiquitination [4]. Cells that are defective in any of the aforementioned response mechanisms are more sensitive to DNA damaging agents, and many of such defects cause cancer, neurological defects and pre-mature aging [1]. The first observation that inefficient DNA repair may increase the probability for cancer initiation was documented in 1968 by James Cleaver who reported that patients suffering from Xeroderma Pigmentosum (XP) have a defect in the DNA repair pathway responsible for repairing UV light-induced DNA damages [5, 6]. Since this discovery, several cancer types stemming from mutations in the DNA damage response have been identified (Table 1).

Table 1: Human disorders associated with defects in genome maintenance and enhanced cancer susceptibility. BER: base-excision repair; ICL: interstrand crosslink; MMR: mismatch repair; RECQ: recombination Q. Table taken and modified from [3]

Disease	Abbreviation	Mutated Gene(s)	Impaired Pathway
Ataxia telangiectasia	AT	ATM	DSB repair
Atypical Werner syndrome	WS	WRN	nuclear structure
Bloom's syndrome	BLS	BLM	RECQ helicase
Dyskeratosis congenita	DKC	DKC1, TERC1	telomere maintenance
Fanconi anemia	FA	FANCA, B, C, D1 (*BRCA2*), D2, E, F, G, I, J (*BRIP1*), L, M, N (*PALB2*), O (*RAD51C*), and P (*SLX4*)	ICL repair
Li–Fraumeni Syndrome		p53	many (p53 tumor suppressor inactivation)
Nijmegen breakage syndrome	NBS	NBS1	DSB repair
Rothmund–Thomson syndrome	RTS	RECQL4	RECQ helicase
Werner syndrome	WS	WRN	RECQ helicase, influencing nuclear structure, DSB repair, ICL repair, MMR, BER
Xeroderma pigmentosum	XP	XPA-G	NER

The plethora of DNA damages induced by endogenous processes and exogenous agents necessitates the coordinated action of multiple distinct DNA repair pathways. Some DNA lesions can be directly reversed; most DNA damages however require the activity of several repair proteins and processing of the damaged DNA [2]. During DNA mismatch repair, a single strand break is induced upon DNA detection and the single strand removed before polymerase and ligase enzymes fill up the gap and reseal the DNA. In base-excision repair, DNA glycosylase enzymes recognize the damaged DNA base, and initiate base removal and repair by nuclease, polymerase and ligase proteins. The nucleotide excision repair (NER) system recognizes helix-distorting base lesions, excises a 22-30 base oligonucleotide and thereby produces a single-stranded DNA that is subsequently repaired. Upon repair of the DNA damage, the DDR proteins dissociate from the damaged and repaired siteallowing the cell to re-enter the cell cycle [2]. If the DNA damage is too severe and cannot be efficiently repaired, cells can either engage in tolerance pathways which permit survival at the cost of mutagenesis [7] or induce cell death by apoptosis [1].

1.2   DNA double-strand break (DSB) repair pathways

For the repair of DSB, two different mechanisms are present in the cell: Non-Homologous End Joining (NHEJ) and Homologous Recombination (HR), which can be both subdivided into several pathways. The HR and NHEJ pathway differ mechanistically in several ways. HR repair is initiated by the generation of a single-stranded DNA (ssDNA) overhang, promoted by the Mre11-Rad50-Nbs1 (MRN) complex and CtBP-interacting protein (CtIP), whereas NHEJ-repair is activated by binding of the Ku70/80 heterodimer to the DSB site. Moreover, NHEJ-repair is considered error-prone, because free DNA ends are quickly rejoined without requiring homology, whereas in HR the homologous sister chromatid is used as a template for copying the genomic information to the damaged DNA strand. Hence, in contrast to NHEJ, HR is not active throughout the cell cycle, but is restricted to late S and G2 phase when the homologous DNA template is available after replication. Thus, NHEJ is the dominant pathway

for DSB repair in mammalian cells [2]. The exact cellular mechanisms determining which of the DNA repair pathway is used to repair a DSB are still not fully understood. However, recent work suggests that the choice between classical NHEJ and HR in replicating cells is regulated by DNA end resection [8]. Two regulators of the DSB repair pathway choice have emerged, namely the tumor suppressor proteins 53BP1 and BRCA1. 53BP1 contributes to NHEJ by interacting with chromatin at the DSB sites, thereby inhibiting DNA end resection and tethering the free DNA ends of the DSB in close proximity to promote their ligation [9]. Moreover, 53BP1 binds dimethylated histone H4 (H4K20me2), thereby stabilizing a chromatin conformation that is non-permissive to nuclease access and limiting DNA end resection of the DSB intermediates [10]. Since in S/G2 phase the level of HR increases, mechanisms exist that block or reduce the activity of 53BP1 at DSB sites in order to promote HR. A key protein in this process is BRCA1; however, the exact mechanisms of BRCA1-mediated 53BP1 inhibition and activation of HR at DSB sites are still unknown. It has been shown that BRCA1 stabilizes CtIP, a key molecule required for DSB end resection that promotes HR [11, 12]. Furthermore, BRCA1 may also influence the local abundance of H4K20me2, which prevents 53BP1 enrichment and NHEJ-repair [13]. Other results suggest that BRCA1 also increases the level of ubiquitylation at DSB sites which may contribute to a chromatin status promoting HR by granting access of the resection machinery [9].

The DNA repair activity and repair kinetics vary in the different compartments of the genome [14]. Densely compacted heterochromatin (HC) was shown to be refractory to DNA repair factors such as γH2AX. Furthermore, DSB sites in heterochromatin, gene poor or pericentromeric regions in the genome are repaired with slower kinetics than DSB in euchromatin [15]. Thus, it has been hypothesized that the slower DSB repair kinetics in HC are due to hindered access of DNA repair factors to these dense chromatin regions [14]. However, Jakob and colleagues demonstrated that H2AX is phosphorylated inside HC, but that the damaged site is subsequently relocated to regions of lower chromatin density to enable efficient DSB repair [16]. Another factor influencing the kinetics of DSB repair is the chemical DNA damage complexity [17]. Naturally-occurring DNA lesions tend to be isolated and homogeneously distributed and can thus be repaired efficiently by the DNA repair pathways. In contrast, clustered DNA damages such as those induced by ion particle irradiation have a reduced reparability compared to that of individual lesions [18].

1.2.1   The Non-Homologous End Joining (NHEJ) Repair Pathway

Non-Homologous End Joining (NHEJ) is the primary DSB repair pathway in the cell and rejoins DNA ends regardless of their homology, which increases the probability for genomic aberrations such as deletions, insertions and translocations [19]. Thus, NHEJ-repair is an important DNA repair pathway which contributes to both genome protection and mutations. Classical NHEJ-repair in mammalian cells was discovered in the 1980s when Mimori and colleagues first described the DNA end binding protein Ku [20]. However, it was not until 1994 when Stamato and colleagues revealed Ku's function in the DDR by showing that cells lacking Ku are sensitive to irradiation [21]. Moreover, several groups observed that the repair of DSB in NHEJ-deficient cells resulted in increased deletion frequencies and use of microhomologies at the repair junctions, which points to an additional NHEJ-like DNA repair mechanism. This second pathway was called microhomology-mediated end joining (MMEJ) or alternative NHEJ (aNHEJ), and occurs at approximately 10% of the frequency of normal or classical NHEJ (cNHEJ). Furthermore, it is less faithful than cNHEJ, since excessive deletions and chromosomal translocations are frequently found at the repaired DSB sites, with some of them leading to oncogenic transformation [22].

Both classical and alternative NHEJ-repair can be divided into four basic steps: (1) DSB recognition, (2) end binding and synapsis formation, (3) end processing, and (4) ligation [22] (Figure 2). Classical NHEJ is initiated by binding of the Ku70/80 heterodimer to free DNA ends at the DSB site. Upon recognition of the DSB, the Ku heterodimer translocates along the DNA from the DSB site in order to allow additional Ku dimers to bind to the DNA, which protects the DNA termini from end resection and acts as a scaffold for the recruitment of additional repair factors including ATM, DNA-PKcs and DNA polymerases. The Ser/Thr protein kinase ataxia-telangiectasia mutated (ATM) is a key regulator of the DNA damage response and activates a plethora of downstream effectors by phosphorylation. Among its target molecules, 53BP1, H2AX, p53 and DNA-PKcs are a few of the most intensively studied repair factors [23]. Another kinase, namely DNA-PKcs (DNA-dependent protein kinase, catalytic subunit) belongs to the same family as ATM and also phosphorylates several proteins such as H2AX, Artemis and XRCC4. Since these proteins are involved in downstream processes of DNA repair, their activation multiplies the DNA damage signal. In addition to activation of effector molecules for DNA repair, the DNA-PK-Ku70/80 complex undergoes autophosphorylation, resulting in a conformational change that increases the accessibility of the DNA for additional DNA processing enzymes and ligases. Aligned and compatible DNA ends can be directly rejoined by the DNA ligase IV. However, complex DNA ends such as those produced by ionizing radiation require processing by additional enzymes to prepare the DNA termini for ligation. One of these processing enzymes is Artemis, which is directly activated by DNA-PKcs. In complex with DNA-PKcs, Artemis removes single-stranded DNA (ssDNA) overhangs that contain damaged nucleotides. Then, XRCC4 interaction with the polynucleotide kinase/phosphatase (PNKP) results in the phosphorylation of 5' OH DNA ends and the removal of the phosphate molecule from the 3' DNA terminus, thereby forming compatible DNA ends for ligation. Subsequently, DNA polymerases fill in the gaps at the DSB site. In the final step of cNHEJ, the two DNA ends are rejoined by the XRCC4-ligase IV-XLF complex. XRCC4 stabilizes ligase IV and stimulates its ligation activity. XLF stimulates the ligation of non-cohesive DNA ends, and is essential for gap filling by the DNA polymerases, suggesting that it plays an important role in the alignment of DNA ends and thus maintaining stability of the broken DNA ends. Upon repair of the DSB, the repair proteins dissociate or are removed from the DNA [22].

In contrast to cNHEJ, the mechanisms and proteins involved in alternative NHEJ remain poorly understood. A characteristic feature of aNHEJ is the use of a 0-10bp microhomologous sequence during the alignment of the broken DNA ends. Furthermore, distinctive DNA repair signatures found at DSB sites repaired by aNHEJ imply that enzymes involved in aNHEJ promote end resection and include nucleases and ligases. Several proteins have been identified to be involved in end resection (Mre11, CtIP) and ligation (PARP1, DNA ligase III) [22] (Figure 2).
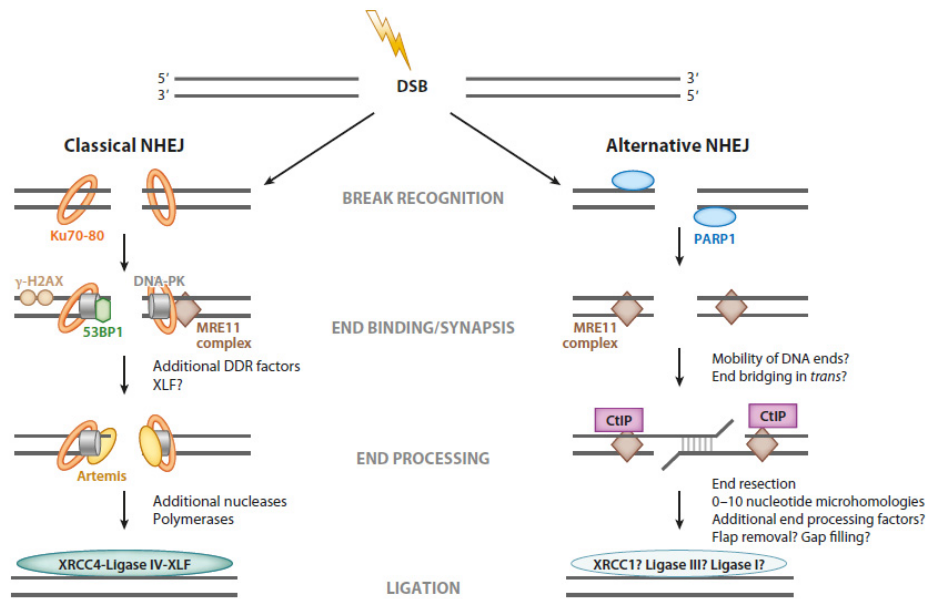
Figure 2: Classical and alternative NHEJ-repair in mammalian cells. During classical NHEJ-repair, DSB are recognized by the Ku70/80 heterodimer, which recruits other repair factors. DNA-PKcs multiplies the damage signal by phosphorylating several downstream targets such as Artemis, which processes the DNA ends for subsequent ligation by the XRCC4-Ligase IV-XLF complex. In contrast to cNHEJ, most repair factors involved in aNHEJ remain unknown. However, PARP1 is known to recognize the DNA break and Mre11 and CtIP are involved in end resection. Image modified from [22].

1.2.2 Methods for the detection of DNA damage and DNA repair activity

The most frequently used method for the detection of DSB and DNA repair activity is immunostaining of DNA repair proteins with fluorescently-labeled antibodies and subsequent visualization by microscopy. Several DNA repair factors that assemble at the DSB site form large complexes that can be microscopically visualized. Among the most prominent molecules are DNA-PK, 53BP1, and γH2AX. The histone variant H2AX is incorporated into the nucleosomes at DSB sites and becomes phosphorylated at its serine residue at position 139 (Ser139) by the kinase activity of DNA-PK, ATM and ATR. Upon phosphorylation, γH2AX forms large foci, termed radiation-induced foci (RIF), which can span up to several megabases around the DSB site in order to facilitate DNA repair and recruit other DNA repair factors. When the DNA damage has been repaired, γH2AX is dephosphorylated and disassembles from the DSB site [2].

Another approach often used to study DNA repair sites and genomic aberrations is whole genome sequencing. In 2005, Roche/454 Life Sciences introduced the first commercially-available next generation sequencing technology [24] that is capable of generating 80–120 Mb of sequence in 200- to 300-bp reads. The process is divided into several sequential steps, starting with the isolation of genomic DNA, followed by a fragmentation step, the ligation of PCR adapters and separation into single DNA strands. The single-stranded DNA fragments are subsequently captured on magnetic beads under conditions that favor one DNA fragment per bead. The beads are incorporated into oil droplets containing all PCR reagents that allow amplification, thereby producing millions of copies of a unique DNA template. Subsequently, the droplets are broken up, and each bead is deposit in a single picoliter-sized fiberoptic well for pyrosequencing. Pyrosequencing is a sequencing-by-synthesis approach in which pyrophosphate (PPi) is released from the nucleotide by the DNA polymerase during nucleotide incorporation. The ATP sulfurylase converts the free PPi molecules into a substrate for the luciferase, which results in light emission and thereby determines the DNA sequence. Similar to the Roche/454 sequencing

platform, Illumina sequencing is also a sequencing-by-synthesis approach. However, it does not require a pre-amplification step of the DNA template, since these are directly bound to the surface of a flow cell. The Illumina flow cell is densely populated with forward and reverse PCR primer adapters which are complementary to adapters ligated to the DNA templates. The DNA templates bind to the forward and reverse primer adapters of the flow cell and form a bridge that serves as the substrate for amplification. For sequencing, modified nucleotides with reversible terminators are used. This allows a single nucleotide to be incorporated in each sequencing cycle [25]. Thus, Illumina sequencing has a higher accuracy than the 454 platform. Each nucleotide carries its own chemically-cleavable fluorescent dye at the 3' OH terminus. The cleavage of the terminator results in fluorescence that is recorded by a camera. Recent advancement of the reversible-terminator sequencing technology is the MiSeq sequencing by Illumina, which produces 2 x 300 paired-end reads in a single run, allowing small genome sequencing and assembly. Further, it allows processing of more samples and generates more reads per run than previous Illumina sequencing platform [26].

1.3 Radio- and chemotherapy

1.3.1 Types of ionizing radiation

Ionizing radiation (IR) is defined as the radiation that has enough energy to remove electrons from atoms, and thereby ionizes them [27]. The IR energy is released inside the cell and is absorbed by various molecules such as proteins, lipids and the DNA and ionizes them. Radiation can be either directly or indirectly ionizing. During direct ionization, the radiation beam hits the target molecule in the cell and disrupts its structure. In contrast, during indirect ionization, the energy of the radiation beam is transferred to an electron of an atom, which is not part of the target molecule. This electron becomes excited, and, due to its higher energy, is lost from the atom. The ionized atom forms a radical that causes damage to different molecules inside the cell [7, 27] such as the DNA.

In general, two types of IR can be distinguished: photon and hadron. X-rays and α-rays belong to the group of photon radiation. X-rays are generated by rapid stopping of highly accelerated electrons, have a wavelength between 0.01 and 10nm and energies in the range of 10^2 to 10^5 electron volt (eV). In contrast, hadron radiation consists of particles including protons, neutrons, and heavy charged ions such as carbon (^{12}C) and ion (^{56}Fe) atoms [7]. The two different types of radiation differ in their biological effectiveness which is expressed by the relative biological effect (RBE) value. The RBE is defined as the dose of a particular radiation divided by the dose of X-rays required for an equal biological effect [27]. The higher the RBE for a specific radiation type, the more damage is induced per unit of energy deposited in the cell/tissue. In comparison to photons, carbon ions have an increased RBE, which is calculated between 2 and 5 depending on the cell type [28]. The different RBE of photons and hadrons result from the distances travelled by the IR in tissues and the pattern of ionizing events along the track. Photons are only sparsely ionizing, dispose their energy in atoms spaced by several hundred nanometers apart and cannot penetrate the tissue deeply. Thus, ionization events occur most at the tissue surface. In contrast, hadrons leave a dense trail of ionized atoms along their path that are spaced only about a tenth of a nanometer apart [29]. Moreover, hadron particles enter into the tissue deeply and only release small amounts of energy at the beginning of their route. At the end of their path, these particles rapidly decelerate and release high amounts of energy before stopping completely [30]. Since hadrons deposit their energy at a depth proportional to the energy of the charged particle [7], hadron irradiation allows a more precise deposition of energy inside the target tissue than photon irradiation.

1.3.2   Radiation-induced DNA damages

In order to determine the density of ionizations along the radiation track and thus the extent of radiation damage the Linear Energy Transfer (LET) is used as a measure. The LET is defined as the average energy deposition per unit length of the ionizing track, and its unit is keV/µm. Photons have a low LET, whereas in general hadrons have a high LET. Both low and high LET radiation induce a variety of DNA lesions including DSB, and the complexity of the DNA damage increases with increasing LET. Isolated DNA damages mainly occur after low LET irradiation, are spaced at a larger distance from other damages and can be repaired efficiently by the DNA repair machinery [29] (Figure 3).
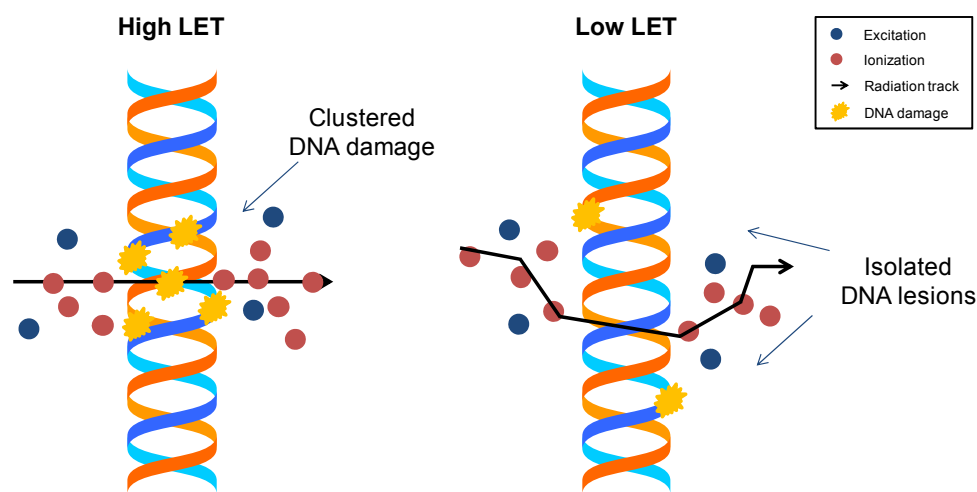
Figure 3: Induction of ionizations and DNA damage by high- and low LET irradiation. High LET radiation causes more ionizations and excitations at high density than low LET radiation that creates more isolated DNA damages.

In contrast, high LET irradiation also induces complex clustered DNA lesions that include two or more lesions within 20bp or one helical turn (Figure 3). Amongst others, these DNA damages comprise abasic sites (AP), damaged bases and single-strand breaks (SSB) that can be converted into potentially lethal DSB when the DNA repair is disrupted. Whether a non-DSB cluster is converted into a complex DSB depends on the type of lesion induced, the distance separating the DNA lesions and whether the additional lesions disrupt the binding of DNA repair enzymes to other DNA damage sites [7]. Thus, in order to reduce the probability of DSB formation during damage repair, opposing DNA lesions are repaired sequentially [31]. Inefficient DNA repair of DNA damage clusters can result in increased levels of mutagenesis. The overexpression of certain DNA repair proteins such as the DNA glycosylase responsible for repair initiation can convert an oxidative DNA damage to a DSB (Figure 4). Non-DSB clusters consisting of opposing AP sites are converted to DSB sites by the activity of the exonuclease Ape1, which stimulates cNHEJ-repair or Ku-independent DNA repair activity (Figure 4). Complex DNA damage clusters consisting of oxidized and abasic DNA molecules can be converted into a complex DSB by the nuclease ApeI, making it highly difficult for the cell to efficiently repair the damage [32]. Such complex clustered DNA damages are mainly induced by high LET radiation, are difficult to repair, have slow repair kinetics, and are thus considered to be more detrimental to the cell than randomly distributed DNA lesion induced by low LET radiation [33], thereby explaining the increase in RBE of high LET. In addition to the formation of complex clustered DNA lesions and radiation-induced DSB, replication-induced DSB sites are formed after ionizing radiation [34, 35], when a replication fork meets an unrepaired non-DSB clustered damage site [35, 36].
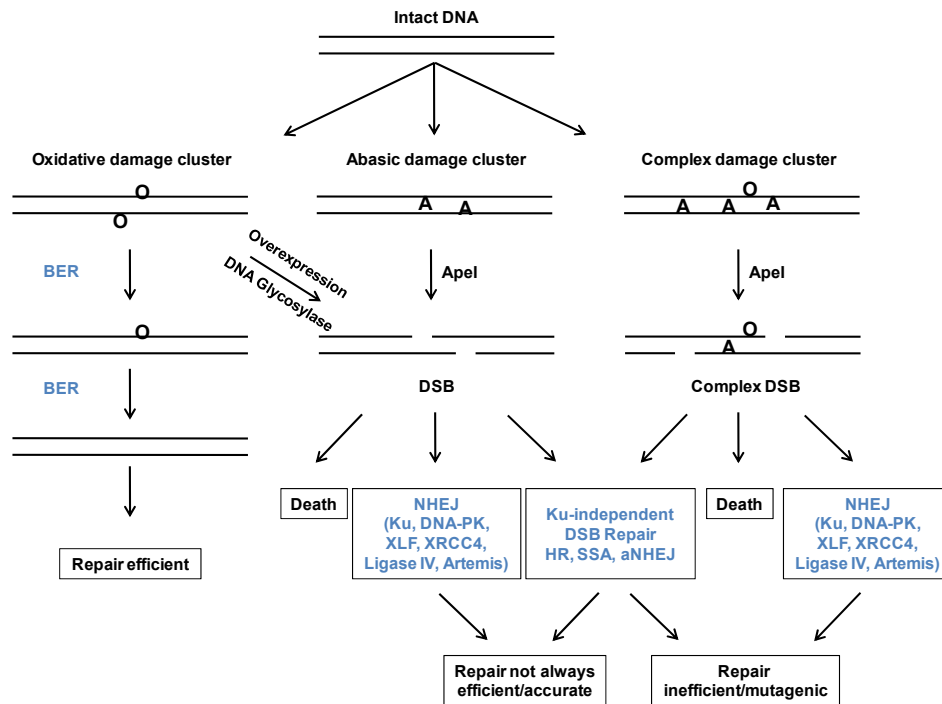
Figure 4: Overview about repair of clustered DNA damages and repair outcomes. Two oxidative DNA damages (O) can be repaired efficiently by sequential base excision repair activity. However, replication through this damage can increase the mutation frequency of the base damage. Furthermore, in the presence of DNA glycosylases, two oxidative lesions can be converted into a DSB. Non-DSB, abasic damage clusters are converted into DSB by the nuclease ApeI. DSB sites can induce cell death, or be repaired by error-prone DSB repair pathways such as classical and alternative NHEJ. Complex DSB clusters resulting from high LET irradiation are more difficult to repair and can result in increased levels of mutagenesis or cell death. A: abasic site; O: oxidative damage; NHEJ: non-homologous end joining; BER: base excision repair; HR Homologous Repair, SSA: Single-Strand Annealing

1.3.3   The topoisomerase family

During transcription and replication, the DNA is locally unwound to allow the transcriptional and replication machinery to proceed. This, however, leads to DNA supercoiling and increased tension in adjacent genomic regions, stalling of the replication and transcription machinery as well as formation of abnormal DNA structures. Topoisomerases are enzymes involved in the unwinding of the DNA during transcription and replication. Upon replication, the replicated DNA strands are interlinked and need to be unlinked before cytokinesis can proceed. Failure to disentangle the two DNA strands can lead to genomic aberrations and induction of cell death. In order to protect the cell from replication-induced damage, DNA topoisomerases also assist in the segregation of the daughter chromosomes before cell division [37].

The human genome encodes six topoisomerases, which are divided into type I and II. Type I topoisomerase (TOP1) only cleaves a single DNA strand, whereas the type II topoisomerase (TOP2) can induce DNA double strand breaks. Both types of topoisomerases cut the phosphodiester backbone of the DNA by a nucleophilic attack from the catalytic tyrosine at the catalytic center of topoisomerase. This reaction is reversible, and the DNA sequence remains unchanged upon religation. Mammalian cells have two TOP2 isoenzymes, namely α and β, which function as homodimers. The expression of TOP2α is closely linked to the cell cycle and the expression increases two- to three-fold during G2/M phase. TOP2β on the other hand, is always expressed in both cycling and non-cycling cells. The TOP2 isoenzymes have low sequence selectivity, but they preferentially recognize and bind DNA knots, supercoils and interlinks [37].

1.3.4   Topoisomerase 2 catalytic cycle and targeting by anticancer drugs

The TOP2 catalytic cycle is initiated by binding of the TOP2 homodimers to the gate (G) DNA segment, through which the transported (T) DNA segment is passed. After binding to the G segments, TOP2 binds the T DNA segment and two ATP molecules. Subsequently, the homodimer changes its conformation to a closed clamp form, called the TOP2 cleavable complex. Upon binding of Mg^{2+}, the two tyrosyl residues of the TOP2 homodimers attack the phosphodiester backbone in the opposite DNA strands of the G segment to induce a DSB. In this stage, each of two topoisomerase monomers is covalently linked to the 5'-terminus of an enzyme-generated DSB. Then, the T segment is rapidly passed through the G segment and released from the enzyme. After strand passage, the G segment is religated by TOP2, and ATP hydrolysis converts the closed clamp conformation of the enzyme to the open form to release the G segment [37] (Figure 5).
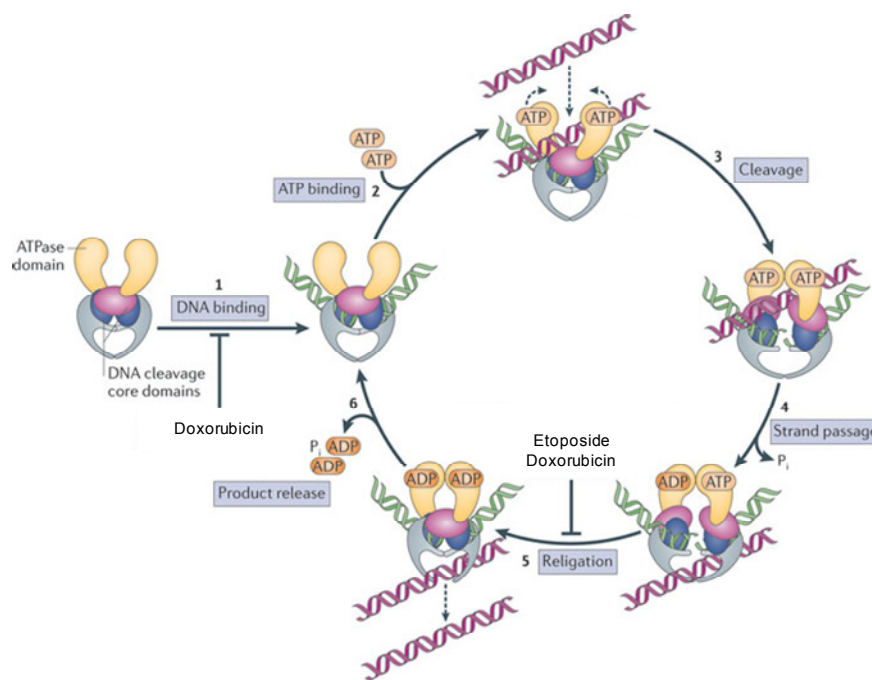
Figure 5: The TOP2 catalytic cycle (1) Upon recognition of a DNA crossover region, TOP2 binds to the DNA (G segment, green). (2) ATP binding leads to a conformational shift to a closed clamp form and binding of the T segment (purple). (3) The G segment is subsequently cleaved and (4) the T strand passes through the G segment. (5) TOP2 religates the G segment and (6) returns to its open conformation after ATP hydrolysis. At low concentrations (<1μM), doxorubicin acts like etoposide and inhibits TOP2 DNA religation, whereas at concentrations above 10μM doxorubicin intercalates into the DNA, thereby blocking TOP2 binding. Figure modified from [38].

Since failure to disentangle two interlinked DNA strands can lead to the induction of cell death, TOP2 enzymes are frequently targeted in cancer chemotherapy. Generally, anti-cancer drugs targeting TOP2 can be classified according to their mechanisms. Molecules that inhibit the religation of the G segment such as etoposide and doxorubicin are termed TOP2 poisons, whereas TOP2 inhibitors such as genistein and azatoxins enhance the formation of the TOP2 cleavable complex. In addition to stimulating DSB formation, TOP2 poisons and inhibitors also induce SSB, affect nuclear processes such as transcription and replication by intercalating into the DNA and by generating reactive oxygen species (ROS). Doxorubicin belongs to the family of anthracyclines, and, at high concentrations (10μM), intercalates into the DNA, which alters the DNA structure and prevents TOP2 from

binding to the DNA. At concentrations below 1µM, doxorubicin acts like etoposide and stabilizes the cleavable complex, thereby prolonging the half-life of the TOP2-DNA-intermediate and increasing the possibility of DSB formation. In order for DSB repair to occur, the covalently attached TOP2 enzyme is removed by cellular end-processing enzymes. The DNA fragment linked to the peptide is excised by nucleases such as the MRN complex, CtIP or Artemis, and the DSB is repaired by NHEJ [39-41].

1.3.5   Cancer therapy-induced delayed genomic instability

In progenitors of radiation-surviving cells, genomic changes can arise several generations of the initial DNA insult. This phenomenon is known as delayed radiation-induced genomic instability, and is characterized by the expression of several radiation-induced effects including apoptosis, gross chromosomal rearrangements, aneuploidy as well as gene mutations and amplification [31]. The mechanisms that initiate and drive delayed genomic instability in progenitor cells are not fully understood yet. However, experimental evidence suggests that increased levels of oxidative stress contribute to radiation-induced genomic instability, since cells with mitochondrial dysfunction and cells exposed to hydrogen peroxide initiate delayed instability, and treatment with antioxidants can greatly reduce these effects [42, 43]. Moreover, it also possible that delayed genomic instability is caused by delayed DSB induction and subsequent misrepair by NHEJ. Delayed DSB induction and illegitimate joining of DNA ends can generate dicentric chromosomes which block segregation during mitosis, thereby inducing additional DSB and accumulation of mutations [44]. Since radiation-induced genomic instability is transmitted through the genome for several generations, it was speculated that the initial ionization event is memorized in the genome [45]. Several studies now suggest that radiation induces alterations in local chromatin structures that can lead to increased levels of replication stress and DSB that can initiate genomic instability in radiation surviving cell populations [44, 45]. Despite these results, the functional consequences of delayed genomic instability on carcinogensis and radiosensitivity have not been described to date [44]. Nonetheless, since the genomic effects of delayed genomic instability are similar to those induced directly by irradiation, it is assumed that these genomic changes initiate or drive carcinogenesis [46-48]. Delayed induced genomic instability has mostly been observed and studied in radiation surviving cells, but, persistent destabilization of the genome following chemical treatment was also reported. For example, bleomycin and neocarzinostatin were equally efficient in inducing delayed genomic instability in progenitors of chemotherapy-surviving cells [49]. Moreover, anti-topoisomerase drugs have been reported to induce structural and numerical chromosome aberrations in surviving cells [50]. Thus, both irradiation and chemotherapy can induce delayed genomic instability, which potentially leads to cancer-therapy resistance and therapy-induced carcinogenesis.

1.4   Lentiviruses

1.4.1   History and phylogeny of lentiviruses

In 1911, Peyton Rous showed that cancer could be transferred from chicken suffering from sarcoma to healthy chicken by using an ultra filter extracts [51]. Later, it was discovered that viruses in the extract were the cancer-inducing agent, and this virus was named after its discoverer Rous Sarcoma Virus (RSV). In 1970, Howard Temin, Satoshi Mizutani and David Baltimore discovered the enzyme reverse transcriptase in RSV, which enables retroviruses to convert their single-stranded RNA genome into double-stranded DNA [52, 53]. In the following years, another characteristic feature of retroviruses, the integration of the proviral genome into the host cell's

genome was discovered [54], which is mediated by the viral integrase enzyme. In the following decade, the first human oncogenic retrovirus (HTLV) [55] and the immunodeficiency causing virus HIV [56] were discovered. These viruses belong to the retroviridae family which is divided into two subfamilies: orthoretroviridae and spumaretroviridae. The subfamily orthoretroviridae further consists of six genera (alpharetrovirus, betaretrovirus, gammaretrovirus, deltaretrovirus, epsilonretrovirus, lentivirus) whereas the spumaretroviridae only has one genus (spumavirus) [57]. Each genus is further subdivided into different types of subspecies. Lentiviruses belong to the subfamily orthoretroviridae and are single-stranded, positive-sense RNA viruses that have a diploid genome with a length of 7-12kb. The virions are enveloped and 80-100nm in diameter. Moreover, these viruses are characterized by two viral enzymes: reverse transcriptase and integrase, which convert the RNA genome into double-stranded DNA and integrate it into the host cell genome. Another characteristic of lentiviruses is their ability to infect both dividing and non-dividing cells. Infection with lentiviruses can result in different diseases such as immunodeficiencies, anemia and encephalitis [58]. Since in this dissertation a HIV-1-derived lentiviral vector was used, further descriptions are based on the HIV-1 genome.

1.4.2   The lentiviral genome

The lentiviral genome exists as two single-stranded positive-strand RNA molecules and has a 5' cap structure and a 3' polyadenylation. The proviral genome is flanked by two, 640bp long terminal repeats (LTR) that are required for transcription, reverse transcription and integration into the host cell genome. Each LTR is made of a U3, R and U5 region and contains several promoter and enhancer elements. The lentiviral genome further contains a primer binding site (PBS) necessary for binding of a tRNA that initiates the reverse transcription of the RNA genome. Located in 3' to the PBS is the packaging and dimerization signal psi (Ψ), which mediates packaging of the RNA genome during virion assembly. At the 3' LTR, the polypurin tract (3'PPT) required for the synthesis of the second DNA strand during reverse transcription is encoded. The viral core proteins, enzymatic proteins, and the envelope glycoproteins shared by all retroviruses are encoded in the *gag*, *pol* and *env* gene. Complex retroviruses, such as lentiviruses, also contain six additional genes (*tat, rev, vif, vpr, vpu,* and *nef*) important for HIV infection and pathogenicity, making it a total of nine open reading frames (Figure 6) [58].

A   **HIV RNA Genome**

↓ Reverse Transcription

B   **Viral DNA Genome**
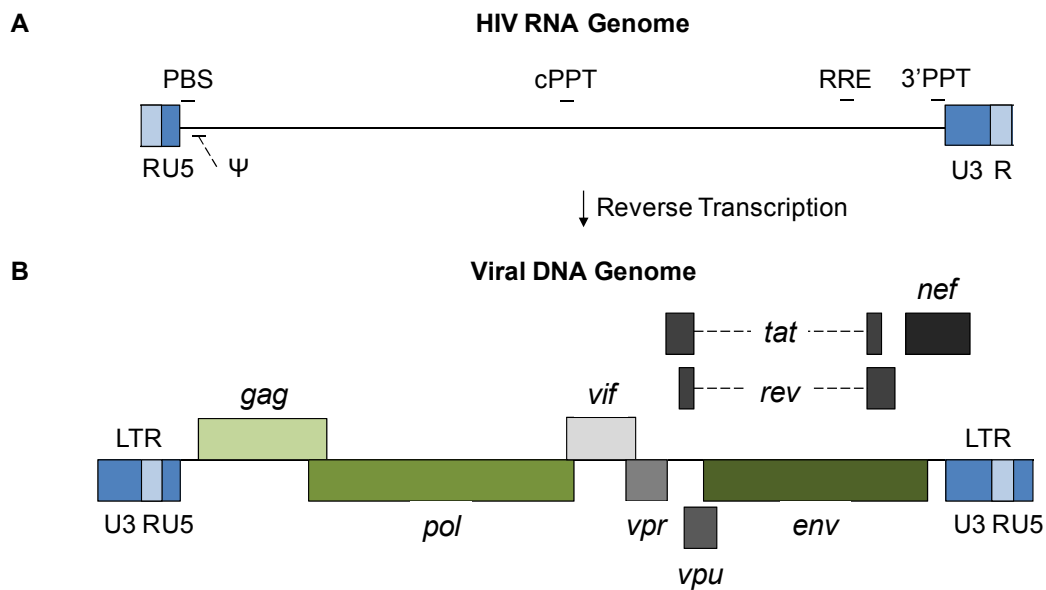
Figure 6: The RNA and proviral genome of HIV-1. (A) Structure of the RNA genome. The 5' and 3' end are modified, carrying a cap structure and a polyadenylation, respectively (not shown). At the primer binding site (PBS) a tRNA for the initiation of reverse transcription is bound (not shown). Psi (Ψ) is the packaging signal, U3, R and U5 mark the end of the HIV genome and form the long terminal repeats (LTR) after reverse transcription. The central polypurine tract (cPPT) is important for transportation of the provirus into the nucleus. (B) Schematic of the proviral genome. During reverse transcription, the U3 region at the 3' end is copied to the 5' end and the 5' U5 region to the 3' end to form the long terminal repeats. The location of the nine lentiviral genes is shown.

The *gag* gene encodes four viral capsid proteins that are derived from a single precursor Gag protein called p55 and that are necessary for the formation and release of viral particles. For lentiviral replication, the viral protease, reverse transcriptase and the integrase are required, which are encoded in the *pol* gene. The *gag* and *pol* genes are synthesized from the same mRNA transcript by a ribosome frameshifting near the 3' end of the *gag* gene. The resulting Gag-Pol precursor protein is processed by the viral protease inside the virions. The *env* gene encodes the viral surface protein gp120 and transmembrane gp41 protein. Similar to the Gag and Pol proteins, these mature proteins are processed from a single precursor protein, gp160, by proteolytic activity. The Gag, Pol and Env proteins are essential for the formation of mature virions and the infection of target cells. One of the accessory proteins encoded in the genome of lentiviruses is the transactivator Tat, which binds to the Tat response element (TAR) located in the LTR and plays a critical role for the transactivation and transcription of proviral DNA as well as the lentiviral replication. Besides, the Rev protein promotes the nuclear export of lentiviral RNA by binding specifically to the Rev response element (RRE) located in the *env* region, thereby increasing the half-life of viral mRNAs. Both Tat and Rev are translated from early transcribed mRNA [58].

1.4.3   The life cycle of lentiviruses

The life cycle of lentiviruses is divided into early and late stages. In the early stages, virus entry, reverse transcription of the RNA genome and integration into the genome occur. The late stages of the lentiviral life cycle are characterized by the transcription of the integrated viral genome, translation of viral transcripts into proteins required for viral assembly and budding from the host cell (Figure 7).

The lentiviral life cycle begins with the recognition and binding of the viral surface protein Gp120 to its primary receptor CD4 and the co-receptor CXCR4 on T-lymphocytes or CCR5 on macrophages and dendritic cells. Binding initiates a conformational change in the viral transmembrane protein Gp41 required for the fusion of the viral membrane with the cell membrane. Upon viral entry into the cell, the capsid proteins are uncoated, and the lentiviral genome, reverse transcriptase, protease and integrase are released into the cytoplasm. Subsequently, the RNA genome is converted into the complementary double-stranded DNA by the enzymatic activity of the reverse transcriptase [58]. The viral enzyme integrase (IN) specifically binds to the U3 region in the 5' LTR and the U5 region in the 3'LTR of the proviral cDNA. IN removes nucleotides from the 3' ends of the viral DNA beyond a conserved 'CA' dinucleotide, thereby creating two single-stranded 5' overhangs. The processed proviral DNA is subsequently imported into the nucleus, where the cellular DNA is cleaved by IN, the 5' dinucleotide overhang in the viral DNA is removed and the 3' DNA ends joined to the 5' ends of the genome [59]. Following stable integration, the host cell transcriptional machinery is hijacked to transcribe the provirus. The 5' LTR has basal promoter activity that is sufficient to initiate transcription by cellular RNA polymerase II. The first mRNA transcripts are multi-spliced mRNAs that are exported from the nucleus into the cytoplasm where they are translated into Rev, Tat and Nef proteins, required for the initiation of subsequent events in viral transcription. The Tat protein binds to the TAR element at the 5' end of viral mRNAs, leading to the transactivation and transcription of other genes from the viral DNA. In addition to the Tat protein, the Rev protein binds to the Rev-responsive element on viral mRNA transcripts and removes them from the splicosome, which leads to the export of singly-spliced and unspliced mRNA and viral genomes into the cytoplasm. Singly-spliced mRNAs are translated at the ribosomes into the proteins Env, Vif, Vpr, and Vpn, whereas unspliced mRNA are translated into the Gag-Pol precursor protein. Following transcription and translation, the viral proteins and the genomic RNA assemble at the cell membrane, where viral Env proteins are integrated into the cell membrane. The unspliced RNA genome copies are then packaged into the virions and released from the cell. Subsequently, the multimerization of Gag and Gag-Pol precursor proteins activates the viral protease, converting immature to mature HIV virions (Figure 7) [58].
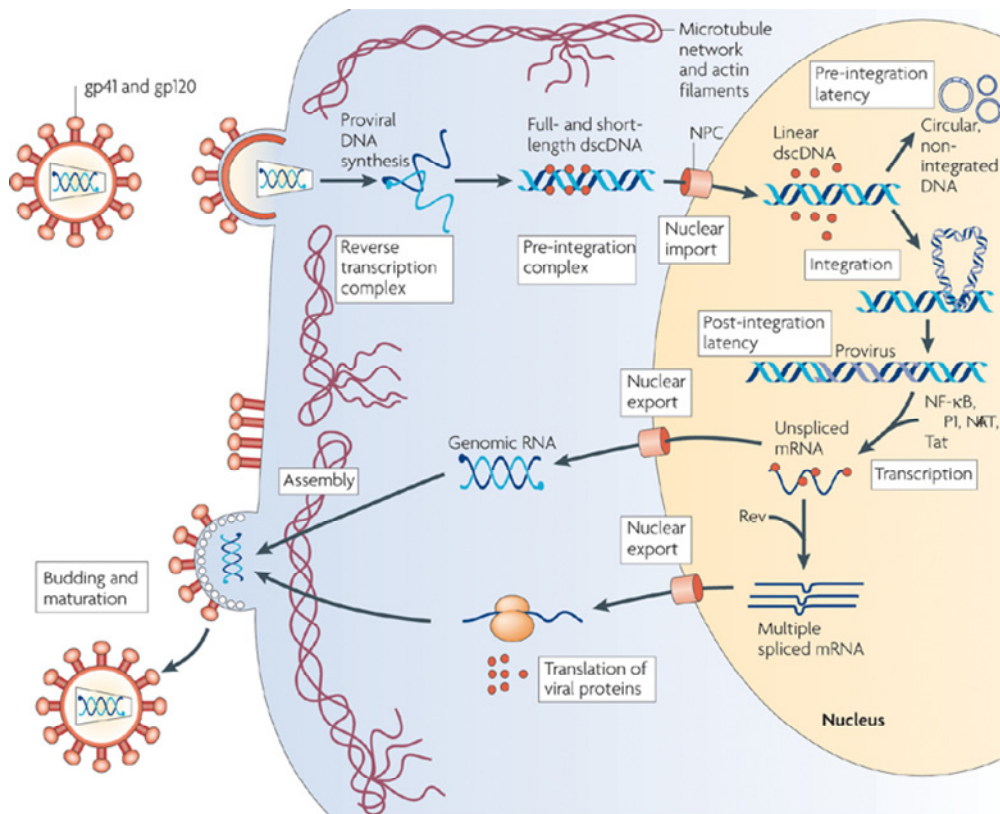
Figure 7: Lentiviral life cycle. In the early stage of viral infection, the virus attaches to the target cell and fuses with the cellular membrane to release the capsid into the cell. The capsid is subsequently uncoated, the RNA genome reverse transcribed, imported into the nucleus and integrated into the genome. After integration, the lentiviral genes are transcribed and exported into the cytoplasm. The viral mRNA is translated, and the lentiviral RNA genome and the proteins assemble into virions at the cell membrane. NPC: nuclear pore complex. Image taken from [60]

The integration profile of HIV in the genome of their host cells is non-random. HIV preferentially integrates into genomic regions with a high gene density and particularly into coding sequences. However, integration upstream of the transcription start site is not favored [61, 62]. Furthermore, HIV integration is influenced by the transcriptional activity: the integration frequency into transcriptionally active genes is increased [63], but genes with high transcriptional activity are less favored [61]. Regions rich in the dinucleotide CpG, termed CpG islands, commonly correspond to gene regulatory regions that frequently contain promoter and enhancers. For HIV, these regions and their surroundings are disfavored [61]. Hence, gene-dense and transcriptionally-active genomic regions, but not CpG islands are preferred sites of HIV integration.

1.4.4 Structure of lentiviral vectors for gene therapy

Because of their ability to infect both dividing and non-dividing cells as well as stably integrate into the genome, lentiviruses are used as vehicles to introduce transgenes into target cells for more than 20 years [58]. In order to generate replication-incompetent viruses for gene transfer, the lentiviral genome sequence was modified. Today, plasmid vector systems are used in which all elements required for production of functional virions (*trans*-acting factors) of the viral genome are divided onto four plasmids to reduce the likelihood of recombination and formation of replication-competent viruses. In principal, the first plasmid encodes the packaging genes *gag* and *pol*. In the second plasmid, the *rev* gene that is required during lentiviral transcription

is encoded. On the third plasmid, the Env protein is often replaced by the VSV-G protein (vesicular stomatitis virus - protein G) to increase the tropism of the lentivirus. In order to prevent that the three plasmids containing the *trans*-acting factors are packaged into mature virions and that replication-competent viruses form, the packaging signal psi and the LTR sequences have been cloned into a fourth plasmid, called the transfer vector. This vector is packaged into the virions, and contains the sequence for reverse transcription and the transgene. Moreover, to further reduce the likelihood of recombinations, the U3 region in the LTR containing the promoter and enhancer elements is truncated [64]. Furthermore, the requirement of lentiviral replication for the Tat protein was eliminated, by replacing the U3 region in the 5' LTR with a constitutively active promoter from the cytomegalovirus (CMV) or RSV. Vectors with these LTR modifications are called self-inactivating (SIN), because they lack LTR promoter activity following integration, thereby reducing the likelihood of transcription activation of proto-oncogenes located in vicinity to the lentiviral integration site. Moreover, increased transduction efficiency and transgene expression *in vitro* and *in vivo* was achieved by placing a central polypurine tract (cPPT) sequence downstream of the RRE in the transfer vector. Another genetic element in the transfer vector is the posttranscriptional regulatory element WPRE (woodchuck hepatitis virus post-transcriptional element), which increases the amount of transgenic mRNA. For production of functional viral particles, the four plasmids are co-transfected into a packaging cell line in which the genomic information is transcribed and translated into mature viral particles, containing the transfer vector.

In addition to integration-competent lentiviruses, episomal-remaining, non-integrating lentiviral vectors have been developed, which enable transient transgene expression in infected cells. These lentiviruses are termed integrase-deficient lentiviruses (IDLV) and are generated by introducing a point mutation at the catalytic core of the integrase gene that changes the amino acid at position 64 from aspartic acid to valine (D64V). This modification was shown to inhibit integration by up to four logs compared to integrase-competent lentiviruses [65, 66], and several studies suggest that the few observed integration events are mediated by cellular DNA repair mechanisms [65]. Indeed, integrated IDLVs showed a high frequency of LTR deletions, a key signature of NHEJ-repair activity, and occurred at sites of DNA damage and repair. Additionally, blocking ATM, a key player in DNA repair, also prevented integration of IDLV at DSB sites [67].

1.5   Scientific aims

Current methods for DSB detection have several limitations, which hamper the analysis of induced and repaired DSB as well as genomic instabilities in irradiated cells: 1) Immunostaining of DNA repair proteins is an indirect detection method that does not enable the identification of DSB sites at single-nucleotide resolution, 2) DNA repair proteins form microscopically-visible foci which disassemble upon DNA repair and do not allow DSB site analysis in surviving cell populations, 3) Immunostaining does not give any information on the relationship between DSB locations, DNA repair pathway choice and cell fate decision. Even though the high-throughput sequencing platforms are frequently used to study the mutational spectra in various cancer types [68], the high costs for sequencing large numbers of whole genomes and the analysis of sequencing data from heterogeneous cell populations make this method impractical for the identification of low frequency genomic events. Moreover, it does not provide any information, whether the identified mutation has functional consequences on cell survival and therapy resistance. Hence, there is little information on how radiation-induced and repaired DSB sites are distributed, how radiation-induced DNA damage is being survived and how these damages induce radioresistance and cell transformation. Analyzing the genomic distribution of radiation-induced DSB in irradiated and expanded cells at single-nucleotide resolution will help to improve our understanding of the

mechanisms of radiation-induced genomic instability and radiation resistance. In this thesis, a new methodology to capture induced and repaired DSB sites in radiation-surviving cell populations has been used to study DSB site distribution. Target cells become transduced with an integrase-deficient lentivirus (IDLV) that carries a point mutation in its integrase gene, preventing integrase-mediated integration of the proviral DNA into the host genome of the cell. Upon DSB induction by irradiation, these IDLV DNA molecules serve as molecular tags, which become stably integrated into the genome by the cellular NHEJ-repair activity, thereby marking the DSB site *in vivo*. The integration of IDLV at induced and repaired DSB sites can be followed by PGK promoter-driven EGFP expression from integrated IDLV by flow cytometry. The DSB repair sites can be amplified and identified using LAM-PCR and deep sequencing (Figure 8).
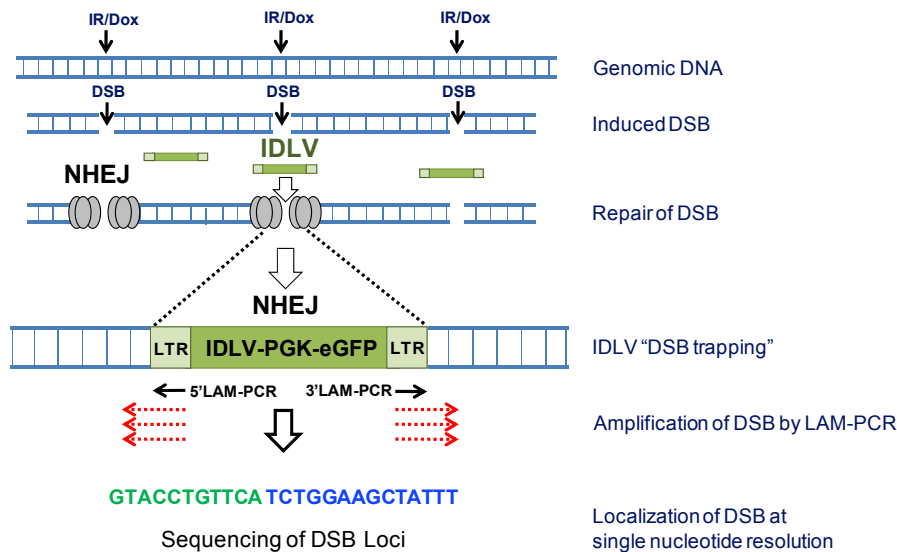
Figure 8: Proposed mechanism of IDLV-based DSB capturing. The double-stranded DNA bait delivered into cells by IDLV is captured by the cellular NHEJ-repair machinery at DSB sites. This leaves a stable genetic tag which enables the mapping and tracking of radiation (IR)- or doxorubicin (Dox)-induced and repaired DSB in the genome of treated cells. The frequency of DSB tagging can be followed by PGK promoter-driven EGFP expression during expansion. Localization of captured DSB is performed by amplifying the vector-genome junction using 5' and 3' LAM-PCR and deep sequencing. DSB: DNA double-strand breaks, IR: irradiation; Dox: doxorubicin; IDLV: integrase-deficient lentivirus; NHEJ: Non-Homologous End Joining, eGFP: enhanced green fluorescent protein; PGK: Phosphoglycerate Kinase 1 promoter; LTR: Long terminal repeat

The following scientific aims were addressed:

1. Can radiation-induced DSB be stably marked, tracked and identified at single-nucleotide resolution during expansion?
2. How do transcriptional and epigenetic states influence DSB site distribution?
3. Do frequently damaged and repaired genomic regions exist that are likely to influence radiation-induced genomic instability and radiotherapy?

The obtained results from this work should bring new insights into the mechanisms that initiate DSB-induced genomic instability, radiotherapy resistance and carcinogenesis.

2. MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals

Chemicals/Reagents	Company
100 bp / 1 kb Marker	Invitrogen
Agarose LE	Sigma
Ampicillin	Roth
Aqua ad iniectabilia (dH$_2$O)	Braun
BD™ Cytometer Setup & Tracking Beads Kit	BD Beckton Dickinson
BD FACS Clean Solution	BD Beckton Dickinson
BD FACS Flow Sheath Fluid	BD Beckton Dickinson
BD FACS Rinse Solution	BD Beckton Dickinson
Bovine Serum Albumine (BSA)	Sigma
Bromphenol blue	Sigma
Deoxyribonucleotic triphosphate (dNTP)	Fermentas (Thermo Fisher Scientific)
Dimethylformamid	Sigma
Dulbecco's Modified Eagle Medium (DMEM)	Invitrogen
Ethanol	VWR
Ethidiumbromid solution (0.07%)	Applichem
Ethylenediaminetetraacetic acid (EDTA)	Applichem
Fetal Calf Serum (FCS)	PAN
Glycerol	Sigma-Aldrich
Guanidin hydrochloride	Sigma
Hexanucleotid Mix (10x)	Roche
Human genomic DNA	Roche
Iscove's Modified Dulbecco's Medium (IMDM)	Invitrogen
Isopropanol	Sigma-Aldrich
LB Agar Miller	US Biological
Lithiumchlorid (LiCl)	Sigma
Loading Buffer (5x)	Elchrom Scientific
Luria-Bertani Broth (LB)	Invitrogen
Magnesiumchlorid (MgCl$_2$)	Sigma
ms2RNA	Roche
Sodium hydroxid (NaOH)	Fluka
Dynabeads M-280 Streptavidin	Dynal
Dynabeads MyOne Streptavidin T1	Dynal
PCR Grade Water	Roche
Penicillin/Streptomycin (Pen/Strep)	Invitrogen
Dulbecco's Phosphate Buffered Saline (DPBS; pH 7,4)	Gibco
Polybrene (1000x; 8 µg/ml)	Sigma
Polyethylenimin (PEI; 1mg/ml)	Sigma-Aldrich
Propidium iodide (1mg/ml)	Molecular Probes (Invitrogen)
Proteinase K	Roche/Qiagen
Puromycin	Invitrogen
RNase/DNase free H$_2$O	Ambion
Tris-borate-EDTA (TBE) Buffer (10x)	Amresco
Trizma-HCl (Tris-HCl)	Applichem

Trypanblue	Invitrogen
Trypsin/EDTA (0,05%)	Life Technologies
Tween 20	Sigma
Vectashield Mounting Medium with DAPI	Vector Laboratories
X-Gal	Sigma

2.1.2   Enzymes

Enzymes	Company
CircLigase ssDNA ligase	Epicentre
Klenow DNA Polymerase	Roche
Quick DNA ligase	Epicentre
T4 DNA Ligase	New England Biolabs
Restriction enzymes with respective buffers	New England Biolabs
SYBRGREEN I Mix	Roche
Taq DNA Polymerase	Qiagen/Genaxxon

2.1.3   Bacteria

Bacterial Strain	Genotype	Company
E.coli TOP10	F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*Z ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Life Technologies
E.coli Stbl3	F$^-$ *mcr*B, *mrr-hsd*S20 (r$_B^-$, m$_{B-}$), *rec*A13, *sup*E44 *ara*-14, *gal*K2, *lac*Y1, *pro*A2, *rps*L20(Str^R), *xyl*-5 λ$^-$*leumtl*-1	Life Technologies

2.1.4   Cell lines and primary cells

Name	Description	Company
A549	Human alveolar adenocarcinoma cell line [69]	ATCC
HEK 293T	Human embryonic kidney cell line, stably expressing the SV40 T-antigen [70]	ATCC
Hela	Human epithelial carcinoma cell line	ATCC
NHDF-A	Adult normal human dermal fibroblasts	PromoCell
PC3	Human prostate cancer cell line [71]	ATCC
U87	Human glioblastoma cell line [72]	ATCC

2.1.5 Antibodies

Name	Description	Company
Alexa Fluor® 647 anti-H2A.X-Phosphorylated (Ser139) Antibody (250ng/ml)	anti γH2AX-specific antibody labeled with Alexa Fluor 647	Biolegend
Alexa Fluor® 647 Mouse IgG1, κ Isotype Ctrl (ICFC)	Isotype control antibody, Antibody labeled with Alexa Fluor 647	Biolegend

2.1.6 Plasmids

Name	Description
pCCLsincPPT.PGK-IRES-eGFP.WPRE (LV106)	Lentiviral transfer vector expressing EGFP under control of human PGK promoter
LV001	Plasmid encoding *gag/pol* genes and carries D64V mutation in integrase gene
LV102	*env*-encoding plasmid
LV103	Plasmid encoding the VSV-G protein
#1211	Plasmid encoding CCR5-specific ZFN monomer 1
#1212	Plasmid encoding CCR5-specific ZFN monomer 2
pCR2.1 TOPO-TA	pUC-derived subcloning vector with multiple cloning site, ampicillin and kanamycin resistance cassette

2.1.7 Oligonucleotides

DNA oligonucleotides were partially biotinylated (B) at the 5' terminus. Additional modifications include dideoxycytosine (ddC) to inhibit ligation of DNA oligos to 3' terminus and phosphothiorate (*) in the phosphate backbone of the last three bases of some DNA oligos. Oligonucleotides were cleaned up by high performance liquid chromatography (HPLC) and lyophilized. Oligonucleotides were purchased from MWG or Sigma Aldrich. B, Biotin; D: degenerated base; LC, Linker cassette; Tit, Titanium 454 primer; $(N)_{2-6}$, Recognition sequence for pyrosequencing.

2.1.7.1 Standard primers for q-RT-PCR

Name	Sequence (5´-3´)
Myo461-439	CTCCCAGTGGCACAGCAGTTAGG
Myo122-143	TGTGCCCCAGGTTTCTCATTTG
GFP2 fwd	TGAGCAAGGGCGAGGAGCTGTT
GFP3 rev	GCCGGTGGTGCAGATGAACT

2.1.7.2   Primers used for linker cassettes in LAM-PCR

Name	Sequence (5´-3´)
LC1	GACCCGGGAGATCTGAATTCAGTGGCACAGCAGTTAGG
LC1 (CATG)	GACCCGGGAGATCTGAATTCAGTGGCACAGCAGTTAGGCATG
LC3 (CG)	CGCCTAACTGCTGTGCCACTGAATTCAGATC
LC2	CCTAACTGCTGTGCCACTGAATTCAGATC
LC3 (AATT)	AATTCCTAACTGCTGTGCCACTGAATTCAGATC
LC3 (TA)	TACCTAACTGCTGTGCCACTGAATTCAGATC

2.1.7.3   Primers used for LAM-PCR

Name	Sequence (5´-3´)
SK-LTR 1bio	B -GAGCTCTCTGGCTAACTAGG
SK-LTR 3bio	B -GAACCCACTGCTTAAGCCTCA
SK-LTR 4bio	B -AGCTTGCCTTGAGTGCTTCA
SK-LTR 5	AGTAGTGTGTGCCCGTCTGT
SK-LTR 5 1/2	GTGTGACTCTGGTAACTAGAG
LCI	GACCCGGGAGATCTGAATTC
LCII	GATCTGAATTCAGTGGCACAG

2.1.7.4   Fusionprimer for Pyrosequencing

Name	Sequence (5´-3´)
Tit3nrLV	CCATCTCATCCCTGCGTGTCTCCGACTCAG$(N)_{6-10}$ GATCCCTCAGACCCTTTTAGTC
Tit3SKLV	CCATCTCATCCCTGCGTGTCTCCGACTCAG$(N)_{6-10}$ TGTGTGACTCTGGTAACTAG
Tit5SKLV	CCATCTCATCCCTGCGTGTCTCCGACTCAG$(N)_{6-10}$ AAGCAGATCTTGTCTTCG
MegaL6/10T	GCCTCCCTCGCGCCATCAG$(N)_{6-10}$GATCCCTCAGACCCTTTTAGTC
MegaL_U3_10/T	GCCTCCCTCGCGCCATCAG$(N)_{6-10}$AAGCAGATCTTGTCTTCG
MiS3nrLV	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT $(N)_{6-10}$GATCCCTCAGACCCTTTTAGTC
Mis3SKLV	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT $(N)_{6-10}$TGTGTGACTCTGGTAACTAG
Tit Linker	B-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGAGTGGCACAGCAGTTAGG
Megalinker	B-GCCTTGCCAGCCCGCTCAGAGTGGCACAGCAGTTAGG
MiS_LK	GACCCGGGAGATCTGAATTCAGTGGCACAGCAGTTAGG$(N)_{16}$CTA

2.1.7.5   Primers and oligos used for direct DSB labeling approaches

Name	Sequence (5´-3´)
L.CDSB 1	B-CCTTGGGAGGGTCTCCTCTGAGTGATTGACAAAAAD
L.CDSB 2	B-CTTGGGAGGGTCTCCTCTGAGTGATTGACAAAAA
L.CDSB 2-1	CTTGGGAGGGTCTCCTCTGAGTGATTGACAAAAA
E.CDSB 3	CTTGGGAGGGTCTCCTCTG
dsGRUV1	P-GTAGTCAATACATCAGAGGAGACCAGTCGGATGCAACTGCAAGATATTG GATACACGGGTACC*C*G*G
dsGRUV2	P-GTAGTCAATACATCAGAGGAGACCAGTCGGATGCAACTGCAAGATATTG GATACACGGGTACCCGGGACCCG*G*G*A
dsGRUVrev	P-CCGGGTACCCGTGTATCCAATATCTTGCAGTTGCATCCGACTGGTCTCCT CTGATGTATTGA*C*T*A-ddC
ssGRUV5-A-Phthio	P-GTAGTCAATACATCAGAGGAGACCAGTCGGATGCAACTGCAAGATATTG GATACACGGGTACC*C*G*G-ddC
GRUV5-2expo	CATCCGACTGGTCTCCTCTGATG

2.1.8   Commercial kits

Kits	Company
Ampure XP beads	Beckman Coulter
Fast-Link Ligation Kit	Epicentre
Fluorescence Biotin Quantitation Kit	ThermoScientific
High Pure PCR Template Preparation Kit	Roche
HiSpeed Plasmid Purification Kit	Qiagen
Microcon-30 Kit	Millipore
QIAamp DNA Mini Kit	Qiagen
QIAfilter Plasmid Purification Kit	Qiagen
QIAprep Spin Miniprep Kit	Qiagen
QIAprep Spin Maxiprep Kit	Qiagen
QIAquick PCR Purification Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
TOPO-TA Cloning Kit	Invitrogen

2.1.9   Buffers, Media, Solutions

Buffers, Media, Solutions	Chemical	Final Concentration
3M LiCl Solution**	Tris-HCl (pH 7.5)	10 mM
	EDTA	1 mM
	LiCl	3 M
6M LiCl Solution**	Tris-HCl (pH 7.5)	10 mM
	EDTA	1 mM
	LiCl	6 M
Acid-isopropanol solution	HCl (1N)	0.04N
	Isopropanol	
Ampicillin**	Ampicillin sodium salt	100mg/ml
	Desalted water	
Cell culture freezing medium	DMSO	15%
	FCS	30%
	Cell culture medium	55%
Denaturation solution	NaOH	0.1 M
DMEM Cell culture medium	DMEM	89%
	FCS	10%
	Pen/Strep	1%
DNA loading buffer (5x)	Tris-HCl (pH 7)	25 mM
	EDTA (pH 8)	150 mM
	Bromphenol blue	0.05%
	Glycerol	25%
DPBS-Tween 20 (DPBST)	Tween 20	0.2%
	DPBS (pH 7.4)	
IMDM cell culture medium	IMDM	
	FCS	10%
	Pen/Strep	1%
LB-Agar*	LB-Miller Agar	3.7%
	Desalted water	
LB-medium*	LB	2%
	Desalted water	
MTT stock solution**	MTT	5mg/ml
	1xDPBS	
Paraformaldehyde solution**	Paraformaldehyde	4%
	DPBS (pH 7.4)	
Triton solution	Triton X-100	0.5%
	DPBS (pH 7.4)	
Washing solution for paramagnetic beads**	DPBS (pH 7.4)	
	BSA	0.10%
X-Gal	X-Gal	20 mg/ml
	Dimethylformamid	

* Autoclaved and addition of antibiotics (Ampicillin, Kanamycin) after cooling to 60°C.

** Solutions were sterilized by filtering through 0.22µm filter

2.1.10 Disposables

Disposable	Company
Sheeting	VWR
Sheeting for LightCycler LC480	Roche
Polypropylene Round Bottom Tube (15 ml)	BD Beckton Dickinson
Polystyrene Round Bottom Tube (5 ml)	BD Beckton Dickinson
Lid for 96-well plates	Greiner bio-one
Falcon, BD™ Falcon™ Tubes (15 ml, 50 ml)	BD Beckton Dickinson
Filter (0,22 µm, 500 ml)	Millipore
Filter System (0,22 µm) Stericup®	Millipore
Gloves (Nitril)	Microflex
Ionoculation Loop	Sarstedt
Parafilm	Brand
PCR-Tubes Softtubes (0,2 ml)	Kisker Biotech
Pipette tips	Starlab
Pipette tips for LightCycler LC480	Starlab
Mixing plates (96-well)	Greiner bio-one
Plates for LAM-PCR (96-well)	Greiner bio-one
Plates for LightCycler LC480 (96-well)	Roche
Photobase paper VM 65 H	Mitsubishi
Reaction tubes (0,5 ml, 1,5 ml, 2 ml)	Eppendorf
RNase/DNase-free reaction tubes (0,5 ml, 1,5 ml, 2 ml)	Ambion
Ultracentrifugation tube	Beckman Coulter
Cell culture flasks (25 cm^2, 75 cm^2, 225 cm^2)	Nunc Brand Products
Cell culture pipettes (2-50 ml)	BD Beckton Dickinson
Cell culture plates (6-, 12-, 24-, 96-well)	BD Beckton Dickinson
Cell culture dish (10 cm, 15 cm)	Greiner bio-one
Cell scraper	BD Beckton Dickinson
Cryotube	BD Beckton Dickinson

2.1.11 Equipment

Equipment	Company
Autoclave	Systec
BD™ LSRII Flow cytometer	BD Beckton Dickinson
Cell culture hood	Heraeus
Centrifuges	Eppendorf
Lysine-coated coverslips	BD Beckton Dickinson
Electrophoresis Power Supply	Pharmacia/Elchrom Scientific
Fluorescence microscope	Carl Zeiss Jena
Freezer -20°C	Liebherr
Freezer -80°C	Sanyo
Fridge	Liebherr
Gel documentation system	Peqlab
Gel electrophoresis chamber	Biometra

Heating Block	Eppendorf
Horizontal shaker (KS 250B)	IKA Labortechnik
Incubator (37°C)	Binder Lifesciences
Ice machine	Ziegra
Infinite 200" plate reader	Tecan
LightCycler LC480	Roche
Liquid Nitrogen Tank	Thaylor-Wharton
Magnetic Particle Collector MPC 96	Dynal
Microplate reader	Biotek
Microscope	Carl Zeiss Jena
Microwave	Siemens
Multi-channel pipette	Eppendorf
NanoDrop spectrophotometer	Thermo Scientific
Neubauer counting chamber	Optik Labor
Optima^{TM} L-90K Ultrazentrifuge	Beckman Coulter
PCR cycler	Biometra
Picofuge	NeoLab
Pipettes (Pipetman P2, P10, P200, P1000)	Eppendorf
Pipette device	Integra Bioscience Pipetboy acu
Proton and Carbon Radiation Source	Heidelberg Ion Therapy Center
Scales	Sartorius
Special accuracy weighing machine	Sartorius
Vacuum pump	NeoLab
Videoprinter	Mitsubishi
Vortexer (MS1)	IKA Labortechnik
Water bath	Thermo Electron Coporation
XRAD320 X-ray device	Precision X-Ray

2.1.12 Software and data bases

Software and data bases	Company
Adobe Acrobat X Pro	Adobe
BLAST Search	http://www.ncbi.nlm.nih.gov/blast
BLAT Search	http://genome.uscs.edu/
FACSDiva Software V6.1.2	BD Beckton Dickinson
Fiji image processing package	http://www.fji.sc/
Galaxy project	in house
USCS Genome Browser	http://genome.uscs.edu/
High-throughput Insertion Site Analysis Pipeline (HISAP)	in house
Ingenuity Pathway Analysis (IPA)	http://www.ingenuity.com/
Lasergene	DNA Star
LightCycler LC480 Software	Roche
Office 2007 (Word, Excel, PowerPoint)	Microsoft
Photoshop CS5.1	Adobe
R-Program (2.13.1)	cran.r-project.org

2.2   Methods

2.2.1   Cell Culture Methods

2.2.1.1   Cell Cultivation

A549 (human alveolar basal epithelial cancer cell line), PC3 (human prostate cancer cell line) and U87 (human glioblastoma cell line) were maintained with twice-weekly subculture in Dulbecco`s Minimal Essential Medium (DMEM) supplemented with 10% fetal calf serum (FCS) and 1% Penicillin/Streptomycin (Pen/Strep). 293THEK (human embryonic kidney cell line) were maintained in IMDM supplemented with 10% FCS and 1% Pen/Strep. Adult normal human dermal fibroblasts (NHDF-A) were cultivated in DMEM supplemented with 10% FCS, 1% Pen/Strep and 1% L-Glutamine. NHDF-A cultivated for more than nine passages were not used for radiation and chemotherapy experiments. All cells were cultured at 37°C and 5% CO_2. For passaging, the adherent cells were washed once with pre-warmed 1xDPBS and subsequently detached using 3ml or 5ml 0.05% trypsin-EDTA for a 10cm or 15cm cell culture dish, respectively. The cells were incubated at 37°C in an incubator for 7 minutes. Trypsin activity was blocked by addition of 7ml (10cm cell culture dish) or 10ml (15cm cell culture dish) cell culture medium supplemented with 10% FCS.

2.2.1.2   Freezing and thawing of cells

In order to freeze cells, cell culture freezing medium containing 3ml DMSO, 6ml FCS and 11ml cell culture medium, was prepared. Cells were detached from the cell culture dish, centrifuged at 1,000rpm and room temperature for 5min, and resuspended in cell culture medium. Subsequently, 400ul aliquots were transferred into cryotubes. To each tube, 400µl freezing medium was added, and the tubes placed in freezing boxes. The freezing boxes were stored at -80°C over night. Finally, cryotubes were placed in a liquid nitrogen tank. For thawing, cryotubes containing cells were incubated in a water bath for two minutes until they were thawn. Then, cells were resuspended in 10ml cell culture medium in 15ml reaction tubes, centrifuged at 1000rpm for 5min and seeded in cell culture dishes of appropriate size.

2.2.1.3   Cell Counting

For cell counting, 10µl medium containing cells detached from the cell culture dish were transferred to a 1.5ml reaction tube and diluted 1:10 in 90µl trypan blue. From this mixture, 10µl were loaded into a Neubauer counting chamber and the average number of cells in the chamber calculated accordingly:

$$Cell\ density\ (cells/ml) = \frac{Average\ Cell\ No.}{Dilution\ Factor\ (10^{-1})}\ x\ 10^4$$

2.2.1.4   Transfection

The term transfection describes several processes for the introduction of nucleic acids into mammalian cells. In this work, transfection was carried out by mixing the cationic lipid polymer poly-ethylene inositol (PEI) with the DNA to produce liposomes, which fuse with the cell membrane to release the DNA into the cell.

Transfection of lentiviral plasmids for virus production was performed on 293T human embryonic kidney (HEK) cells (chapter 2.2.1.6). Therefore, $1x10^6$ 293T HEK cells were seeded in a 15cm cell culture dish with 15ml IMDM 24h prior to transfection. In total, 20 dishes were used for virus production. Following incubation, the cell culture medium was removed and replaced with fresh 10ml IMDM supplemented with 15% FCS.

	Reagent	Volume for one 15cm dish
PEI solution	PEI [1mg/ml]	167.25µl
	IMDM without FCS	ad 2.5ml
DNA solution	LV001 [1mg/ml]	12.5µg
	LV102 [1mg/ml]	6.25µg
	LV103 [1mg/ml]	9µg
	LV106 [1mg/ml]	28µg
	IMDM without FCS	ad 2.5ml

For transfection of CCR5-specific ZFN (kindly provided by Dr. Richard Gabriel), $1x10^6$ A549 or 293T HEK cells were seeded in a 15cm cell culture dish with 15ml IMDM 24h prior to transfection. Following incubation, the cell culture medium was removed and replaced with fresh 10ml IMDM supplemented with 15% FCS.

	Reagent	Volume for 15cm dish
PEI solution	PEI [1mg/ml]	18µl
	IMDM without FCS	ad 2.5ml
DNA solution	#1211 [1mg/ml]	3µl
	#1212 [1mg/ml]	3µl
	IMDM without FCS	ad 2.5ml

The Zinkfinger nuclease consists of two monomers (#1211, #1212) which dimerize inside the nucleus at the target DNA sequence to induce a DSB. Each monomer has a DNA binding domain which can be targeted basically against every locus in the genome and a FokI-nuclease domain that introduces the DSB upon dimerization.

The transfection solutions were mixed together and incubated at room temperature for 30 minutes prior to addition to the cell culture dish. Then, the cells were incubated with the transfection mixture over night. The next morning, the medium was discarded and replaced with fresh 12.5ml IMDM (10% FCS).

2.2.1.5   Transduction

The introduction of DNA into target cells by viral delivery systems is termed transduction. Here, lentiviruses carrying the IDLV vector were used to transduce target cells. During transduction, the lentiviruses bind to and fuse with the cell membrane. The viral capsid enters the cytoplasm, where the nucleic acid is released. The lentiviral RNA is reverse transcribed into double-stranded DNA and transported into the nucleus.

For transduction, $1x10^6$ cells in 15ml medium were plated in 15cm cell culture dishes 24h prior to transduction. Then, the cell culture medium was removed and replaced with fresh 10ml medium containing 1x polybrene. Polybrene increases the transduction efficiency by neutralizing the electrostatic charges between the virion and the negative cell surface. The virus was pooled in a single reaction tube, added to the cells and replaced with fresh medium after 24h incubation.

2.2.1.6   Virus production

For virus production, 293T HEK cells were co-transfected (chapter 2.2.1.4) with the four plasmids encoding the Gag, Pol (LV001), Rev (LV102) and VSV-G (LV103) proteins as well as the transfer vector carrying the *eGFP* gene (LV106). Following 48h incubation, the virions were isolated from the cell culture medium. The medium was removed from the cell culture dish and filtered through a 0.22µm filter. The flowthrough containing the virions was then split into six vials and concentrated by ultracentifugation at 20°C and 20,000 x g for 2h. The pellet in each vial containing the virions was resuspended in 70µl 1xDPBS, incubated for 20 minutes at RT, and the virions pooled in a single 1.5ml reaction tube. The tube was placed in a rotator for 20 minutes and finally stored in 10µl aliquots at -80°C.

2.2.1.7   Determining the lentiviral titer on Hela cells

In order to determine the efficiency of the virus production and to calculate the multiplicity of infection (MOI) for transduction of the target cells, the titer of the lentivirus production was determined. The MOI is the number of infectious particles that can infect a specific number of cells: MOI1 means that for a defined number of cells the same number of virus particles is present in the solution.

Since the IDLV vector used in this study encodes EGFP under control of the human PGK promoter, the lentiviral titer was determined by FACS analysis. First, Hela cells were seeded in all wells of a 6-Well plate at a density of 5 x 10^4 cells per well and incubated for 24h at 37°C, allowing each cell to divide once. The medium was then removed and replaced with 500µl DMEM containing 2µl 1000x polybrene. A virus dilution series in 1ml DMEM was prepared and 500µl transferred to each well containing cells. The cells were incubated with IDLV for 48h before FACS analysis. To calculate the number of infectious particles (Titration units, TU) per ml the following formula was used:

$$Lentiviral\ Titer\ [^{TU}/_{ml}] = \frac{^{GFP^+Cells\ [\%]}/_{100}}{Dilution\ Factor} \times Number\ of\ Cells\ (10^5)$$

2.2.1.8   MTT Assay

The MTT assay is a quantitative colorimetric assay for mammalian cell survival and proliferation [73]. The tetrazolium salt MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide) is a yellow substance that is metabolized into purple, insoluble salt crystals by dehydrogenase enzymes in mitochondria of living cells exclusively. Thus, the degree of crystal formation is proportional to the number of living cells.

The MTT reaction solution (10µl MTT in 100µl medium) was added to all assay wells containing cells in a 96-well plate, and plates were placed in the incubator at 37°C for 4h. Subsequently, the MTT-reaction solution was removed, replaced with 100µl acid-isopropanol solution (0.04N HCl in isopropanol) and mixed thoroughly to dissolve all crystals. The solution was incubated at RT for 5min, and finally the absorption measured in a microplate reader at wavelength of 610nm. Triplicate measurements for each substrate concentration were performed. Here, the MTT assay was used to determine the LD50 values of etoposide, doxorubicin, Nu7441 and Mirin on A549 and NHDF-A.

2.2.1.9   Immunostaining of γH2AX foci

Immunostaining of cellular proteins involved in the DNA damage repair with fluorescent-labeled antibodies is the most frequently used method to study DSB induction and repair. Target proteins used in immunostaining include 53BP1 (53 binding protein 1), ATM (Ataxia telangiectasia mutated) and γH2AX (phosphorylated histone H2AX). Especially immunostaining of the phosphorylated histone variant γH2AX is often used as a sensitive marker for DSB induction and repair. H2AX becomes phosphorylated at its Serine 139 (Ser139) residue after DSB induction to form γH2AX. Responsible for this modification are kinases of the phosphatidylinositol-3 kinase (PI3K) family, namely ATM, ATR and DNA-PKcs. The phosphorylated H2AX localizes to the DSB site and forms microscopically-visible foci in a range of 1-2 mega base pairs around the DSB site. Upon DSB repair, the phosphate residue is removed from γH2AX and the foci dissociate. Hence, analyzing the dynamics in γH2AX phosphorylation and dephosphorylation is often used to study the kinetics of DSB induction and repair.

For γH2AX immunostaining, cells were seeded on coverslips suitable for cell cultivation in 6-Well plates for 24h prior to DSB induction by irradiation or chemotherapy. Subsequently, cells were washed twice with 500µl 1xDPBS per well. Then, the cells were fixed with 500µl 4% paraformaldehyde for 10min at RT, rinsed twice with 500µl 1xDPBS for 5min each and subsequently permeabilized with 0.5% Triton X-100 in 1xDPBS at RT for 7min. To remove excess Triton X-100, the cells were rinsed twice with 500µl 1xDPBS for 5min each. Next, unspecific binding sites were blocked with 500µl 1% BSA in DPBST for 2h at RT, and γH2AX foci were stained with 200µl γH2AX-specific antibody labeled with Alexa Fluor 647 (100ng/ml) in 1% BSA-DPBST solution at 4ºC overnight on a horizontal shaker. As a negative control for immunostaining, treated cells were stained with IgG-Alexa Fluor 647 antibody. After incubation, cells were washed twice with 500µl DPBST for 5min each. Finally, one drop of vectashield containing DAPI was added to each coverslip and the slips were viewed upside down under confocal microscope. For counting of γH2AX, the Fiji platform was used [74].

2.2.1.10   Inhibition of NHEJ-repair activity by Nu7441 and Mirin

Blocking DNA repair activity with chemical compounds increases radio- and chemosensitivity and can be used to study DNA repair mechanisms [75]. Here, NHEJ-repair activity was blocked by Nu7441 (kindly provided by Dr. Friederike Herbst) and mirin. Nu7441 inhibits DNA-PK activity and thus blocks classical NHEJ-repair pathway. The MRN complex inhibitor mirin blocks the Mre11-mediated activation of ATM following DNA damage, thus inhibiting end-resection, a key step in alternative NHEJ [75]. Either Nu7441 or mirin or both together were added to IDLV-transduced NHDF-A at LD10 3h prior to irradiation. The cell culture medium containing the NHEJ inhibitors was removed from the cells 6h after irradiation. Cells were pooled into triplicates (3×10^5 cells each)

and expanded until constant EGFP expression was observed. At passages 4, 6, and 8 (end point) one third of the cell populations was subjected to DNA isolation in order to analyze the clonal dynamics during expansion.

2.2.1.11   Irradiation of cells

Irradiation was performed in collaboration with the Max-Eder-Junior Research Group Translational Radiation Oncology at the National Center for Tumor Diseases (NCT) and German Cancer Research Center (DKFZ). Prior to irradiation, 6×10^6 cells were transduced with IDLV at MOI3 in 15cm cell culture dishes for 48h. Cells were subsequently transferred into 96-Well plates at a density of 1×10^5 cells/well (total of 1×10^6 cells per radiation dose) 12h before irradiation. In order to induce DSB and DNA repair activity, cancer cell lines were irradiated with various doses of photon (X-rays) or hadron (protons, carbons) radiation beams ranging from 0.125Gy to 10Gy. Primary human dermal fibroblasts (NHDF-A) were irradiated with X-rays only. Photon irradiation was delivered by a XRAD320 X-ray device (Precision X-Ray, CT) at 320 KeV, while proton and carbon beam therapy was performed at the Heidelberg Ion Therapy Center (HIT) with the horizontal beamline using a raster scanning technique [76]. For the detection of induced and repaired DSB sites within 24h after irradiation, 1×10^6 cells were detached at 1h, 4h, 8h, and 24h after irradiation and subjected to DNA isolation, LAM-PCR and deep sequencing. In order to analyze the DSB distribution in expandable cell populations, 3×10^5 cells were pooled into triplicates in 6-Well plates immediately after irradiation and expanded in cell culture until constant EGFP expression levels from stably integrated IDLV were observed. The cells were then subjected to DNA isolation and LAM-PCR.

2.2.1.12   Inhibition of Topoisomerase 2 in mammalian cells with doxorubicin and etoposide

Doxorubicin and etoposide belong to the drug class of topoisomerase 2 poisons. Both chemicals inhibit the active TOP2 dimer from religation of the G segment. The stalled TOP2 is therefore removed by proteolytic enzymes and nucleases from the DNA, thereby exposing the DNA double strand break to the repair machinery.

Prior to TOP2 poisoning, 3×10^6 NHDF-A and Hela cells were transduced with IDLV (MOI3) in 15cm cell culture dishes for 48h and subsequently seeded in 96-Well plates at a density of 1×10^5 cells per well. Then, NHDF-A and Hela cells were treated with doxorubicin or etoposide for seven subsequent days. The DMEM cell culture medium containing the TOP2 poisons was replaced with fresh poison containing cell culture medium on days 3 and 5. Following treatment, 3×10^5 cells were pooled into triplicates in 6-Well plates and expanded in cell culture until constant EGFP expression levels from stably-integrated IDLV were observed. NHDF-A cells were subjected to DNA isolation and LAM-PCR.

2.2.1.13   Preparation for FACS

Fluorescent Activated Cell Sorting (FACS) is a special method of flow cytometry. It allows cells to be sorted based upon specific physical properties such as light scattering or fluorescent characteristics of the cell. A single-cell suspension passes a laser with a distinct wave length and the emission is recorded. The cells can be either stained by fluorescent-tagged antibodies binding to cell-type specific target molecules or the cells express a fluorescent protein themselves. In addition to fluorescence intensity, the cell size and granularity can be

determined. In this thesis, enhanced green fluorescent protein (EGFP)-expression from lentiviral vectors was measured in order to determine transduction efficiency and follow integration of IDLV during cultivation. Additionally, flow cytometry was used to determine cell cycle phase in synchronized NHDF-A. EGFP is excited at a wave length of 488nm and has a single major emission peak at 509nm.

For flow cytometry measurement of EGFP, the cells were washed with 1xDPBS, detached from the cell culture dish with 0.05% trypsin at 37°C for 7min and transferred to a fresh polypropylene tube. Cells were centrifuged at 800 rpm for 5min at RT. Subsequently, the cells were washed once with 1xDPBS, centrifuged, and dead cells stained with propidium iodide in 2% BSA-DPBS solution. Then, cells were centrifuged at 800 rpm for 5min at RT, and resuspended in 500µl 2% BSA-DPBS. Cells were stored on ice until FACS analysis.

For determining the cell cycle phase of NHDF-A, the cells were washed once with 1xDPBS, detached from the cell culture dish with 0.05% trypsin at 37°C for 7min and transferred to a fresh propylene tube. Cells were centrifuged at 1500xg for 5min at RT. Subsequently, the cells were washed once with 1xDPBS, centrifuged, and fixed by adding 1ml ice-cold 70% ethanol dropwise 0.3ml cell suspension in DPBS. Cells were left on ice for 2h. Subsequently, cells were centrifuged, the supernatant discarded and cells washed twice with 1ml DPBS. Since PI does not exclusively bind to nuclear DNA, but also to double-stranded RNA, the RNA needed to be degraded. Therefore, NHDF-A were incubated with 75µl ribonuclease A (Rnase A) (10mg/ml) at 37°C for 30min. The cells were subsequently mixed with 1ml DPBS, centrifuged and stained with propidium iodide in 2% FCS-DPBS solution on ice for 30min. Then, cells were centrifuged at 1200xg at RT for 5min, and resuspended in 500µl 2% FCS-DPBS. Cells were stored on ice until flow cytomtric analysis. Flow cytometry was performed on a LSRII from BD Biosciences.

2.2.1.14   Synchronization of NHDF-A in G1/G0, S and G2/M phase of the cell cycle

Cell synchronization is a process by which cells in different cell cycle phases are brought to the same phase. Several different strategies including addition of chemical substances or depletion of factors in the cell culture medium are used to achieve synchronization. Here, NHDF-A were synchronized in G1/G0, S and G2/M phase of the cell cycle by serum starvation, double thymidine block and nocodazole treatment, respectively.

In order to synchronize cells, the cells need to be seeded at a confluency of 30-50% to enable cell division and prevent contact inhibition. Elimination of serum from the culture medium results in the accumulation of cells in G1/G0 phase. Therefore, DMEM supplemented with 10% FCS was removed from the cells and replaced with DMEM containing 1% FCS for 72h. In order to synchronize cells in S phase, DNA synthesis can be blocked by using the inhibitor thymidine. The cell culture medium was removed from the NHDF-A and replaced with DMEM supplemented with 10% FCS and 2mM thymidine for 48h. This first block was followed by a 16h incubation period with DMEM supplemented with 10% FCS. Then, a second thymidine block was induced by replacing the medium with DMEM supplemented with 10% FCS and 2mM thymidine for 72h. Nocodazole interferes with the polymerization of microtubules at the end of G2 phase and beginning of mitosis, thereby inhibiting entry of cells into mitosis. Here, NHDF-A were synchronized in G2/M by addition of nocodazole at a concentration of 100ng/ml to DMEM supplemented with 10% FCS. Efficiency of synchronization was analyzed by flow cytometry.

2.2.2   Molecular Biology Methods

2.2.2.1   Isolation of genomic DNA from cultivated cells

For the isolation of genomic DNA, the High Pure PCR Template Preparation Kit from Roche was used according to the manufacture's protocol with a single modification. The DNA was eluted twice with 50µl dH$_2$O each into a single 1.5ml reaction tube. The DNA concentration was measured at the Nanodrop-1000 and stored at -20°C.

2.2.2.2   Determining the DNA concentration

The concentration of the genomic DNA was measured using either the Nanodrop-1000 or Qubit. For Nanodrop DNA analysis, 1.5µl DNA were transferred to the Nanodrop-1000 machine. The DNA concentration was determined spectrometrically at a wavelength of 260nm. For Qubit measurements, the Qubit dsDNA HS assay kit was used according to manufacturer's protocol.

2.2.2.3   Polymerase Chain Reaction (PCR)

The Polymerase Chain Reaction (PCR) enables the exponential amplification of a target DNA for cloning, induction of mutations or sequencing [77]. The PCR can be subdivided into three basic steps, which are repeated for 20-50 cycles. Each cycle starts with the denaturation of the double-stranded DNA by heating the DNA to 95°C. Prior to the first PCR cycle, the DNA is usually heated for a longer time to guarantee that the two DNA strands are separated completely before the cycling begins. Subsequently, the DNA is cooled to allow primers to anneal to the target sites. The exact temperature for primer annealing is calculated based on the primer length and its base composition. Usually, annealing temperatures are about 60°C. If the annealing temperature is too low, primers can bind to unrelated DNA sites and result in the amplification of unspecific sequences. In the final step, called elongation phase, the temperature is increased up to 68°C-72°C and the DNA is amplified by the DNA polymerase starting from the 3' end of the primers. During elongation, the primer becomes integrated into the amplified DNA (amplicon). The repetition of the three phases (denaturation, annealing, elongation) results in the exponential amplification of the target DNA sequence.

Reagent	Amount
H$_2$O bidest	Ad 50µl
10x Taq DNA Polymerase Buffer	5µl
dNTP mix (dATP, dGTP, dCTP, dTTP) (10mM each)	1µl
Sense Primer (10pmol/µl)	20pmol
Antisense Primer (10pmol/µl)	20pmol
Taq DNA Polymerase (1U/µl)	2µl

For amplification, the following PCR cycling program was used for a *Thermus aquaticus* (Taq) DNA polymerase:

Description	Temperature	Time	Repetitions
Initial Denaturation	95°C	5min	
Denaturation	95°C	30sec	
Annealing	60°C	30sec	30 cycles
Elongation	72°C	1min	
Final Elongation	72°C	5min	
Cooling	4°C		

2.2.2.4   DNA agarose electrophoresis

Gelectrophoresis is a frequently used method for the separation and visualization of DNA. The DNA fragments are separated according to their molecular mass in an agarose gel when an electric charge is applied. Since the DNA is negatively charged, it is pulled towards the anode. According to the size of the DNA fragments to be separated, 0.8-2% agarose concentrations are used.

First, the agarose was dissolved in 100ml 1xTris-borat-EDTA (TBE) buffer and subsequently heated in a microwave until the agarose powder was dissolved completely. Then, one drop of ethidiumbromide solution was added to the liquid agarose gel, and the gel was poured into a gel chamber. Upon solidification, the gel chamber was placed into an electrophoresis chamber with 1xTBE buffer. Then, 8µl DNA were mixed with 2µl 5x loading buffer and transferred into the wells of the gel. Finally, the DNA fragments were separated at 130V for 45 minutes. Since the ethidiumbromide present in the gel binds to the DNA and fluorescates in UV light, the DNA fragments was visualized upon UV light exposure. The DNA fragment sizes were determined by comparing the DNA bands to a marker of known size.

2.2.2.5   DNA isolation from agarose gel

For isolating DNA after gel electrophoresis from agarose gels, a piece of the gel containing the DNA fragment of interest was excised and transferred into a 1.5ml reaction tube. Subsequently, the DNA was isolated by using the QIAqucik Gel Extraction Kit from Qiagen according to manufacturer's protocol.

2.2.2.6   Absolute quantitative real-time PCR (q-RT-PCR)

Real-time quantitative PCR is special type of PCR, which enables amplification and quantification of DNA simultaneously. Quantification is performed by measuring the incorporation of a fluorescent dye into newly synthesized double stranded DNA during amplification. Since the amount of DNA increases during each PCR cycles and thus the fluorescence, the fluorescence intensity can used to calculate the amount of DNA in each reaction. Here, the fluorescent dye used was SYBR green I, which is measured after each PCR cycle. SYBR Green is excited 488nm and it emits green light at a wavelength of 522nm. By comparing the fluorescent levels of the DNA samples to a standard dilution series of known concentration the DNA concentration in the sample can be calculated. To ensure that different samples can be compared to each other, the expression of the target gene is normalized to the expression of a reference gene present in every sample. Here, the target gene was the *EGFP* locus in IDLV, and the genomic myoglobin (*MYO)* gene locus was used as a reference.

Prior to each absolute q-RT-PCR, qPCR standards for the target and reference locus were synthesized. Therefore, the *EGFP* locus and the *MYO* locus were amplified by PCR using an annealing temperature of 60°C. Subsequently, the PCR fragments were analyzed by gel electrophoresis and isolated from the agarose gel. Then, the DNA concentrations were determined at the Nanodrop-1000, and the amplicon copy number (n) calculated using the following formula:

$$n = \frac{DNA\ for\ qPCR\ (ng) \times\ 6.0223x10^{23}\ (\frac{molecules}{mol})}{Length\ of\ PCR\ product\ \times 10^{9}\left(\frac{ng}{g}\right) \times 650\ (\frac{g}{mol})}$$

For the EGFP (326bp) and MYO (339bp) standards, $1x10^{8}$ DNA copies were inserted into the q-RT-PCR reaction, which is equal to 0.0352ng and 0.0367ng, respectively. Subsequently, the EGFP and MYO standard dilutions series ranging from 10^{6} to 10^{0} copies were prepared in 5µl ms2RNA solution per reaction. For each DNA sample obtained from IDLV-transduced and irradiated cells, 30ng genomic DNA diluted in 5µl ms2RNA solution were inserted for the amplification of the *MYO* genomic locus and the *EGFP* gene encoded in the IDLV. Subsequently, the following mastermix was prepared for each primer pair:

Reagent	Amount
H$_2$O bidest	4µl
Primer 1 (10mM)	0.5µl
Primer 2 (10mM)	0.5µl
2x SYBR Green I	10µl

For each standard dilution and DNA sample, triplicate reactions were prepared. The q-RT-PCR reactions were prepared in 96-Well plates specifically designed for q-RT-PCR. After the addition of the mastermix to each well, the 96-Well plate was sealed and centrifuged for 3min at 800xg. The qPCR was performed in a Lightcycler 480 using the following program:

Description	Target Temperature	Time	Ramp Rate	Repetitions
Pre-Incubation	95°C	15min	4.4°C/sec	
	95°C	10sec	2.2°C/sec	
Amplification	58°C	5sec	4.4°C/sec	40 cycles
	72°C	25sec	4.4°C/sec	
	95°C	5sec	4.4°C/sec	
Melting Curve	65°C	1min	2.2°C/sec	
	97°C		0.11°C/sec	
Cooling	40°C	30sec	2.2°C/sec	

2.2.2.7   Linear-Amplification Mediated Polymerase Chain Reaction (LAM-PCR)

To detect and analyze the integration site distribution of retroviruses in the genome, linear-amplification mediated polymerase chain reaction (LAM-PCR) was developed [78, 79]. In the first step of LAM-PCR, the retroviral vector-genome junction is amplified by a linear PCR step. Biotinylated primers bind to the LTR sequence of the IDLV and are subsequently elongated by the Taq polymerase. The resulting single-stranded amplicons are immobilized on paramagnetic beads which carry streptavidin on their surface, resulting in the

separation of the amplified sequences from the remaining genomic DNA. The following steps are performed with the solid phase. The single-stranded DNA is converted into double-stranded DNA using the bacterial Klenow Polymerase. Afterwards, the immobilized dsDNA is digested by a restriction enzyme that cuts in the genomic part of the amplicon to yield a single-stranded DNA overhang. This overhang is used to ligate a DNA linker of known sequence called the linker cassette to the amplicon. The non-biotinylated DNA strand is subsequently separated by denaturation and amplified by two rounds of nested PCR with vector- and linker-specific DNA primers. The amplified DNA is then subjected to pyrosequencing (Figure 9).
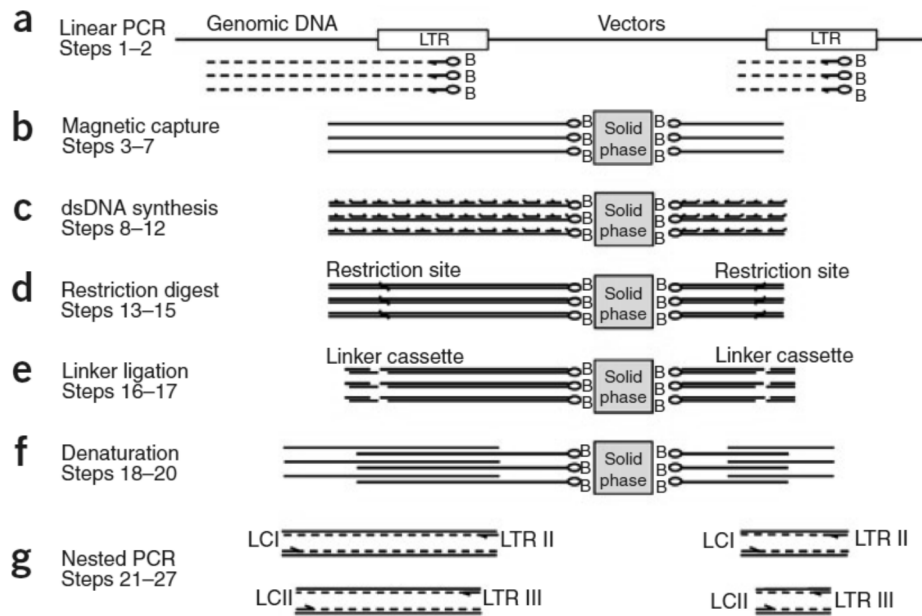
Figure 9: Schematic of 5' LAM-PCR for the amplification of retroviral integration sites. (a) Linear amplification of the vector-genome junction using biotinylated primers which bind in the U3 region of the LTR. (b) The amplicons from the linear amplification are immobilized to paramagnetic beads. (c) The single-stranded amplicon is converted into double-stranded DNA using Klenow polymerase from E.coli. (d) Enzymatic restriction digest of the genomic part of the amplicons to create single-stranded overhangs. (e) Ligation of DNA linker sequences to the ssDNA overhangs. (f) Alkaline denaturation detaches the non-biotinylated DNA, which is subsequently used for two rounds of exponential amplification (g). Since LTR are present at both ends of the retroviral genome, two fragments are amplified: vector-genome junction and an internal vector sequence. B: biotinylated primer; LTR: long terminal repeat; LTR I: oligonucleotide for the linear amplification; LTR II: oligonucleotide for the first exponential PCR; LTR II: oligonucleotide for the second exponential PCR; LCI: DNA linker-specific oligonucleotide for first exponential PCR; LCII: DNA linker-specific oligonucleotide for first exponential PCR. Image taken from [78]

Creation of linker cassettes

Two single-stranded DNA oligos that are complementary to each other can be hybridized to form a double-stranded DNA molecule. This dsDNA linker cassette is used in the ligation reaction. According to the restriction enzyme used during LAM-PCR, the sequence of second linker cassette is chosen and used for hybridization.

Basically, the two oligonucleotides are mixed in a hybridization buffer:

Reagent	Amount
LC 1 (100mM)	40µl
LC 3 (100mM)	40µl
MgCl$_2$ (100mM)	10µl
Tris-HCl pH7.5 (250mM)	110µl

The solution was prepared in a 1.5ml reaction tube and heated to 95°C for 5min in a heating block. After the incubation, the heating block was switched off and the tube was cooled in the heating block over night. Subsequently, the solution was filled up to a total volume of 500µl with H_2O bidest., transferred to a Microcon YM-30 centrifugation filter and centrifuged at 14,000 x g for 10min. The linker cassette was then eluted from the inverted filter tube into a fresh 1.5ml reaction tube at 1,000 x g for 2min. Finally, the volume of the oligonucleotide solution was adjusted to 80µl, and stored at -20°C in 10µl aliquots.

Linear PCR

For linear amplification, 500ng or 1µg of genomic DNA was inserted. To increase the total number of IDLV integration sites, triplicates for each sample were inserted into LAM-PCR. Following the first 50 cycles of linear amplification, 1µl Taq DNA polymerase (5U) was added to each sample, and amplification was repeated for another 50 cycles. IDLV transduced samples were amplified with the biotinylated primers SK-LTR 1bio and SK-LTR 3bio (0.167µM each). As controls for LAM-PCR, non-transduced human genomic DNA from Roche and H_2O bidest were used.

Reagent	Amount
H_2O bidest	Ad 50µl
10x Taq DNA Polymerase Buffer	5µl
dNTP mix (dATP, dGTP, dCTP, dTTP) (10mM each)	1µl
Primer 1 (0.167µM)	0.25µl
Primer 2 (0.167µM)	0.25µl
Taq DNA Polymerase (5U/µl)	0.25µl
DNA	500ng/1µg

Description	Temperature	Time	Repetitions
Initial Denaturation	95°C	2min	
Denaturation	95°C	45sec	
Annealing	60°C	45sec	2x50 cycles
Elongation	72°C	1min	
Final Elongation	72°C	5min	
Cooling	4°C		

Magnetic Capture of amplicons

For each PCR reaction, 20µl magnetic beads were prepared. The magnetic beads were transferred into a fresh 1.5ml reaction tube and placed in a magnetic particle concentrator (MPC) for 1min. The storage solution was discarded and the beads resuspended in 40µl 0.1% BSA-DPBS solution per reaction. Again, the tube was placed in the MPC and the supernatant discarded. The washing step was subsequently repeated. Afterwards, the beads were washed with 20µl of the 3M LiCl solution per reaction. Finally, the beads were resuspended in 50µl 6M LiCl solution per reaction and mixed with 50µl linear PCR product in the PCR tube. To allow biotinylated DNA amplicons to bind to the streptavidin-coated paramagnetic beads, the tubes were placed in a horizontal shaker at 300 rpm and RT over night.

Hexanucleotide priming

During hexanucleotide priming, the ssDNA amplicons from the linear amplification were converted into dsDNA by the bacterial Klenow polymerase. The reaction mastermix was prepared on ice as followed:

Reagent	Amount per reaction
H$_2$O bidest	16.5µl
10x Hexanucleotide Mix	2µl
dNTP mix (dATP, dGTP, dCTP, dTTP) (10mM each)	0.5µl
Klenow DNA polymerase	1µl

To each PCR tube, 80µl H$_2$O bidest was added and the beads transferred into a 96-Well plate. Subsequently, the beads were exposed to a MPC for 1min and the supernatant discarded. Washing was repeated once with 100µl H$_2$O bidest and the 96-Well plate was put in a cooling rack. Finally, 20µl hexanucleotide mastermix was added to each well and incubation was performed at 37°C for 1h.

Restriction digest of amplicons

The DNA-paramagnetic particle complexes were washed once with 80µl and once with 100µl H$_2$O bidest. The bead solution was resuspended in 20µl of the restriction digest mastermix.

Reagent	Amount per reaction
H$_2$O bidest	8.7µl
10x CutSmart buffer	1µl
Restriction Enzyme (5U/µl)	0.3µl

Restriction enzymes, which do not have a recognition site in the LTR of IDLV were chosen for LAM-PCR. Here, MseI and MluCI/Tsp509I were used. The restriction mixture was incubated in a PCR cycler at 37°C (MseI, MluCI) or 64°C (Tsp509I) for 2h.

Ligation of linker cassette

To each restriction digest mixture, 80µl H$_2$O bidest were added and the 96-Well plate placed on a MPC for 1min. The supernatant was discarded and the washing step repeated with 100µl H$_2$O bidest The DNA-particle mix was resuspended in 10ul ligation mastermix.

Reagent	Amount per reaction
H$_2$O bidest	5µl
10x Fast Link Ligation Buffer	1µl
ATP (10µM)	1µl
Fast Link DNA ligase (2U/µl)	1µl
Linker cassette	2ul

The ligation was performed at RT for 5min. To stop ligation reaction, the 96-Well plate was placed on ice. For Illumina MiSeq pyrosequencing, linker cassettes with unique DNA recognition sequences, so called barcodes, for each sample were used in the ligation reaction.

Denaturation

The total volume in each well was filled up to 100µl with 90µl H$_2$O bidest. Then, the 96-Well plate was placed on a MPC, the supernatant discarded and washing repeated with 100µl H$_2$O bidest. Subsequently, the beads were resuspended in 5µl 0.1N NaOH solution and incubated at 300rpm on a horizontal shaker and at RT for 10min.

Again, the 96-Well plate was placed on a MPC and the supernatant transferred into fresh 0.5ml reaction tubes. The DNA was either stored at -20°C or used immediately for exponential amplification.

Nested PCR for exponential amplification of vector-genome junctions

The exponential PCR was performed using nested primers that bind specifically to the LTR and linker cassette sequence. For the first nested PCR, 1µl of linear PCR product and the biotinylated primers SK-LTR 4bio (3' LAM-PCR) as well the linker cassette specific primer LCI were inserted.

Reagent	Amount per reaction
H$_2$O bidest	Ad 25µl
10x Taq DNA Polymerase Buffer	2.5µl
dNTP mix (dATP, dGTP, dCTP, dTTP) (10mM each)	0.5µl
Primer 1 (16.7µM)	0.125µl
LCI (16.7µM)	0.125µl
Taq DNA Polymerase (5U/µl)	0.25µl
DNA	1µl

Description	Temperature	Time	Repetitions
Initial Denaturation	95°C	2min	
Denaturation	95°C	45sec	
Annealing	60°C	45sec	35 cycles
Elongation	72°C	1min	
Final Elongation	72°C	5min	
Cooling	4°C		

For capturing of the biotinylated PCR products, 20µl magnetic beads per PCR reaction were prepared as described (see magnetic capture of amplicons). The beads were resuspended in 25µl 6M LiCl solution per reaction and mixed with 25µl PCR product in the PCR tube. To allow biotinylated DNA amplicons to bind to the streptavidin-coated paramagnetic beads, the tubes were placed in a horizontal shaker at 300 rpm for 2h. Afterwards, 80µl H$_2$O bidest was added to each PCR reaction tube and the beads transferred into a new 96-Well plate, which was subsequently placed on a MPC. The supernatant was discarded and the washing step repeated twice with 100µl H$_2$O bidest. Next, beads were resuspended in 5µl 0.1N NaOH, and the plate was incubated at 300rpm in a horizontal shaker at RT for 10min. The plate was placed on a MPC and the supernatant transferred into a fresh 0.5ml reaction tube. For the second exponential PCR, 1µl denaturation product was inserted. The linker-specific primer used in this PCR was LCII, and the vector-specific primer was either SK-LTR 5 (3' LAM-PCR).

Reagent	Amount per reaction
H$_2$O bidest	Ad 50µl
10x Taq DNA Polymerase Buffer	5µl
dNTP mix (dATP, dGTP, dCTP, dTTP) (10mM each)	1µl
Primer 1 (16.7µM)	0.25µl
LCII (16.7µM)	0.25µl
Taq DNA Polymerase (5U/µl)	0.5µl
DNA	1µl

Description	Temperature	Time	Repetitions
Initial Denaturation	95°C	2min	
Denaturation	95°C	45sec	
Annealing	60°C	45sec	35 cycles
Elongation	72°C	1min	
Final Elongation	72°C	5min	
Cooling	4°C		

To analyze the LAM-PCR products, 2µl of the amplicons were mixed with 8µl loading dye and loaded onto a 2% agarose gel.

2.2.2.8   Cleaning-Up of PCR products using AMPure XP beads

In order to attach sequencing adaptors to PCR amplicons or to subject samples to pyrosequencing, the LAM-PCR products were cleaned up. Here, AMPure XP beads from Beckman Coulter were used according to manufacturer's protocol. For the attachment of sequencing adaptors, the beads-to-PCR product ratio was 1.1:1. For pyrosequencing, the ratio of beads-to-PCR product was changed to 0.65:1.

2.2.2.9   Fusionprimer-PCR

Fusionprimer-PCR was performed to attach Roche/454- or Illumina-specific sequencing adapters to the LAM-PCR amplicons. For Roche/454 sequencing, an adaptor sequence was attached to the linker sequence in order to enable binding of the PCR amplicons to specific beads during sequencing. At the vector sequence, an adaptor was attached in order to enable sequencing. For Illumina sequencing, two primers were attached to the vector- and linker sequence to facilitate binding of the amplicons to an array surface during sequencing. The vector-specific adaptor contained an additional 6 or 10bp long unique recognition sequence, called barcode, which enables massive parallel sequencing and later sorting of the sequences. For 454 and Illumina sequencing, 50ng cleaned-up LAM-PCR products were inserted into PCR. Each Fusion-PCR primer consists of the specific sequencing-adaptor and a vector- or linker-specific sequence.

Reagent	Amount per reaction
H$_2$O bidest	Ad 50µl
10x Taq DNA Polymerase Buffer	5µl
dNTP mix (dATP, dGTP, dCTP, dTTP) (10mM each)	1µl
Fusionprimer forward (5µM)	0.5µl
Fusionprimer reverse (5µM)	0.5µl
Taq DNA Polymerase (5U/µl)	0.5µl
LAM-PCR amplicon	50ng

Description	Temperature	Time	Repetitions
Initial Denaturation	95°C	2min	
Denaturation	95°C	45sec	
Annealing	60°C	45sec	15 cycles
Elongation	72°C	1min	
Final Elongation	72°C	5min	
Cooling	4°C		

The PCR products were subjected to gel electrophoresis, cleaned up, and measured spectrometrically using the Nanodrop-1000.

2.2.2.10   SureSelect Target Enrichment for Illumina Multiplexed Sequencing

The SureSelect Target Enrichment protocol is used to enrich target regions of the genome from unrelated sequences and prepare the samples for Illumina sequencing. Prior to enrichment, pre-capture libraries were prepared that target specific sequences. These target sequences among others include SIN lentiviral vectors. The protocol is divided into sample preparation, hybridization, sample processing for multiplexed sequencing, and sequencing (Figure 10).
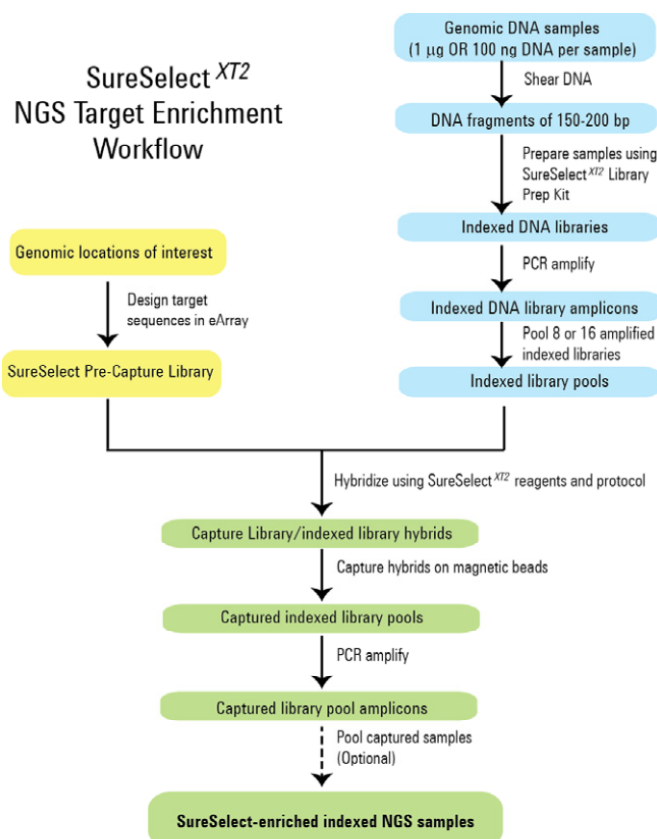
Figure 10: SureSelecet Target enrichment workflow for Illumina sequencing. In the first step, the genomic DNA is sheared into fragments of 150-250bp size. Each sample is subsequently indexed, amplified and pooled with other indexed libraries. Target sequences are then enriched using a SureSelect Pre-Capture Library, captured and amplified by PCR. The indexed library is finally subjected to Illumina HiSeq sequencing. Image taken from Agilent Technologies manual (SureSelect XT2 Target Enrichment System for Illumina Multiplexed Sequencing, Version D.0, September 2013).

Sample Preparation

Step 1: Shear genomic DNA

A total of 1µg genomic DNA of IDLV-transduced, irradiated and expanded NHDF-A was diluted in 50µl H_2O bidest. The solution was transferred into a Covaris microtube and subsequently sheared using the following settings: Duty Factor: 10%, Intensity Level: 5, Cycles per Burst: 200, Treatment Type: 120sec, yielding DNA fragments with an average size of 250bp. Then, the DNA fragments were transferred into fresh 0.5ml reaction tubes and stored at -20°C until further usage.

Step 2: Assess quality of sheared DNA (skipped)

Step 3: Repair the ends

In order to repair the DNA ends, 50µl SureSelect End-Repair Mastermix was added to each sheared DNA sample and mixed well by pipetting the solution up and down. The mixture was incubated at 20°C in a thermal cycler without using the heated lid. Finally, the mixture was cooled down and kept at 4°C.

Step 4: Purify the samples using AMPure XP beads

Prior to purification, the AMPure XP beads were warmed at RT for 30 minutes. For each sample, 500µl 80% ethanol was prepared. Then, 160µl beads were mixed with 100µl end-repaired DNA (beads-to-DNA ratio: 1.6:1) and mixed by pipetting up and down ten times, followed by 5min incubation at RT. Then, the 0.5ml reaction tubes were placed into a MPC for 4 minutes and the supernatant removed. While washing each sample twice with 200µl 80% ethanol, the reaction tubes were kept in the magnetic stand. The residual ethanol was removed, the tubes removed from the MPC, 22µl nuclease-free H_2O added to each sample and incubated for 2min at RT before placing the tubes back into the MPC. The DNA-containing solutions were transferred to fresh 0.5ml reaction tubes.

Step 5: Adenylate 3' ends of DNA fragments

To add a single adenine nucleotide to the 3' end of the DNA fragments, 20µl dA-Tailing Master Mix was added to each sample, mixed and incubated in thermal cycler at 37°C for 30 minutes without using the heated lid. The mixture was cooled down and kept at 4°C.

Step 6: Ligate the pre-capture indexing adaptors

In this step, DNA adaptors carrying unique DNA barcodes, called index, were ligated to each DNA sample to generate an indexed library. To each A-tailed DNA sample, 5µl Ligation Master Mix and 5µl of the appropriate Pre-Capture Index Solution were added. The mixtures were vortexed for 5 seconds, spinned briefly in a table-top centrifuge, and incubated in thermal cycler at 20°C for 15min without using the heated lid.

Step 7: Purify the indexed DNA using AMPure XP beads

Purification was performed as described above (step 4) with two changes. To each index sample, 36µl beads were added (Beads-to-DNA ratio: 1.2:1), and the purified DNA was resuspended in 50µl nuclease-free H_2O.

Step 8: Amplify the indexed library

For amplification, the following Pre-capture PCR reaction mix was prepared on ice. For each indexed sample, 1µl XT2 Primer Mix and 25µl Herculase II PCR Master Mix were mixed and combined with 24µl of each indexed library in a fresh 96-Well PCR plate. The following program was run in a PCR thermal cycler:

Temperature	Time	Repetitions
98°C	2min	
98°C	30sec	
60°C	30sec	5 Cycles
72°C	1min	
72°C	10min	
4°C	hold	

Step 9: Purify the amplified library with AMPure XP beads

Purification was performed as described above (step 8) with a single change ion protocol: to each amplified library, 50µl beads were added (Beads-to-DNA ratio: 1:1).

Step 10: Assess quality with the 2100 Bioanalyzer DNA 1000 Assay

To assess the quality of each indexed library, 1µl of each sample and 1µl D1000 Ladder were mixed with 3µl D1000 Sample Buffer each in a tube strip, vortexed for 3 sec and placed in a 2200 TapeStation. Finally, gel electrophoresis was run and quality of amplified libraries analyzed.

Hybridization

Step 1: Pool indexed DNA samples for hybridization

For each hybridization reaction, a total of 2000ng indexed DNA library was required. Here, 19 amplified libraries were combined (111ng of each library) in a single 1.5ml reaction tube. Subsequently, the samples were completely dried in a vacuum concentrator at 37°C for 3h. Then, the DNA was resuspended in 7µl TE buffer (PCR purification kit, Qiagen) and dissolved for 4h at 300rpm and RT in a horizontal shaker.

Step 2: Hybridize DNA library pools to the SureSelect Capture Library

To each pooled genomic DNA (gDNA) library, 9µl SureSelect Blocking Mix was added and the solution transferred into a fresh PCR reaction tube. The tube was placed in PCR cycler at 95°C for 5min and then held at 65°C. In the meantime, the Capture Library/RNase Block Mix was prepared by mixing 1µl SureSelect Capture Library and 0.5µl SureSelect RNase Block with 5.5µl H_2O bidest. Subsequently, 37µl SureSelect hybridization buffer was added to the 7µl Capture Library/RNase Block Mix. While maintaining the gDNA pool at 65°C, 44µl Capture Library Mix was added and mixed by pippetting up and down. The hybridization mixture was kept at 65°C for 67h with the heated lid at 99°C.

Step 3: Prepare streptavidin magnetic beads for DNA hybrid capture

First, SureSelect Wash buffer #2 was pre-warmed at 65°C in a water bath. Second, 50µl Dynabeads MyOne Streptavidin T1 were transferred to a fresh 0.5ml reaction tube and washed by adding 200µl SureSelect Binding buffer. The tube containing the beads was vortexed for 5sec, and the supernatant removed from the beads on a

MPC. This washing step was repeated for a total of three washes. Finally, beads were resuspended in 200µl SureSelect Binding buffer.

Step 4: Capture the hybridized DNA using streptavidin beads

The hybridization mixture was kept at 65°C while the entire hybridization mixture was transferred into the tube containing the 200µl washed Dynabeads from step 3. The solution was mixed, and incubated on a horizontal shaker at RT and 300rpm for 30min. The tube was placed in a MPC and the supernatant removed. The beads were resuspended in 200µl SureSelect Wash buffer #1, mixed, and the tube placed back into the MPC. The supernatant was subsequently removed and the beads washed with 200µl pre-warmed SureSelect Wash buffer #2. The beads were resuspended, incubated at 65°C for 5min in a heating block, and the tube placed into a MPC. The supernatant was discarded and washing with buffer #2 was repeated for a total of 6 washes. Finally, the beads were mixed with 30µl nuclease-free H_2O and stored at 4°C.

Post-Capture Sample Processing for Multiplexed Sequencing

Step 1: Amplify the captured libraries

For amplification of the captured library, the following PCR reaction mix was prepared on ice. For each library, 25µl Herculase II Master Mix, 1µl XT2 Primer Mix and 9µl nuclease-free H_2O were mixed and added to 15µl of the captured library pool bead suspension in a PCR reaction tube. The following program was run in a PCR thermal cycler:

Temperature	Time	Repetitions
98°C	2 min	
98°C	30 sec	
60°C	30 sec	14 cycles
72°C	1 min	
72°C	10 min	
4°C	hold	

Step 2: Purify the amplified captured libraries using AMPure XP beads

Purification was performed as described above (Sample Preparation, step 4) with two changes: To each amplified and captured library, 80µl beads were added (Beads-to-DNA ratio: 1.6:1). The beads were resuspended in 15µl H_2O bidest, the tube was placed in a MPC and the supernatant transferred to a fresh reaction tube.

Step 3: Assess quality with the 2100 Bioanalyzer High Sensitivity DNA assay

For assessing the quality of the amplified and captured library, 1µl of each library and 1µl D1000 Ladder were mixed with 3µl D1000 Sample Buffer each in a tube strip, vortexed for 3sec and placed in a 2200 TapeStation. Finally, gel electrophoresis was run and quality of amplified libraries was analyzed.

Step 4: Prepare samples for multiplexed sequencing

To sequence the DNA using HiSeq 2000, 3µl of the DNA solution was diluted 1:10 in H_2O bidest. Sequencing was performed at the DKFZ Genomic and Proteomics Core Facility.

2.2.2.11   Tdt-mediated labeling of DSB sites

For the direct labeling of free DNA ends the TUNEL assay (TdT-mediated dUTP-biotin nick end labeling) protocol was modified.

Fixation and permeabilization

In the first step, $1x10^6$ A549 cells were washed once with 1xDPBS and centrifuged at 1,000rpm and RT for 5min. Then, the cells were fixed with 4% paraformaldehyde (PFA) at 37°C and 600rpm for 10min in a heating block. Subsequently, the fixed cells were washed with 1xDPBS before permeabilization with 0.25% Triton X-100 solution for 20min in a heating block at 37°C and 600rpm.

dUTP labeling of free DNA ends

Residual PFA and Triton X-100 solution were removed by washing the cells twice with H_2O bidest. Centrifugation was performed at 1,000rpm for 5min at RT. In order to label free DNA ends with UTP nucleotides, the following Tdt master mix was prepared in a 1.5ml reaction tube is shown in the table below:

Reagent	Amount
H_2O bidest	53.5 µl
5x Tdt reaction buffer	20 µl
$CoCl_2$ solution (25mM)	20 µl
Biotin-16-dUTP (1mM)	5 µl
dUTP (100mM)	0.5 µl
Tdt (400U/µl)	1 µl

As controls for the labeling reaction, EcoRI- or PvuI-digested LV902 plasmid and H_2O bidest were included. The cells and control labeling reactions were incubated with the Tdt master mix at 37°C and 600rpm in a heating block for 60min. In order to block Tdt labeling activity, the 1.5ml was placed on ice and 10µl 0.2M EDTA (pH8) added. Then, the total volume was filled up to 200µl with 1xDPBS.

DNA isolation

To isolate the genomic DNA from the fixed cells, the cells were lysed with 200µl lysis buffer (4M urea, 200mM Tris-HCl, 20mM NaCl, 200mM EDTA, pH7.4) and 40µl proteinase K (1µg/ml) for 1h at 55°C. The DNA was subsequently precipitated by adding 440µl isopropanol. Next, the DNA was centrifuged for 15min at full speed in a table top centrifuge. The supernatant was carefully removed and the DNA washed with 100µl 70% ethanol. The centrifugation step was repeated, the ethanol removed and the DNA dried at RT for 5min. Finally, the DNA was resuspended in 50µl TE buffer.

Magnetic Capture of amplicons

For each DNA sample, 20µl magnetic beads were transferred into a fresh 1.5ml reaction tube and placed in a magnetic particle concentrator (MPC) for 1min. The storage solution was discarded and the beads resuspended

in 40µl 0.1% BSA-DPBS solution per reaction. Again, the tube was placed in the MPC and the supernatant discarded. The washing step was repeated once. Afterwards, the beads were washed with 20µl of the 3M LiCl solution per reaction. Finally, the beads were resuspended in 50µl 6M LiCl solution per reaction and mixed with 50µl labeled genomic DNA in a 96-Well plate. To allow biotinylated DNA to bind to the streptavidin-carrying paramagnetic beads, the plate was placed in a horizontal shaker at 300 rpm and RT over night.

Restriction digest of captured DNA

The DNA-paramagnetic particle complexes were washed once with 100µl 0.1% Tween-20 solution and once with 100µl H_2O bidest. The bead solution was resuspended in 20µl of the restriction digest mastermix.

Reagent	Amount per reaction
H_2O bidest	17.6µl
10x NEBuffer 1	2µl
Tsp509I (5U/µl)	0.4µl

The restriction mixture was incubated in a PCR cycler at 65°C (Tsp509I) for 1h.

Ligation of linker cassettes

To block restriction enzyme activity, 80µl H_2O bidest were added to each reaction, and the 96-Well plate placed on a MPC for 1min. The supernatant was discarded and the washing step repeated with 100µl H_2O bidest. The DNA-particle mix was resuspended in 10µl ligation mastermix.

Reagent	Amount per reaction
H_2O bidest	5µl
10x Fast Link Ligation Buffer	1µl
ATP (10µM)	1µl
Fast Link DNA ligase (2U/µl)	1µl
Linker cassette	2µl

The ligation was performed at RT for 5min. To stop ligation reaction, the 96-Well plate was placed on ice. The linker cassettes inserted into the ligation reaction were the same ones used for LAM-PCR and had an AATT DNA overhang complementary to the free DNA overhang introduced by the restriction enzyme Tsp509I.

Nested PCR for exponential amplification of linker-genome junctions

The total volume in each well was filled up to 100µl with 90µl H_2O bidest. Then, the 96-Well plate was placed on a MPC, the supernatant discarded and washing repeated with 100µl H_2O bidest. The exponential PCR was performed using nested PCR primers that bind specifically to the polyU tail and linker cassette sequence. For the first PCR, 2µl DNA-beads solution and the polyA-primer L.CDSB2-1 as well as the linker cassette specific primer LCI were inserted.

Reagent	Amount per reaction
H$_2$O bidest	Ad 25µl
10x Taq DNA Polymerase Buffer	2.5µl
dNTP mix (dATP, dGTP, dCTP, dTTP) (10mM each)	0.5µl
L.CDSB 2-1 (16.7µM)	0.25µl
LCI (16.7µM)	0.25µl
Taq DNA Polymerase (5U/µl)	0.25µl
DNA-bead solution	2µl

Description	Temperature	Time	Repetitions
Initial Denaturation	95°C	2min	
Denaturation	95°C	45sec	
Annealing	60°C	45sec	35 cycles
Elongation	72°C	1min	
Final Elongation	72°C	5min	
Cooling	4°C		

For the second exponential amplification, 5µl PCR product from the first PCR reaction was directly inserted. The linker-specific and polyU-specific primers used in this PCR were LCII and E.CDSB 4, respectively.

Reagent	Amount per reaction
H$_2$O bidest	Ad 50µl
10x Taq DNA Polymerase Buffer	5µl
dNTP mix (dATP, dGTP, dCTP, dTTP) (10mM each)	1µl
E.CDSB 4 (16.7µM)	0.5µl
LCII (16.7µM)	0. 5µl
Taq DNA Polymerase (5U/µl)	0.5µl
DNA	5µl

Description	Temperature	Time	Repetitions
Initial Denaturation	95°C	2min	
Denaturation	95°C	45sec	
Annealing	60°C	45sec	35 cycles
Elongation	72°C	1min	
Final Elongation	72°C	5min	
Cooling	4°C		

To analyze the PCR products, 2µl of the amplicons were loaded onto a 2% agarose gel.

2.2.2.12   Biotin Quantification

In order to accurately measure the biotinylation level of labeled DNA in the Tdt-mediated DSB labeling assay and to determine the optimal amount of biotin-16-dUTP, the TermoScientific Fluorescence Biotin Quantitation Kit was used.

The Fluorescence Biotin Quantitation Kit is a colorimetric assay to estimate the biotin-to-molecule ratio. It consists of two reagents: HABA (4'-hydroxyazobenzene-2-carboxylic acid) and avidin. In a premix (DyLight Reporter), fluorescent avidin is mixed with HABA, a dye that weakly interacts with avidin. The premix is added to the solution containing the biotinylated DNA. Since biotin has a higher affinity for avidin, biotin displaces the HABA, allowing the avidin to fluorescate. The level of fluorescence is proportional to the amount of biotin in the sample and is measured in a microplate reader by comparing the fluorescence to a biocytin standard curve.

      Preparation of a diluted biocytin standard

In order to quantify the biotinylation level in each sample, a standard curve based on a biocytin standard dilution series was prepared and measured in a microplate reader (excitation wavelength: 495nm). First, 1xDPBS (diluent) was prepared by adding 9.5ml ultrapure water to 0.5ml 20xDPBS.

Vial	Volume of diluent	Volume of biocytin standard	Final biocytin concentration [pmol/10µl]
A	198µl	2µl of biocytin control	100
B	10µl	40µl of vial A	80
C	20µl	30µl of vial A	60
D	30µl	20µl of vial A	40
E	40µl	10µl of vial A	20
F	45µl	5µl of vial A	10
G	47.5µl	2.5µl of vial A	5
H	50µl	0	0

      Preparation of DyLight Reporter Working Reagent (DWR)

To determine the total volume of DWR required the following formula was used:

$$(\# \ standards \ + \ \# \ unknowns) \ \times \ (\# \ replicates) \ \times \ (90 \ \mu l \ DWR) \ = \ total \ volume \ DWR$$

Immediately before use, 14 parts of 1xDPBS were mixed with 1 part of DyLight Reporter to obtain the calculated amount of DWR solution.

      Biotin quantification

For each standard and unknown sample, 10µl triplicates were transferred into a microplate well and 90µl 1xDWR solution was added. After mixing, the plate was wrapped in aluminum foil to protect the solution from light exposure and incubated at RT and 200rpm for 5min in a horizontal shaker. Finally, the microplate was placed in a microplate reader and the fluorescence measured after excitation with a laser at a wavelength of 495nm. Then, a reverse plot of the linear range (10-60pmol biocytin/10µl reaction) of the standard curve was

prepared and a linear regression equation generated. The resulting equation was used to determine picomoles of biotin in the unknown samples by inserting the sample's average fluorescence intensity.

2.2.2.13   Linker-Amplification-Mediated DSB Trapping (LAM-DST)

In order to label DSB sites *in situ*, a ligation approach with DNA linkers of known sequence was developed. First, the DNA break is labeled and subsequently enriched on streptavidin-coated paramagnetic beads. The captured DNA is then amplified and identified by LAM-PCR and next-generation sequencing, respectively.

Fixation and permeabilization

In the first step, 1×10^6 293T cells transfected with CCR5-specific ZFN or with the plasmid LV902 as a control for 72h, detached from the 6-Well plate, washed once with 1xDPBS and centrifuged in a 15ml reaction tube at 1,000rpm and RT for 5min. Then, the cells were fixed with 4% PFA at 37°C for 10min with occasional shaking. Subsequently, the fixed cells were washed with 10ml 1xDPBS before fixation with 0.25% Triton X-100 solution at RT for 20min. The cells were then washed three times with 1xDPBS and resuspended in 800µl 1xDPBS. The cells were subsequently split into 8 samples (100µl each) into a 96-Well plate on ice.

Ligation of DNA linker adapters

Before the ligation of the linker adapters to the DSB sites, the 1xDPBS was removed and the cells incubated with 1xT4 DNA ligase or 1xCircLigase buffer (100µl) incubated at 37°C for 15min. During this pre-incubation step, the ligation mastermix for single-stranded linker and double-stranded linker ligation was prepared as follows:

Reagent	Amount per reaction
10x CircLigase ligation buffer	3µl
MnCl$_2$ (50mM)	1.5µl
ATP (1mM)	1.5µl
PEG 8,000	3µl
ssGRUV5-A-Phthio linker (100mM)	2µl
CircLigase ssDNA Ligase (100U/µl)	2µl
H$_2$O bidest	17µl

Reagent	Amount per reaction
10x T4 DNA ligation buffer	5µl
dsGRUV2 linker (100mM)	1µl
T4 DNA Ligase (100U/µl)	2µl
H$_2$O bidest	42µl

The pre-incubation buffer was removed and the cells resuspended in the ligation mastermix. Ligation of the ssDNA linker was performed at 60°C for 2h, whereas the dsDNA linker ligation was incubated in a PCR cycler at 16°C for 16h. As a ligation control, DNA was replaced with water.

DNA isolation and measuring DNA concentration

For DNA isolation, the Roche High Pure PCR Template Preparation Kit was used according to the manufacturer's protocol, and the DNA concentration of each sample was determined using the Nanodrop-1000.

CCR5 locus-specific PCR

In order to show successful labeling of ZFN on-target activity sites, two subsequent rounds of nested CCR5 locus-specific PCR were performed using the primers CCR5fwd1 and 2 and U5VI bio primers.

Reagent	Amount per reaction
H_2O bidest	ad 25µl
10x Taq DNA Polymerase Buffer	2.5µl
dNTP mix (dATP, dGTP, dCTP, dTTP) (10mM each)	0.5µl
CCR5 fwd 1 (10µM)	0.25µl
U5VIbio (10µM)	0.25µl
Taq DNA Polymerase (5U/ul)	0.5µl
DNA	200ng

The mastermix was prepared on ice and subjected to the following amplification program:

Description	Temperature	Time	Repetitions
Initial Denaturation	95°C	2min	
Denaturation	95°C	45sec	
Annealing	58°C	45sec	30 cycles
Elongation	72°C	30sec	
Final Elongation	72°C	2min	
Cooling	4°C		

Following the first 30 cycles, 5µl of the PCR product were subjected to a second round of amplification using nested PCR primers. The mastermix is shown in the table below. The PCR cycling program was identical to the program in the first PCR.

Reagent	Amount per reaction
H_2O bidest	37µl
10x Taq DNA Polymerase Buffer	5µl
dNTP mix (dATP, dGTP, dCTP, dTTP) (10mM each)	1µl
CCR5 fwd 1 (10µM)	0.5µl
U5VIbio (10µM)	0.5µl
Taq DNA Polymerase (5U/µl)	1µl
DNA	5 µl

The PCR products were loaded onto a 2% agarose gel for gel electrophoresis. Successful labeling of the CCR5 locus results in the amplification of a 200bp DNA fragment. In addition, to verify site-specific labeling on the sequence level, the PCR amplicon was cloned into the pCR2.1-TOPO TA cloning vector and sent to GATC (Konstanz, Germany) for Sanger sequencing.

LAM-PCR for the detection of genome-wide DSB

In addition to CCR5 locus-specific PCR, the DNA samples were subjected to LAM-PCR. For linear amplification of the linker-genome junctions, the primers U5V bio and U5VI bio (1.6µM each) were inserted. During nested PCR for exponential amplification, U5VI bio and GRUV5-2expo (16.6µM each) were used.

2.2.2.14   Pyrophosphate sequencing

Pyrophosphate sequencing of LAM-PCR products and the SureSelect libraries was performed by GATC (Konstanz, Germany) or the DKFZ Genomic and Proteomics Core Facility (Heidelberg, Germany). For sequencing, the amplicons carrying sequencing-specific adapters were mixed in specific ratios according to the desired sequencing depth (for 454: 20ng = 1,000 sequences; for Illumina MiSeq: 3µg = $1x10^7$ sequences).

2.2.2.15   Cloning of PCR amplicons using the TOPO-TA Cloning Kit

For fast and efficient cloning of single PCR products, the commercial TOPO-TA-Kit was used. It is based on the observation that DNA polymerases do not create blunt-ends during PCR, but introduce single adenosine bases at the 3' end of the PCR product [80]. Hence, the linearized cloning vector, pCR2.1-TOPO TA, has two single thymidine overhangs at the 3' ends, enabling complementary pairing of the vector and PCR product. Two molecules of the topoisomerase from vaccinia virus are covalently attached to the vector. The topoisomerase has a ligation activity, which mediates the ligation of the PCR product into the vector. During this reaction, the topoisomerase is released from the vector and a circular vector-PCR product is formed.

For cloning of PCR products, the following reaction was mixed:

Reagent	Amount
PCR product	2µl
Salt Solution	1µl
pCR2.1-TOPO TA vector	1µl
H_2O bidest	Ad 5µl

The reagents are mixed on ice and incubated for 5min at RT. Until transformation into chemically-competent *E.coli*, the mixture was stored on ice.

2.2.2.16   Transformation of circular DNA into chemically-competent *E.coli*

Methods for the introduction of DNA molecules into bacterial cells are called transformation. Here, plasmid DNA molecules were introduced into chemically-competent *E.coli* cells by a heat-shock step. The *E.coli* strains stbl3 and OneShot TOP10 were purchased from Qiagen.

Prior to transformation, *E.coli* cells stored at -80°C were thawed on ice for 3min. Subsequently, half the volume of cells was transferred into a fresh, sterile 1.5ml reaction tube, kept on ice and also used for transformation.

Next, between 1-5µl plasmid DNA was added to the cells and incubated for 30min on ice, before being exposed to a 30sec heat-shock at 42°C. This forces the DNA to enter the cells through either cell pores or the damaged cell wall. The cells are subsequently placed on ice for 2min. Then, 200µl S.O.C. medium was added to each reaction tubes, and the cells incubated at 37°C for 2h, before platting on LB-agar plates supplemented with ampicillin (100µg/ml). The plates were incubated at 37°C over night

2.2.2.17   Mini- and maxipreparation of plasmid DNA

Plasmid preparations are used to extract and purify plasmid DNA from bacteria. According to the size of bacterial culture and plasmid yield, the method is subdivided into mini-, midi- or maxiprep. Here, plasmid preparations from mini and maxi cultures were performed, which involve three steps: growth of bacterial culture, harvesting and lysis and plasmid purification. For mini- and maxiprep, the QIAprep spin miniprep and QIAprep spin maxiprep kits were used according to manufacturer's protocol.

2.2.2.18   Enzymatic DNA restriction digest

Restriction digest is a method to prepare plasmid DNA for analysis and cloning. It takes advantage of naturally-occurring enzymes that recognize and cleave DNA at specific sequence sites, thereby generating DNA fragments of specific size or producing compatible ends for cloning. The most commonly-used restriction enzymes are Type II restriction endonucleases with 4-8 bp long palindromic DNA recognition sequences. Here, restriction enzymes were used to control cloning efficiency and lentiviral plasmid preparations.

Below, a typical DNA restriction digest is shown:

Reagent	Amount
Plasmid (1µg/ml)	1µl
10x Digestion buffer	1µl
Restriction enzyme	1µl
H_2O bidest	Ad 10µl

The digestion was performed at 37°C for 2h and the fragments analyzed by gel electrophoresis.

2.2.3   Bioinformatical Methods

2.2.3.1   Automated Sequence Analysis (HISAP)

To map and analyze the sequences obtained from LAM-PCR and deep sequencing, the raw sequences were processed [81, 82]. To annotate IDLV insertion site informations, the high-throughput integration site analysis pipeline (HISAP) was used [83]. Starting from raw 454 or Illumina sequences, HISAP removes all vector- and LAM-PCR- specific sequence parts and clusters identical sequences, which are subsequently aligned to the human genome (GRch37/hg19) using BLAT [81] in order to identify the genomic DSB loci. Additional information such as nearby RefSeq genes, their distance and orientation to the integration site, and distance to the closest CpG

island are also calculated. DSB sites from different samples mapping to the same genomic location were called collisions and processed according to their sequence count. If the read count of one the DSB sites was more than four-times higher than the other, the DSB site with the lower sequence count was deleted. DSB sites with the same sequence count were both excluded from the DSB data set. Further, DSB sites exclusively located in repetitive genomic regions were also excluded.

2.2.3.2   A549 mRNA expression analysis

For A549 gene expression analysis, triplicate microarray data sets were used. These data sets are publicly available at the GEO database server and have the GEO accession numbers GSM661204, 661205 and 661206. Prior to correlation of the expression data with the DSB data set, the mRNA data were bioinformatically processed. First, the expression data were binned according to their expression activity. Therefore, a threshold was set which was twice the background signal intensity (dark corner). Then, the mRNA microarray probes were broadly divided according to their signal intensity into active (>2x threshold) and silent (<2x threshold) transcriptional mRNA signatures. Subsequently, active mRNA signatures were further subdivided into ten bins according to their signal value ranging from low to high expression activity. Moreover, the mRNA expression signatures were correlated to the A549 DSB sites. A DSB site was termed mRNA signature associated when one or more probe signals were found 250kb up- or downstream of the DSB site. Further, a DSB site was considered to be located in a transcribed genomic region, if more than four active mRNA signatures were found within 250kb of the DSB site.

2.2.3.3   DNaseI Hypersensitive Sites

DNaseI Hypersensitive Sites (DHS) are genomic regions with increased accessibility for the endonuclease DNaseI, and hence represent open chromatin. For DHS analysis, 117,818 and 372,215 DHS peak data for A549 and NHDF-A with the GEO sample accession numbers GSM816649, GSM736567 and GSM736520 were obtained from the ENCODE project repository. These data were subsequently correlated with the A549 and NHDF-A DSB sites. A DSB was considered to be DHS associated when it was located in between the starting and ending position of the DHS.

2.2.3.4   Histone modifications and Transcription Factor Binding Sites

For histone modification analysis, the following A549 and NHDF-A histone peak data were downloaded from the ENCODE project repository: H3K4me1 (GEO accession numbers: GSM1003495, GSM1003453, GSM1003526), H3K4me2 (GSM1003511, GSM1003496, GSM733753), H3K4me3 (GSM1003542, GSM1003561, GSM733650), H3K9ac (GSM1003544, GSM733709), H3K27ac (GSM1003493, GSM1003578, GSM733662), H3K36me3 (GSM1003494, GSM1003456, GSM733733), H4K20me1 (GSM1003458, GSM1003486), H3K9me3 (GSM1003454, GSM1003553), H3K27me3 (GSM1003577, GSM1003455, GSM733745), H3K79me2 (GSM1003543, GSM1003512, GSM1003554), and H2A.Z (GSM1003580, GSM1003546,GSM1003505). For the CTCF binding site analysis, CTCF binding site peaks for A549 (GSM1003581, GSM1003582) and NHDF-A (GSM733744) were used. H3K4me1/2/3, H3K9ac, H3K27ac and H4K20me1 are markers for euchromatin, whereas H3K9me3, H3K27me3

and H3K79me2 are considered to be markers for heterochromatin. Genomic regions containing both eu- and heterochromatic histone modifications were termed border regions. Histone peaks were correlated with A549 and NHDF-A DSB sites. A histone modification was considered to be DSB associated when it was found within 2kb of a DSB site. To obtain a DSB enrichment value for every histone modification the average number of histone modifications per DSB site was calculated and the histone frequencies compared to the random DSB data set.

2.2.3.5   Ingenuity Pathway Analysis (IPA)

In order to detect and analyze the enrichment of genes in a specific cohort, network analyses are frequently used. Here, to analyze whether genes in which induced and repaired DSB sites were identified, can be assigned to specific functions or diseases, the program "Ingenuity Pathway Analysis" (IPA) was used. For the analysis, the "NM_Accession" number for each gene marked by IDLV was saved in a separate .txt-file and uploaded. The resulting networks "Function&Disease" and "Upstream Regulators" with a p-value smaller than 0.05 were analyzed. In order to calculate the enrichment of specific categories, the frequency of genes associated to the respective category was calculated and divided by the frequency for the random DSB data set.

2.2.3.6   Identification of DSB site clusters in the genome

To investigate, if DSB sites cluster in genomic regions, statistical methods previously described for modeling the non-random distribution of retroviral integration sites in the genome were applied [84]. A genomic region is termed a DSB hotspot when the number of IDLV-tagged DSB sites exceeds the number of DSB sites estimated by a uniform distribution of DSB in the genome. Statistically, two, three, four and five captured DSB located in a 30kb, 50kb, 100kb and 200kb window, are considered to be non-randomly distributed and defined as hotspots. The higher the number of captured DSB is in a maximum of 200 kb genomic window the higher is the intensity and significance of the hotspot.

3.   RESULTS

3.1   IDLV-mediated capturing of radiation-induced DSB sites *in vivo*

To detect and analyze radiation-induced DSB sites *in vivo*, a new method was used to stably mark DSB sites *in vivo* (Figure 11). Prior to irradiation, target cells are transduced with an integrase-deficient lentiviral vector, which becomes stably integrated into the genome by the cellular NHEJ-repair machinery at DSB sites. Thus, the lentiviral vector places a stable genetic tag at the DSB repair site. Cells that successfully repair the induced damage and survive radiotherapy are subsequently expanded in cell culture. In order to analyze the distribution of induced and repaired DSB, the DNA is isolated and subjected to LAM-PCR and deep sequencing. The resulting vector-genome amplicons are then bioinformatically processed to obtain the genomic landscape of DSB sites and genomic instabilities (**Figure** 11).
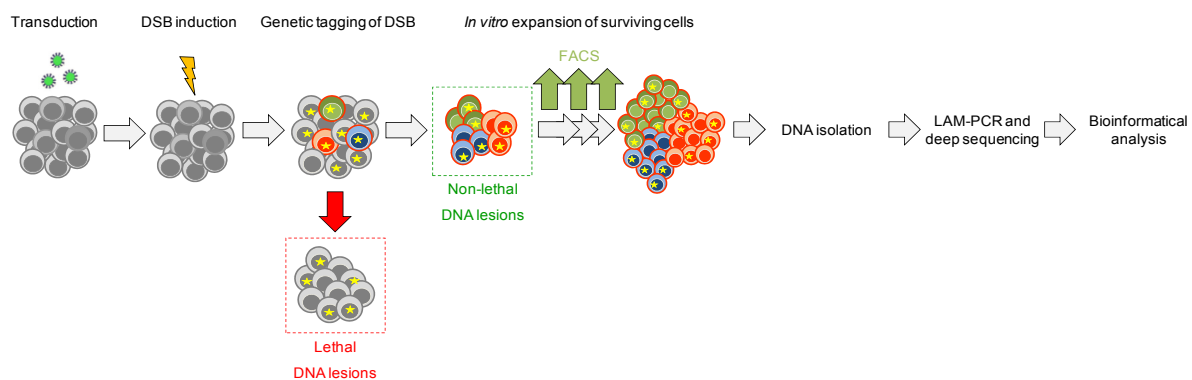
Figure 11: Experimental scheme for the induction, tagging and detection of radiation-induced DSB sites in surviving cell populations. Prior to DSB induction by radiation, the cells are transduced with IDLV. Lethal and non-lethal DNA lesions become stably marked by IDLV, and the incorporation of IDLV at DSB sites is followed by flow cytometry during expansion of cells. The DSB sites can be amplified and identified by LAM-PCR and deep sequencing, respectively.

3.1.1   Immunostaining of γH2AX foci in irradiated cells

One of the most frequently used methods for the detection and analysis of DSB sites in radiation-exposed cells is immunostaining of DNA repair proteins using fluorescently-labeled antibodies. These DNA repair proteins engage and disengage from the DSB site within a 24h period when damaged DNA is sensed and repaired. In order to determine the kinetics and intensity of DSB repair activity following irradiation with 1 and 4Gy photon beams (X-rays), immunostaining of γH2AX foci in A549 and NHDF-A was performed. In A549 cells, a single 1Gy and 4Gy photon beam led to the formation of 16 and 37 γH2AX foci per cell 1h post-irradiation, respectively. The level of phosphorylation subsequently decreased until reaching background intensity at about 8h after irradiation (Figure 12**,** Supplementary Figure 1). In NHDF-A, the same kinetics of phosphorylation and dephosphorylation were observed. However, only 13 and 32 γH2AX foci formed 1h after irradiation with 1Gy and 4Gy photon beams (Figure 12**,** Supplementary Figure 1). Hence, DSB site analysis by immunfluorescence is restricted to the first 8h after irradiation and does not provide any information about cell fate decision.
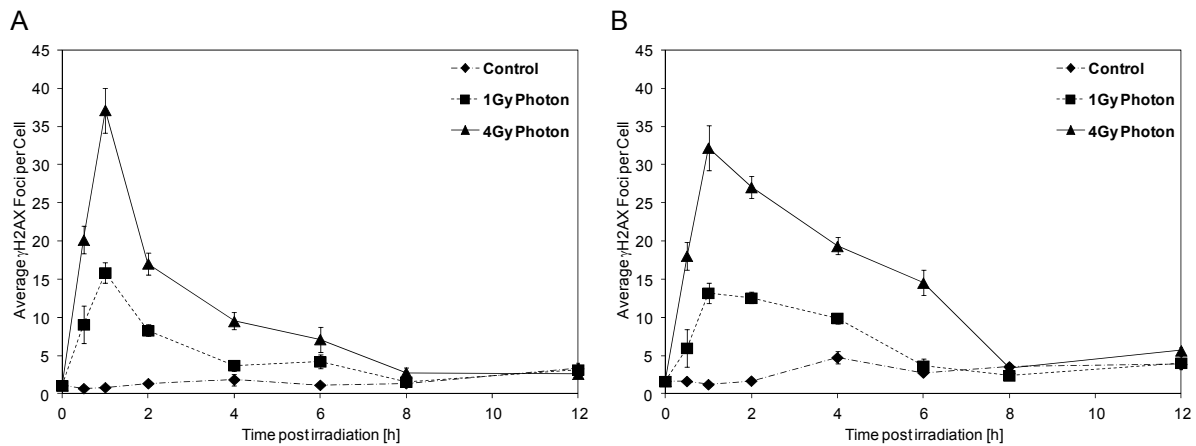
Figure 12: Kinetics of DSB induction and repair by immunostaining of γH2AX foci in (A) A549 and (B) NHDF-A after irradiation with 1 and 4Gy X-rays. The maximum number of γH2AX foci was reached after 1h and terminated at 8h post irradiation. Control: non-irradiated cells.

3.1.2   MTT assay to determine lethal dose values for etoposide and doxorubicin

As a control for radiation-induced DSB sites, etoposide- and doxorubicin-induced topoisomerase 2 (TOP2) DSB sites were determined, too. First, TOP2 inhibitor concentrations for efficient DSB induction and cell survival were determined by using the MTT assay. The lethal dose 50 (LD50) values of etoposide and doxorubicin (i.e. the dose at which 50% of the cells die) for Hela and NHDF-A cells was resolved by incubating the cells with various concentrations of doxorubicin and etoposide, ranging from 5 - 700 nM (doxorubicin) and 0.5 - 150 μM (etoposide) for seven subsequent days. Subsequently, the MTT assay was performed, in which active mitochondria in living cells reduce the tetrazolium dye MTT to the insoluble, purple-colored formazan. The absorbance at a wavelength of 610nm is proportional to the mitochondrial activity and thus to the number of viable cells (Figure 13, Table 2).
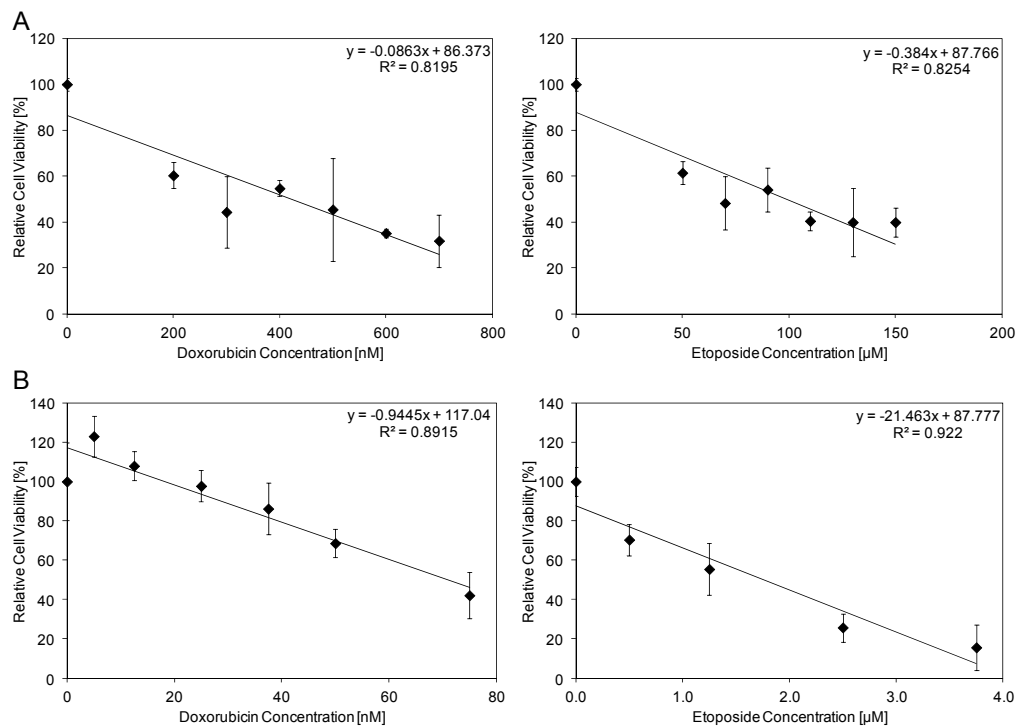
Figure 13: Relative cell viability of NHDF-A (A) and Hela (B) after seven day treatment with etoposide and doxorubicin (MTT Assay). Absorbance was measured at a wavelength of 610nm in a microplate reader and the cell viability was normalized to untreated control cells. Triplicate measurements for each concentration were performed. From the linear regression, the LD50 values were calculated. Error bars represent the standard deviation of the mean cell viability.

Table 2: Calculated LD50 values for doxorubicin and etoposide in NHDF-A and Hela.

Cell type	LD50 Doxorubicin [nM]	LD50 Etoposide [µM]
NHDF-A	421.5	98.4
Hela	70.8	1.8

3.1.3   IDLV-Delivered DNA-baits tag radiation-induced and repaired DSB sites

In order to stably trap radiation-induced and repaired DSB sites, three cancer cell lines and primary human fibroblasts (NHDF-A) were transduced with IDLV prior to irradiation with a single dose of photon, proton or carbon beams. Following irradiation, the cells were expanded in cell culture and the frequency of incorporated proviral DNA was measured by EGFP expression from the IDLV vector during passaging. After 25-30 days which represents six passages in cell culture, stable EGFP expression was observed (Figure 14, Supplementary Figure 2). With the serial passaging of the cells, episomal IDLV vector copies were gradually lost, and the EGFP expression was stabilized by the incorporation of the IDLV vector at radiation-induced DSB sites in the genome. Irradiation of the human alveolar adenocarcinoma cell line A549 with increasing doses of a photon, proton or carbon beam led to a dose-dependent increase in the IDLV integration frequency (Figure 14). Moreover, differences in the DSB capture frequency for different radiation sources were observed (Figure 14). In order to investigate whether IDLV trapping is also suitable to capture DSB in different cell types, five different cancer cell lines were transduced with IDLV (MOI 200) and subsequently irradiated with a single beam of carbons, protons or photons. A549 cancer cells showed the highest IDLV integration frequency of all five cell lines with an increase up to 18-fold compared to naturally-occurring DSB after irradiation with 4Gy photon, 4Gy proton or 1Gy carbon beams. In

contrast, the A431 epidermal carcinoma cell lines showed the lowest IDLV integration frequency (Figure 14). Hence, IDLV-mediated DSB trapping enables stable tagging of radiation-induced and repaired DSB sites in surviving cell populations over time.
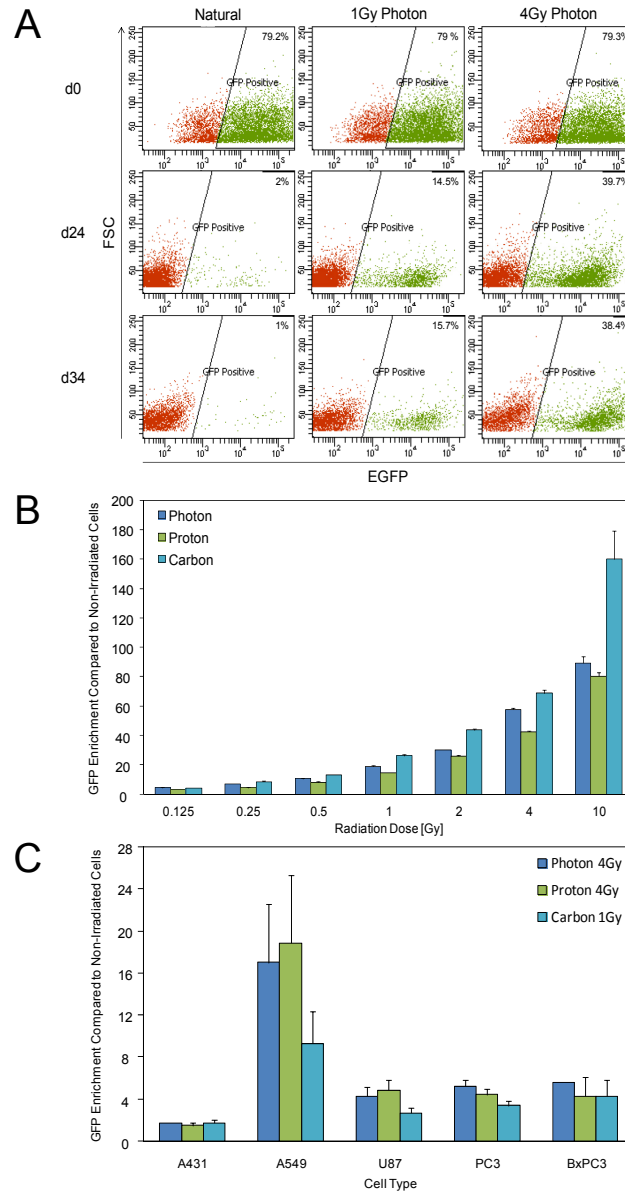
Figure 14: IDLV-trapping of radiation-induced DSB sites in cancer cell lines and human primary fibroblasts. (A) Gradual loss of IDLV-based DNA baits during *in vitro* expansion of NHDF-A. The IDLV integration frequency and loss of episomal vector copies was measured by EGFP expression from viral DNA baits encoding EGFP under the control of the PGK promoter. EGFP expression decreased following irradiation (d0) until reaching a stable percentage in expanded cells (d18 in A549 / d34 in NHDF-A). The frequency of IDLV incorporation correlates with an increase in DSB induction and NHEJ-repair. The numbers indicate the frequency of EGFP[+] cells. (B) The frequency of DSB trapping is measured by normalizing the level of stable EGFP expression compared to untreated cells. Increasing doses of photon, proton or carbon beams lead to a radiation source- and dose-dependent increase in the IDLV integration frequency up to 180- fold higher than naturally-occurring DSB frequencies in A549. (C) IDLV efficiently labels radiation-induced DSB sites in various cell types. The difference in DSB trapping efficiencies indicates that NHEJ-repair activity various among different cell types. FSC: Forward scatterer, EGFP: Enhanced green fluorescent protein

3.1.4   IDLV DSB trapping in NHDF-A with impaired NHEJ-repair activity

In order to analyze, if IDLV becomes incorporated at radiation-induced DSB by NHEJ-repair activity [85, 86], classical and alternative NHEJ-repair activity were blocked in NHDF-A by Nu7441 (LD5) and mirin (LD5) prior to irradiation, respectively (Figure 15). During expansion of irradiated cells, the frequency of EGFP[+] cells with impaired NHEJ-repair was significantly increased compared to NHDF-A with active NHEJ-repair (Figure 15). A significant increase in EGFP[+] cells was observable from passage 3 after irradiation, suggesting that back-up DNA repair mechanisms (over-) compensate for the NHEJ-repair defect in NHDF-A. In non-irradiated control cells with impaired NHEJ-repair activity no increase in the frequency of EGFP[+] cells compared to NHEJ-proficient cells was observed (Supplementary Figure 3), indicating that impairment of NHEJ-repair alone does not lead to increased DNA damage levels and more IDLV integration events.
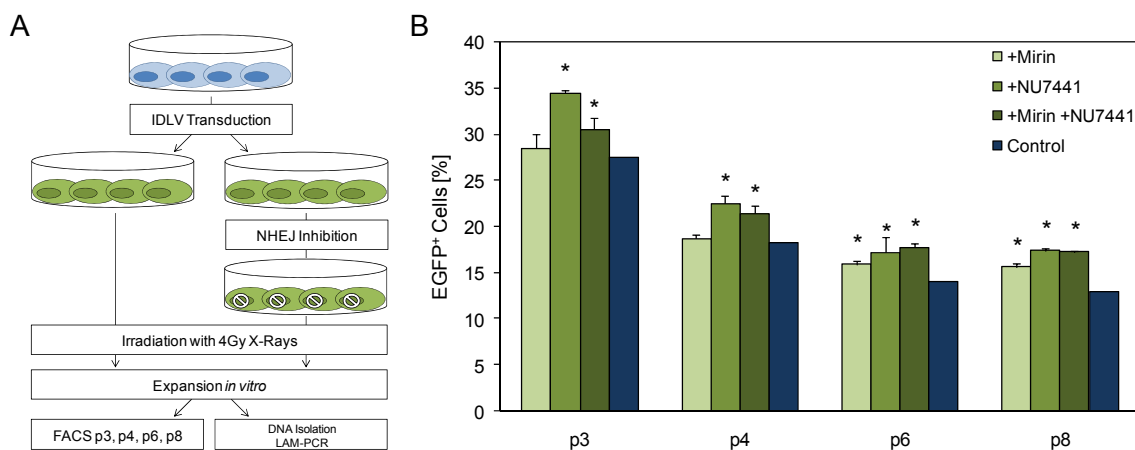
Figure 15: Chemical inhibition of NHEJ-repair activity increases integration of IDLV in irradiated cells. (A) Experimental scheme: NHDF-A were transduced with IDLV 48h before irradiation. NHEJ-repair activity was impaired by Mirin, Nu74441 or both together 3h prior to irradiation. NHDF-A were expanded for up to eight passages in culture. Integration of IDLV at radiation-induced and repaired DSB sites was monitored by measuring frequency of EGFP[+] cells by flow cytometry. (B) Starting from passage 3, the frequency of irradiated EGFP[+] cells with impaired NHEJ-repair activity was significantly increased compared to irradiated NHDF-A with active NHEJ-repair (Control). Error bars represent the standard deviation of the average frequency of EGFP[+] cells. Asterix: p-value $< 5 \times 10^{-4}$ (Fisher's exact test).

3.1.5   Trapping and mapping of TOP2 poison-induced DSB

As a control for trapping of radiation-induced DSB, capturing of TOP2 activity sites at which the chemotherapeutic reagents etoposide and doxorubicin block religation of the DNA after DSB induction was performed and compared to radiation-induced DSB sites. Therefore, Hela cells were transduced with IDLV and subsequently treated with doxorubicin (LD25 and LD50) for seven days. The experiment was cancelled after 14 days, because more than 90% of the Hela cells died or entered senescence following doxorubicin treatment (data not shown). The frequency of EGFP[+] cells was not significantly increased compared to untreated Hela cells at day 21 (Supplementary Figure 4). Thus, lower concentrations of doxorubicin and etoposide (LD10, LD5) were used for DSB induction and IDLV trapping. To ensure efficient DSB induction at LD10 and LD5, immunostaining of γH2AX in NHDF-A was also performed (Supplementary Figure 5). Up to 15 and 16 foci per cell were detected in cells treated with LD5 and LD10 doxorubicin, respectively, compared to four foci per cell in untreated control cells. Hence, doxorubicin induces efficient levels of DNA lesions for IDLV-trapping. In contrast to doxorubicin, the

number of γH2AX foci induced by etoposide treatment was not significantly increased compared to untreated control cells. Thus, IDLV-trapping of TOP2 activity sites was performed using doxorubicin.

NHDF-A were transduced with IDLV and subsequently treated for seven days with doxorubicin at LD5 and LD10. To distinguish doxorubicin-induced DSB from naturally-occurring DSB, DSB sites were also captured in untreated NHDF-A. The integration of IDLV was monitored by flow cytometry at days 28, 36 and 42 after the beginning of doxorubicin treatment. The percentage of EGFP[+] NHDF-A cells was stable at these three time points (Figure 16). Concentration-dependent differences in the IDLV-trapping frequency were also observed. In average, 35.9% and 51.1% of cells treated with LD5 and LD10 doxorubicin were EGFP[+], respectively, compared to 1.5% of the untreated control cells. This increase was statistically significant (p-value < 5x10^{-3}, Fisher's exact test), showing that IDLV-delivered DNA baits are efficient in capturing doxorubicin-induced DSB sites.
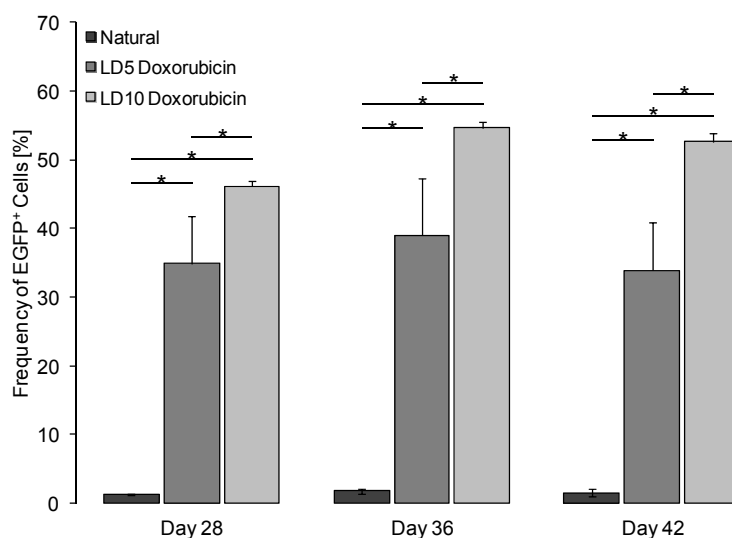
Figure 16: Frequency of IDLV-transduced EGFP[+] NHDF-A treated with LD5 and LD10 doxorubicin. Cells treated with LD10 doxorubicin showed the highest IDLV integration frequency. The enrichment of EGFP[+] doxorubicin-treated compared to untreated cells as well as LD5-treated compared to LD10-treated cells was statistically significant. Error bars represent the standard deviation of the average frequency of EGFP[+] cells. Asterix: p-value < 5x10^{-3} (Fisher's exact test)

3.1.6   Calculating the number of integrated IDLV copies per irradiated cell

In order to quantify the number of DSB sites marked by IDLV in radiation-surviving cells, an absolute quantitative real-time PCR (q-RT-PCR) with primers binding specifically to the *EGFP* locus in the IDLV vector as well as the genomic *MYO* gene locus were performed (Table 3).

Table 3: Quantification of IDLV integration events. Conc.: Concentration; SD: Standard deviation of the mean

Cell type	Radiation Source	Radiation Dose	Mean Conc. MYO	SD MYO	Mean Conc. EGFP	SD EGFP	EGFP/MYO ratio
A549	Natural	0Gy	68.20	3.75	1.13	0.13	0.017
	Photon	4Gy	73.33	4.57	18.93	0.69	0.258
	Proton	4Gy	80.7	3.09	20.27	0.42	0.251
	Carbon	2Gy	67.7	3.26	14.80	1.69	0.219

Since every A549 cancer cell carries two copies of the *MYO* locus per cell, the number of *EGFP* copies per cell can be calculated by simply multiplying the *EGFP/MYO* ratio by 2. In average, 0.034 EGFP copies per cell were identified in non-irradiated A549 control cells. The number of EGFP copies was increased about 15-fold to 0.516 and 0.502 copies per cell in photon- and proton-irradiated samples, respectively. A single 2Gy carbon ion beam led to the incorporation of 0.438 EGFP copies per cell, showing that 2Gy carbon irradiation induces similar ionization and DNA damage levels compared to 4Gy photon and proton beams.

3.1.7   SureSelect Target Enrichment for analyzing IDLV vector integrity

Irradiation does not only induce damage to genomic DNA, but also to the lentiviral episomal DNA. Similar to damaged genomic DNA, the broken fragments of the lentiviral vector may be stably integrated at DSB sites in a non-LTR-dependent manner. Since q-RT-PCR only covers a small portion of the lentiviral vector, the results from q-RT-PCR may be an underrepresentation of the total number of marked DSB sites. Moreover, q-RT-PCR does not provide any information about the vector integrity after integration. Thus, in order to analyze the IDLV integrity in irradiated cells, the SureSelect Target Enrichment protocol for Illumina Sequencing was applied to genomic DNA isolated from IDLV-transduced and irradiated NHDF-A. This method enriches the complete lentiviral vector sequences and enables the identification of IDLV vector-genome junctions which represent vector breakage sites by pyrosequencing (Table 4).

Table 4: Overview about results of HiSeq sequencing of SureSelect DNA samples. V-V: vector-vector; V-G: vector-genome

Radiation Dose	No. of reads	No. of filtered trimmed reads	Vector aligned reads	Correctly vector aligned reads	No. of V-G junctions	Sequence Count (V-G)
0Gy	1,662,774	1,620,487	12,675	10,872	13	203
1Gy	1,990,260	1,981,804	70,102	58,280	99	340
4Gy	1,284,307	1,270,338	64,174	61,228	235	590

About 1% of the filtered and trimmed reads aligned to the vector. Of these sequences, the majority (>95%) were correctly aligned vector-vector reads, indicating that IDLV is stable upon integration at radiation-induced DSB. However, 345 vector-genome junctions in irradiated and non-irradiated NHDF-A were also identified (Table 4). In non-irradiated cells, the majority of vector-genome junctions were located in the LTR regions and one junction at the packaging signal psi, the RRE as well as the cPPT inside the lentiviral vector (Figure 17). In contrast, vector-genome junctions in irradiated NHDF-A were evenly distributed throughout the vector, showing that the continuity of IDLV was disrupted and that the resulting fragments were integrated into the genome in a non-LTR-mediated manner (Figure 17).
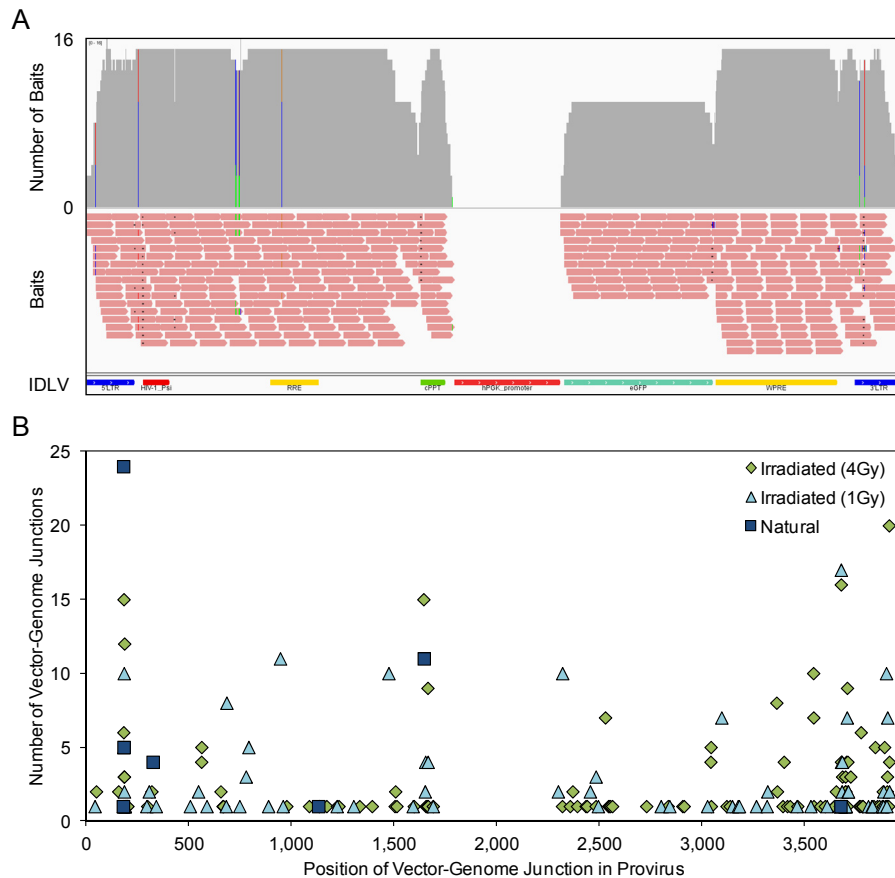
Figure 17: Distribution of vector-genome junctions in IDLV vector (A) Design and distribution of IDLV-vector baits.No baits were designed for the human PGK promoter sequence in the capture library. (B) Distribution of vector-genome junctions in NHDF-A. Vector-genome junctions in non-irradiated control cells were identified mainly at the LTR. In contrast, vector-genome junctions in irradiated NHDF-A occurred in almost every site inside the vector, indicating that the vector continuity is disrupted upon irradiation. Natural: Non-irradiated cells

However, a more detailed analysis of IDLV integrity and identification of additional DSB sites from this data set could not be performed at this time, since the total number of vector-genome junctions in irradiated and control cells was too low. The experimental and bioinformatical analysis pipeline for the SureSelect platform is currently under development in our group. Thus, more vector-genome junctions in both non-irradiated and irradiated samples may be identified in the future.

3.2   Analyzing early DSB repair events and kinetics of DSB induction at single nucleotide resolution

As shown in the previous chapters, IDLV-based trapping of radiation-induced and repaired DSB sites is suitable to identify and analyze the genomic landscape of DSB and genomic instabilities in radiation-surviving cells. Thus, IDLV-mediated DSB trapping might also be useful to identify radiation-induced DSB sites within the first few hours after DSB induction by IR and follow the DSB site distribution over time. In addition, new approaches to label DNA lesions in real time were developed to study DNA repair kinetics. The development of these methods is described in chapters 3.2.3 and 3.2.4. In order to study DSB induction and repair kinetics in different cell cycle phases, protocols to synchronize cells in G1, S and in G2 phase were established and are also described in the following sections.

3.2.1   Genomic tagging of early-repaired radiation-induced DSB sites by IDLV

To analyze early DSB repair events, i.e. within 24h after irradiation, and to trace DSB sites over time, NHDF-A were transduced with IDLV 48h prior to irradiation. Following irradiation with 4Gy X-rays (photon), the DNA from irradiated and non-irradiated control cells was isolated at 1h, 4h, 8h, and 24h post irradiation. LAM-PCR and deep sequencing identified a total of 1731 unique DSB sites in irradiated and control cells (Table 5).

Table 5: Number of DSB sites identified at 1h, 4h, 8h and 24h after irradiation

Radiation Source	1h	4h	8h	24h
Photon	137	48	71	101
Natural	1183	49	71	71

The number of unique DSB identified in irradiated NHDF-A was similar to the number of DSB in non-irradiated control cells at all four time points after irradiation. A possible reason for this observation is that the majority of IDLV copies in the cell were not yet integrated into the genome, but present as episomes. These vector molecules also become amplified during LAM-PCR and are thus overrepresented in the sequences. Indeed, only 30% of the sequences that carried the correct sequencing megaprimer sequence and LTR sequence parts also mapped to genomic locations. The remaining 70% were not related to genomic sequences, but to lentiviral vector parts. Moreover, only 1% of the sequences mapping to the genome were IDLV integration sites, indicating that the majority of the genomic sequences map to the same locus. Hence, analyzing the DSB site distribution and repair kinetics in NHDF-A within the first 24h following irradiation was not possible.

3.2.2   Synchronization of NHDF-A and Hela cells in G1, S and G2 phase

In order to study the influence of cell cycle phases on DSB induction and repair kinetics, cells need to be synchronized efficiently. Therefore, protocols to block cell cycle progression in G1 and S phase as well as at the G2/M checkpoint were established. The efficiency of each synchronization method was analyzed by staining the nuclear DNA with propidium iodide (PI) foolowed by flow cytometry. For synchronizing cells in G1 phase, serum starvation was performed. Therefore, NHDF-A were incubated with DMEM containing 1% FCS instead of 10% for 72h (Figure 18). Up to 93% of the NHDF-A were arrested in G1/G0 phase. Synchronizing NHDF-A in S phase was achieved by using a double thymidine block. Cells were cultivated in DMEM containing 10% FCS and 2mM thymidine for 48h. Subsequently, the block was released by incubating cells in DMEM (10% FCS) for 16h before cell cycle was blocked for a second time with DMEM containing 10% FCS and 2mM thymidine for another 72h (Figure 18). Double thymidine treatment enables the synchronization of about 74% of all cells in S phase. Nocodazole is a chemical substance frequently used for blocking entry into mitosis. Therefore, cells were arrested in G2 phase by adding Nocodazole at a concentration of 100ng/ml to the cell culture medium for 72h (Figure 18). At these conditions, 46% of all cells were arrested in G2/M phase. Since nocodazole is an anti-neoplastic agent that interferes with microtubule formation, the frequency of dead cells increased up to 6%.
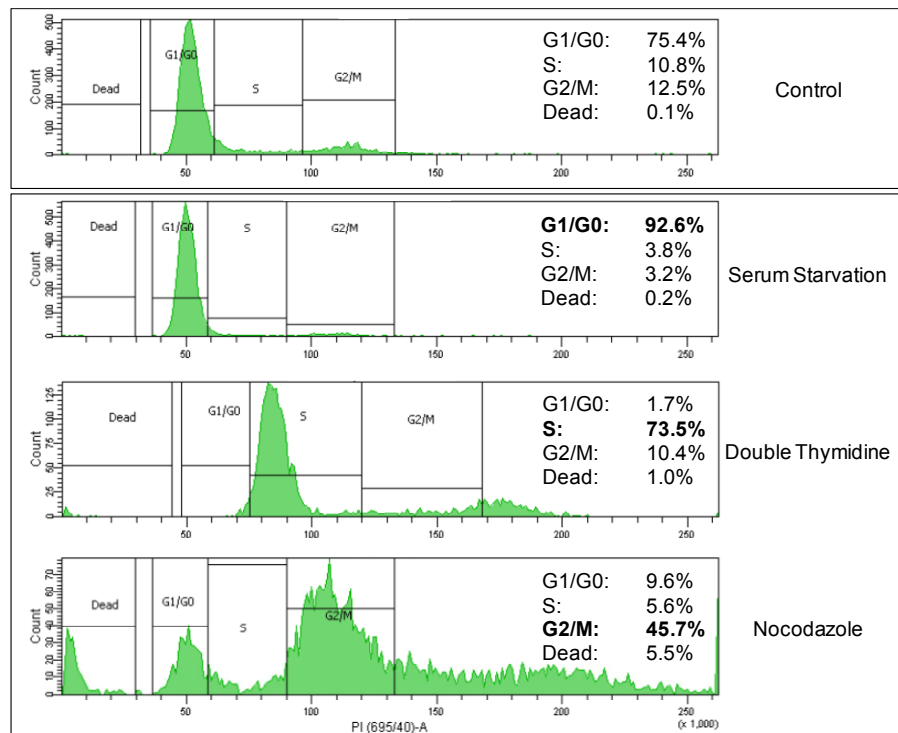
Figure 18: Synchronization of NHDF-A in G1/G0, S and G2/M phase of the cell cycle. The method used for synchronization and the distribution of the cells in the different cell cycle phase is indicated at the right side. The PI fluorescence values on the x-axis indicate the DNA content. Control: Non-synchronized cells.

3.2.3   Tdt-mediated labeling of radiation-induced DSB sites

Since IDLV-labeling of DSB sites does not enable the study of early DSB repair events and repair kinetics. Since there is a lack of methods to map DSB with high specificity and resolution genome-wide, a new method to identify rapidly-repaired DSB sites is required. Thus, an *in situ* DSB labeling approach was developed and established. In principal, irradiated cells can be fixed using 4% paraformaldehyde and permeabilized with 0.25% Triton X-100. Subsequently, the free DNA ends produced by irradiation are labelled by the activity of the terminal deoxynucleotidyl transferase (Tdt), which catalyses the addition of the nucleotide dUTP to the 3' terminus of DNA molecules. In addition to dUTP, modified dUTP covalently linked to biotin (Biotin-16-dUTP) are also included in the labeling reaction to enable capturing of the DSB sites by streptavidin-coated paramagnetic beads. The captured DNA can subsequently be digested by a restriction enzyme and short DNA linkers of known sequence ligated to the DNA overhangs. The DNA fragments are finally exponentially amplified by PCR (Figure 19).
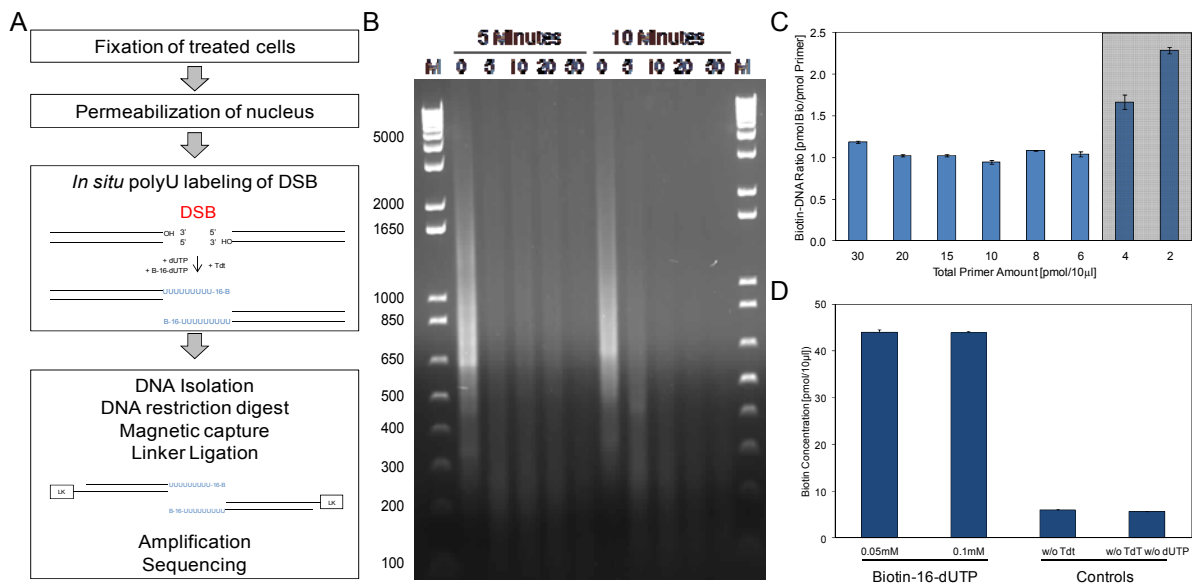
Figure 19: Tdt-mediated polyU labeling of radiation-induced DSB sites. (A) Workflow of Tdt labeling and amplification. (B) MNaseI digestion of genomic DNA *in situ* shows that fixation of cells with 4% PFA and permeabilization with 0.25% Triton X-100 enables enzymes to enter the cell nucleus. Numbers at DNA ladder indicate the DNA fragment size. The numbers above every lane represent the units MNaseI inserted and the incubation time. (C) Determining the biotin detection limit using the Biotin Quantificatioin Kit from ThermoScientific. The lowest amount of biotin detectable in 10µl reaction volume is 6pmol. Grey area indicates concentrations below detection limit. (D) Efficient *in vitro* biotin-dUTP labeling of genomic DNA by Tdt with two different concentrations of biotin-16-dUTP. Tdt: Terminal deoxynucleotidyl transferease; w/o: without; DSB: DNA double strand break; M: 1.5kb plus DNA ladder

First, standard reaction conditions were established and evaluated. Thus, different protocols and solutions were tested for fixation and permeabilization. Non-treated A549 cancer cells were fixed with 4% paraformaldehyde (PFA) and permeabilized with 0.25% Triton X-100. Subsequently, different units of MNase were added for 5 and 10 minutes incubation (Figure 19). The endonuclease MNase recognizes and cleaves double stranded DNA located in between the nucleosomes [87], thereby reducing the median DNA size to several hundred nucleotides in length. MNase digestion of genomic DNA in fixed cells led to smaller DNA fragments, showing that the reaction conditions chosen for fixation and permeabilization are efficient to enable enzymes to enter and become active in the cell nucleus (Figure 19).

Next, to evaluate the activity of Tdt, the DNA biotinylation level was measured. In order to determine the detection limit of biotinylated DNA, the biotinylated DNA primer SK-LTR 5bio was diluted (2pmol -15pmol per 10µl reaction). Since every DNA primer molecule carries a single biotin molecule, the ratio of biotin-to-DNA is 1. Hence, the biotin concentration determined is equal to the inserted primer concentration. The biotin detection limit of the biotin quantification assay was the lowest concentration at which the biotin-to-DNA ratio was still 1. As shown in figure 32C, the biotin detection limit per DNA molecule was determined to be 6pmol in 10µl reaction volume. Morover, the reaction conditions for the Tdt polymerase were optimized. Therefore, two different biotin-16-dUTP concentrations (0.05mM and 0.1mM) were used for *in vitro* Tdt-mediated labeling of genomic DNA. Compared to non-labeled DNA, a 7.6-fold increase in biotinylation levels was observed (Figure 19). The level of biotinylation in the two samples with 0.05mM and 0.1mM biotin-16-dUTP was similar, showing that the chosen concentrations are suitable to label free DNA ends. Furthermore, *in situ* DSB labeling was performed and the level of biotinylation determined. Therefore, 4×10^6 293T cells were irradiated with 20Gy X-rays, fixed and permeabilized. Subsequently, 1×10^6 cells were incubated with the Tdt mastermix for 1h at 37°C to label the free DNA ends. Following DNA isolation, the biotinylation level of the DNA was determined. No increase

in biotinylation levels compared to the control reaction was observed. By reducing the PFA concentration, introducing more stringent washing steps following permeabilization, incubation with Tdt reaction buffer prior ro DSB labelling and prolonged labelling time did not increase the biotinylation to detectable levels (data not shown), showing that the Tdt activity *in situ* cannot be measured.

Still, it could be posssible that Tdt activity was efficient to label DSB, but that the biotinylation levels were below the detection limit. Thus, in order to demonstrate that induced DSB sites can be labelled *in situ* by Tdt, site-specific DSB sites were induced in the genome of 293T cells and subsequently labelled. Therefore, $4x10^6$ 293T cells were transfected with a ZFN targeting the CCR5 locus [85] and incubated for 24h and 72h. Subsequently, cells were fixed, permeabilized, and the DSB sites labeled by Tdt. Following DNA isolation, LAM-PCR was performed, resulting in the amplification of DNA fragments in CCR5 ZFN-treated samples as well as untreated control samples (Figure 20). In order to analyze the specificity of the Tdt-labeling reaction and amplification, the PCR amplicons were purified and subjected to deep sequencing.
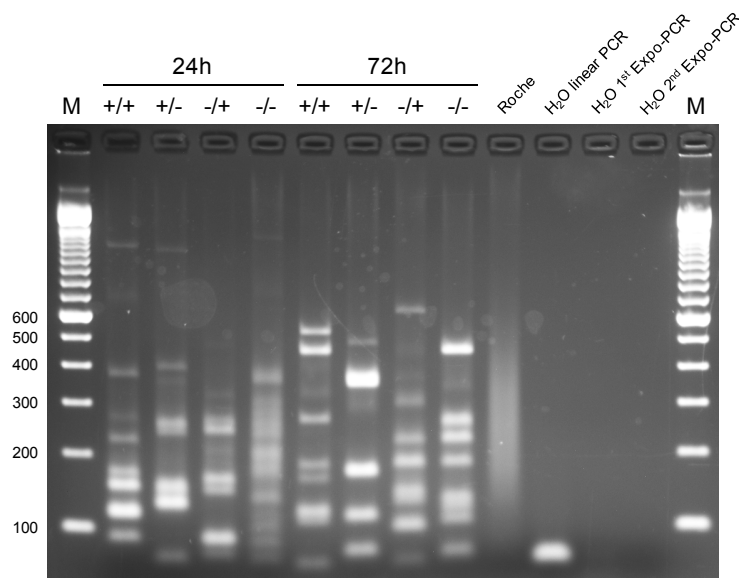
Figure 20: Gel picture of LAM-PCR amplicons. Numbers at the side of DNA ladder and on top of the gel picture indicate the amplicon size and the time of incubation after transfection, respectively. +/+: +CCR5 ZFN / +Tdt labeling; +/-: +CCR5 ZFN / without Tdt labeling; -/+: without CCR5 ZFN / +Tdt labeling; -/-: without CCR5 ZFN / without Tdt labeling; Roche: negative control (human genomic DNA); H_2O: water control; M: 100bp DNA ladder

Table 6: Summary of sequencing results from LAM-PCR of CCR5-ZFN treated and Tdt-labeled DSB samples. ZFN: Zinkfinger Nuclease; Seq: sequence; w/o: without; Tdt: Terminal deoxynucleotidyl transferase, #: Number

Sample	Tdt Labeling	#Raw Seq	#Megaprimer found	#DSB	#DSB On-target	#DSB Off-target
CCR5 ZFN	+Tdt	8531	8475	48	0	0
CCR5 ZFN	w/o Tdt	5562	5516	40	0	0
Untreated	+Tdt	1299	1283	29	0	0
Untreated	w/o Tdt	4608	4590	53	0	0

The CCR5 locus-specific ZFN is known to bind and induce a DNA double-strand break at position 46,414,558 on chromosome 3. LAM-PCR-mediated Tdt-labeling site analysis of the CCR5 ZFN-treated samples showed that no Tdt-label was identified at the intended on-target and off-target loci [85]. All of the sequences identified were

unrelated genomic sites, revealing a high level of unspecific Tdt-labeling and/or DNA amplification during LAM-PCR. Sequence analysis of the amplified genomic sequences revealed that more than half of the sequence carried a thymidine nucleotide at the first two positions after the junction. An increase in the frequency of the two nucleotides adenine and thymidine and a depletion of cytidine as compared to the frequency of these nucleotides in the human genome (A=T=0.29; G=C=0.21) was observed (Supplementary Figure 17). These sequencing results suggest that unspecific DNA sequences were amplified during LAM-PCR. Hence, *in situ* DSB labeling by Tdt polymerase does not seem to be efficient for the identification of induced DSB sites.

3.2.4   Linker-Amplification-Mediated DSB-Trapping for DSB labeling in real-time

Since Tdt-mediated DSB labeling did not efficiently identify induced DSB sites *in situ* and suffered from high amplification rates of unspecific DNA sequences, a new method based on direct ligation of DNA adaptors *in situ* was developed and termed Linker-Amplification-Mediated DSB-Trapping (LAM-DST). To implement LAM-DST, 293T HEK cells were transfected with the Zinkfinger nuclease targeting the CCR5 locus [85]. Following a 48h incubation, the cells were fixed with paraformaldehyde (PFA) to stabilize chromatin and prevent artificial DSB formation and subsequently permeablized with Triton X-100 on ice. Next, the fixed and permeabilzed cells were pre-incubated with 1x CircLigase ssDNA Ligase reaction buffer or 1xT4 DNA Ligase buffer for 15min to allow the buffer to enter the nucleus. The ZFN-induced DSBs were directly ligated to DNA linkers that are covalently linked to biotin at 60°C for 2h (CircLigase) or 12°C for 16h (T4 DNA ligase) (Figure 21). The DNA linkers were designed to be compatible with (nr)LAM-PCR (Supplementary Figure 19). Additional modifications were included in the DNA linker. The 5' terminus was phosphorylated to enable linker ligation at free 3'OH DNA termini. The 3' DNA terminus of the DNA linker carries a dideoxycytidine, a pyrimidine analogue, which prevents linker-linker ligations at the 3' DNA termini. Furthermore, the phosphate backbone of the last three nucleotides at the 3' DNA terminus were replaced with a thiophosphoester. Following DNA linker ligation, the genomic DNA was isolated. To show that LAM-DST labels CCR5 ZFN-induced DSB, a CCR5 locus-specific PCR with linker- and locus-specific PCR primers was performed. A PCR fragment of expected size (210bp) was amplified (Figure 21). Cloning of the 210bp PCR fragment into the pCR2.1-TOPO TA vector and Sanger sequencing verified the amplification of the linker-genome junction at the CCR5 locus.
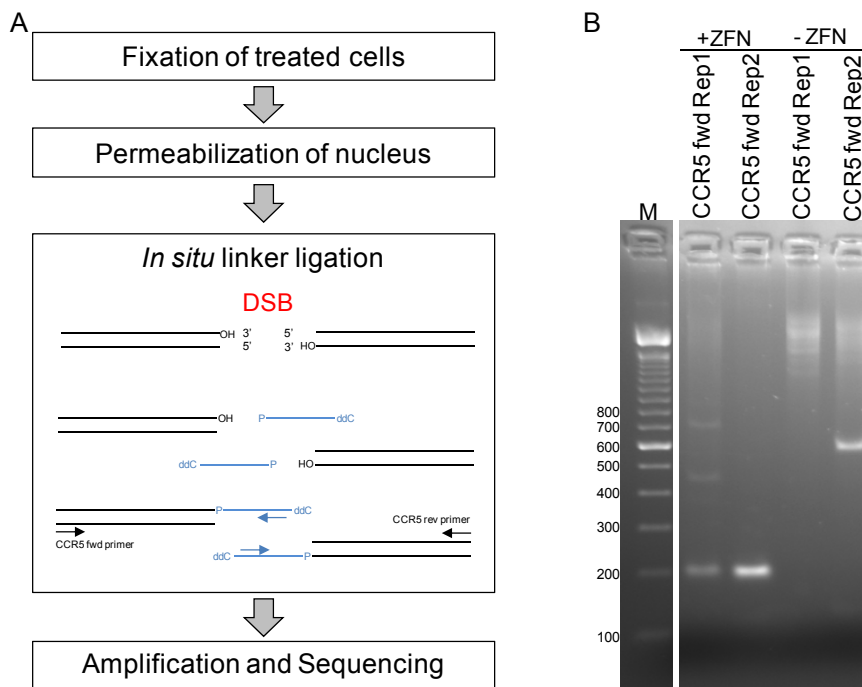
Figure 21: LAM-DST for *in situ* labelling of DSB sites. (A) Workflow and principal of *in situ* ligation by LAM-DST. (B) CCR5-locus specific PCR using the primers indicated in (A) shows that LAM-DST is suitable to label and identify CCR5 ZFN-induced DSB sites. Numbers at DNA ladder represent the DNA size. OH: hydroxyl group at 3' DNA terminus; P: phosphorylated end of DNA linker; ddC: dideoxycarbon at 3' DNA terminus; M: 100bp DNA ladder; fwd: forward primer; rev: reverse primer; Rep: replicate; ZFN: CCR5-locus specific Zinkfinger nuclease

In addition to CCR5 locus-specific PCR, LAM-PCR and deep sequencing were performed on ssDNA labelled samples to show that LAM-DST enables the genome-wide identification of induced DSB sites. The LAM-PCR was performed with the restriction enzymes MluCI and MseI and showed a polyclonal DNA band pattern (Figure 22). Subsequently, the DNA was subjected to pyrosequencing.
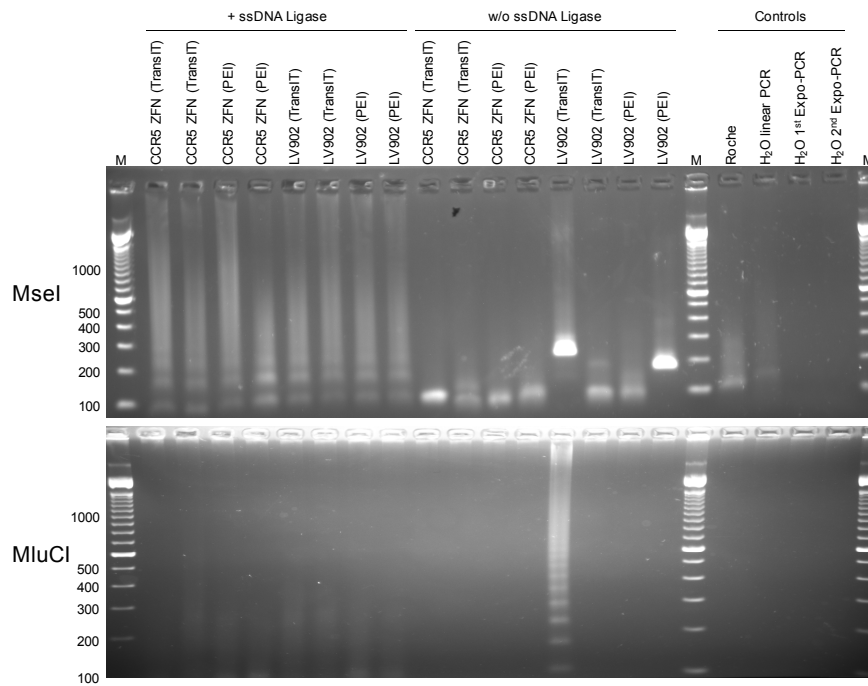
Figure 22: LAM-PCR on ssDNA linker-labelled DSB sites in CCR5 ZFN-treated 293T cells. Transfection of CCR5-specific nuclease was performed with either TransIT or PEI. As a control for the nuclease, the plasmid LV902 that does not encode any nuclease, was also transfected into 293T cells. LAM-PCR was performed with MseI and MluCI endonucleases. A polyclonal DNA pattern was observed for samples that have been treated with the ssDNA ligase. Unspecific amplification was observed in samples that were not labeled with the ssDNA ligase. ZFN: Zinkfinger nuclease; PEI: Polyethylene-inositol; M: 100bp DNA ladder, w/o: without; Roche: genomic DNA; H_2O: water control

In total, 203,319 raw sequences were obtained for LAM-PCR samples transfected with CCR5 ZFN and labeled by ssDNA ligase. Of these, 4,369 linker-genome junctions were identified. In CCR5 ZFN-treated samples that have not been labeled by ssDNA ligase 10,954 linker-genome junctions were obtained. LAM-DST-mediated analysis of the CCR5 ZFN-treated samples revealed that no ssDNA ligase site was identified at the intended on-target and off-target loci [85]. All sequences were unrelated to the CCR5 ZFN activity, reflecting a high level of unspecific DSB labeling and/or DNA amplification.

3.3   DSB site distribution in radiation-surviving and expandable cell populations

3.3.1   Identification of radiation-induced DSB sites by LAM-PCR

In order to determine the distribution of radiation-induced and repaired DSB sites in radiation-surviving cell populations, IDLV-DSB trapping was chosen. Following expansion of irradiated cells for more than four passages and subsequent genomic DNA isolation, triplicate LAM-PCR with the two restriction enzymes MluCI and MseI were performed. For each replicate 1µg DNA was inserted into the linear amplification step. The LAM-PCR amplicons were subsequently prepared for pyrosequencing. For the doxorubicin-treated cell populations, duplicate LAM-PCRs with the restriction enzymes MseI and MluCI using 500ng genomic DNA input were performed for samples at day 36 and 42. The LAM-PCR amplicons were subjected to MiSeq sequencing in the DKFZ core facility.

After sequencing, the linker- and vector-parts were removed bioinformatically from the sequences, and the remaining sequence was mapped to the genome using BLAT. From these sequences, collisions with other sequences and non-mappable DSB sites in repetitive genomic regions were removed. In total, more than 30,000 unique radiation-induced and repaired DSB sites in A549, U87, PC3 and NHDF-A were obtained. In doxorubicin-treated NHDF-A, a total of 7739 unique DSB sites and 267 naturally-occurring DSB in untreated control cells were obtained. In NHDF-A with impaired NHEJ-repair activity, a total DSB data set of 22,382 sites from three passages (p4, p6, p8) was acquired (Supplementary Table 1). In order to determine, if the DSB distribution in radiation-surviving cells reflects a random distribution, the unique DSB sites were compared to a set of 5,000 randomly-distributed, *in silico*-generated DSB sites.

3.3.2   DSB are not enriched on chromosomes, in genes and gene-regulatory regions

First, the chromosomal distribution of radiation-induced DSB was determined and compared to naturally-occurring and doxorubicin-induced DSB sites. The number of captured DSB per chromosome correlated with the chromosome size, and did not show any difference to the random DSB locations (Figure 23). This was also observed in the irradiated cancer cell lines A549, PC3 and U87, in irradiated NHDF-A with impaired NHEJ-repair activity and in doxorubicin-treated NHDF-A (Supplementary Figure 6).
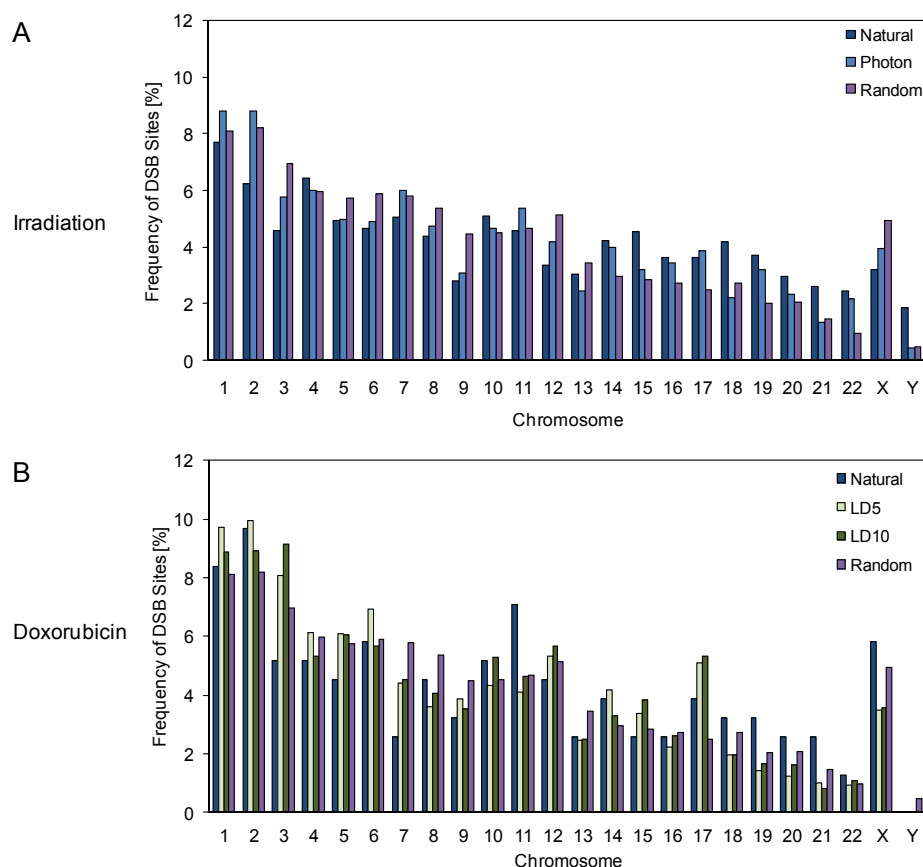
Figure 23: Chromosomal distribution of (A) radiation- and (B) doxorubicin-induced DSB in NHDF-A. The number DSB sites on each chromosome correlates with the chromosome size, and no enrichment of induced and repaired DSB sites on any chromosome was observed. Natural: DSB sites from untreated cells; Random: randomly-distributed, *in silico*-generated DSB sites; LD: lethal dose

Next, the distribution of radiation-induced and repaired DSB sites with respect to gene-coding regions was analyzed. In average, the frequency of radiation-induced and repaired DSB was 51.4% in NHDF-A compared to

47.5% and 44.6% for the cancer cell lines and randomly-distributed DSB, respectively. Moreover, 5.5% and 2.7% of the DSB sites in irradiated primary fibroblasts and cancer cell lines were located in transcriptional control regions located within 10kb upstream of the transcription start site (TSS), compared to 4.7% for the randomly distributed DSB (Figure 24, Supplementary Figure 7, Supplementary Table 2). In doxorubicin-treated NHDF-A, the DSB frequency at TSS was comparable to radiation-induced DSB in NHDF-A, cancer cell lines and the randomly distributed DSB. However, an increase of induced and repaired doxorubicin-related DSB sites in RefSeq genes was observed (56.3%) (Figure 24, Supplementary Table 2). Likewise, the frequency of DSB sites in genes and at the TSS in irradiated NHDF-A with impaired NHEJ-repair in all three passages was similar to the DSB frequency in cells with active NHEJ-repair and to the random DSB frequency (Supplementary Figure 7, Supplementary Table 2). Furthermore, the distribution of DSB sites in cancer cell lines with respect to genomic regions with a high content of CpG dinucleotides, termed CpG islands, was compared. These genomic regions are generally associated with gene regulatory elements such as promoters and enhancers. In irradiated cancer cell lines, the DSB frequency at CpG islands [≤1kb] was 3.2% compared to 2.6% for the randomly-distributed DSB locations (Supplementary Table 2) and decreased with increasing distance from the CpG island, reflecting a random distribution.
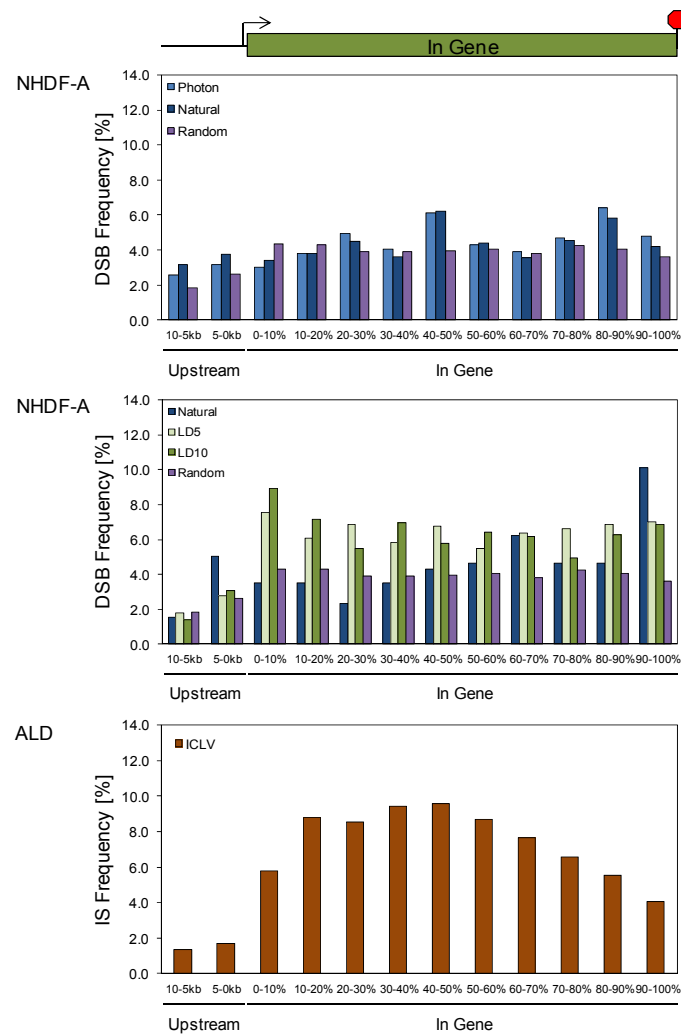
Figure 24: DSB frequency in RefSeq genes and at the transcription start site (TSS). Radiation- (top) and doxorubicin-induced (middle) as well as naturally-occurring DSB sites in NHDF-A were not enriched at TSS and in RefSeq genes. In contrast to IDLV, integrase-competent lentiviruses (ICLV) from an ALD clinical study showed preferred integration in gene-coding regions (bottom). LD5: Lethal Dose 5 Doxorubicin, LD10: Lethal Dose 10 Doxorubicin, Natural: Untreated control cells, Random: Random DSB data set, ICLV: Integrase-competent lentivirus; ALD: Adrenoleukodystrophy

In order to analyze whether the observed IDLV integration profiles are not mediated by residual integrase activity, the DSB distribution pattern was compared to an integrase-competent lentivirus (ICLV) data set retrieved from two patients from a X-linked adrenoleukodystrophy (ALD) clinical gene therapy [88]. The ICLV frequency in RefSeq genes was higher than expected by chance and higher than the IDLV frequency in irradiated or doxorubicin-treated cells (Figure 24, Supplementary Table 2).

Topoisomerase II poisons bind at specific sites at the interface of topoisomerase-DNA complex. Previous analysis of the DNA sequence at the doxorubicin-poisoned TOP2 cleavable complex has revealed that doxorubicin has sequence selectivity for inhibiting TOP2 activity, namely an adenine immediately upstream of the cleaved nucleotide [89]. Therefore the genomic sequence adjacent to the vector-genome junction was analyzed. Indeed, the frequency of the A nucleotide at the first genomic position at doxorubicin-induced DSB sites was significantly increased (35%) compared to naturally-occurring (28%) (p-value < 0.05, Fisher's exact test) and randomly-distributed genomic DSB locations (31%) (p-value < $3x10^{-6}$, Fisher's exact test) (Supplementary Figure 8).

3.3.3   IDLV integration at radiation-induced DSB sites is mediated by NHEJ-repair

To further analyze the DNA repair mechanism by which IDLV becomes integrated at transiently-induced DSB sites in the genome, the LTR deletion signatures of the integrated vector were investigated, since deletions are a typical sign of error-prone NHEJ-repair. In average, 31% and 53% of IDLV joined into the genome in untreated and irradiated NHDF-A, respectively, showed deletions of the LTR. In doxorubicin-treated NHDF-A, the frequency of LTR deletions was decreased compared to untreated control cells (26% vs. 39%), and deletions varied from a single base up to 40 bases in length. NHDF-A with impaired NHEJ-repair activity also showed signs of error-prone DNA repair activity. In all three passages, the average LTR deletion frequency in non-irradiated NHDF-A with impaired NHEJ-repair was similar to the frequency of LTR deletions in NHDF-A with active NHEJ-repair. The frequency of LTR deletions varied between the different NHEJ-repair inhibitors. For example, in p4, 64% of the LTR sequences in mirin-treated NHDF-A carried at least a single deletion, whereas 49% of the LTR in mirin- and Nu7441-treated NHDF-A were deleted. The frequency of LTR deletions in NHDF-A with active NHEJ-repair in the same passage was 55%, indicating that integration of IDLV is mediated by error-prone DNA repair activity. In contrast to IDLV, integrase-mediated lentiviral vector insertions showed LTR deletions only in 2% of all integration events (Supplementary Figure 9).

The 3' terminus of each strand of unprocessed lentiviral DNA ends in the four nucleotide sequence 'CAGT'. During infection, the integrase processes the 3' terminus and removes the terminal 'GT' dinucleotide. Since IDLV is integrase-deficient, the 'GT' dinucleotide should remain present at the vector-genome junctions. As expected, the dinucleotide was frequently found at vector-genome junctions in non-irradiated NHDF-A (61%) and irradiated NHDF-A (30%). Considering a random DSB site distribution, the 'GT' dinucleotide frequency would be 5% (Supplementary Table 3). These results and the integration profile of ICLV shown in chapter 3.3.2 indicate that the incorporation of IDLV at naturally-occurring and induced DSB sites is not mediated by integrase activity, but by the cellular NHEJ-repair activity.

3.4   Analyzing the influence of transcriptional activity, chromatin status and gene classes and networks on DSB site distribution

3.4.1   Trapping of Radiation-Induced DSB is not influenced by transcriptional activity

The transcriptional activity of the genome has been reported to influence DSB induction and the genomic accessibility of the DNA repair machinery [14]. Thus, in order to evaluate the influence of the transcriptional activity on DSB site distribution and genomic instability, A549 DSB sites were correlated with mRNA microarray probes located within 250kb up- and downstream of the DSB sites. Microarray probes with signal intensities above the threshold (2x background intensity) were considered to be active transcriptional mRNA signatures, whereas mRNA signals falling beneath the threshold were classified as silent transcriptional mRNA signatures.
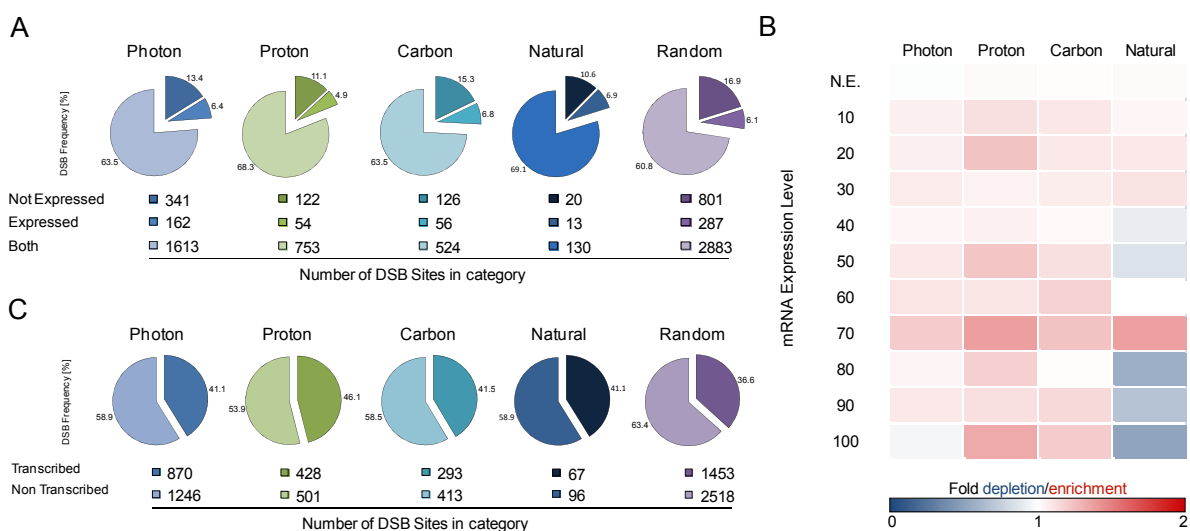
Figure 25: Transcriptional activity does not influence DSB site distribution in irradiated A549 cells. (A) More than 80% of DSB sites were associated to mRNA transcriptional signatures of which the majority of DSB sites (65.7%) was associated to both expressed and not-expressed mRNA signatures. (B) Radiation-induced and naturally-occurring DSB sites are not statistically significant enriched at mRNA signatures of specific activity. Numbers indicate the mRNA expression level. (C) In transcribed genomic regions, the DSB frequency was increased up to 1.25-fold compared to randomly distributed DSB. For statistics, Fisher's exact test was used. Asterix: p-value <0.05; N.E. = Not expressed

Using these criteria, more than 80% of the DSB sites identified in irradiated and cultivated A549 were associated to mRNA transcriptional signatures. Of these, in average, 6.2% were associated exclusively to active and 12.5% exclusively to silent transcriptional mRNA signatures. The majority of radiation-induced and repaired DSB sites (65.7%) were associated to both active and silent mRNA signatures (Figure 25). In order to determine if DSB sites are more frequently associated with highly active rather than weakly active transcriptional mRNA signatures, active mRNA signatures were further subdivided into ten bins according to their signal intensity. Subsequently, the frequency of DSB sites for each mRNA bin was calculated and normalized to the DSB frequency for the randomly distributed DSB locations. An increase in the radiation-induced DSB frequency at highly active mRNA signatures was observed (mRNA expression activity ≥50%) (Figure 25). However, this increase was not statistically significant. Further, no statistically significant enrichment or depletion for naturally-occurring DSB sites at active mRNA signatures was observed. Finally, to analyze whether transcribed genomic regions have a higher probability for DSB induction and repair than non-transcribed regions, the DSB frequency in transcribed genomic regions was calculated and compared to the randomly distributed DSB data set. A genomic region was

found to be transcribed when more than four active mRNA signatures were located 250kb up- or downstream of the DSB site. Radiation-induced DSB showed a minor, up to 1.25-fold enrichment in transcribed genomic regions compared to the randomly-distributed DSB locations (Figure 25). Taken together, these results show that the transcriptional activity of the genome at the point of irradiation does not influence DSB site distribution.

3.4.2 The location of radiation-induced DSB sites is composed of histone modifications defining active chromatin

The accessibility of the genome is regulated by modifying different histone variants. These histone modifications influence DNA repair pathway choice and the activity of the DNA repair machinery [2]. The distribution of radiation-induced and repaired DSB sites and genomic instabilities with respect to histone modifications and genome accessibility in radiation- surviving cell populations has not been analyzed to date. Therefore, the DNase I hypersensitive sites (DHS), histone modifications and transcription factor binding sites (TFBS) at radiation-induced DSB site in expanded cell populations were analyzed.

DHS are genomic regions sensitive to cleavage by the endonuclease DNaseI. In these genomic regions, the chromatin is not condensed and the DNA exposed. Generally, these regions are associated to transcriptional activity, since this chromatin state is necessary for transcription factor binding [90]. To investigate whether induced and repaired DSB sites are associated to DHS, DHS data available from the ENCODE project repository were correlated with DSB sites from irradiated A549 and NHDF-A. A DSB site was termed DHS-associated, if the DSB site was identified within the start and end position of the DHS. The DSB frequency was then normalized to the random DSB frequency, in order to obtain an enrichment value. In irradiated A549 cancer cells, a 2-fold higher frequency of DSB sites located in DHS (p-value $< 2 \times 10^{-4}$, Fisher's exact test) was observed compared to randomly-distributed DSB locations. In irradiated NHDF-A, no enrichment of DSB sites in DHS was observed (Figure 26, Supplementary Table 4). In contrast, the average frequency of doxorubicin-induced DSB sites in DHS was 4.3%, representing a 1.7-fold increase in DSB sites in DHS compared to the random DSB data set (Figure 26, Supplementary Table 4). In irradiated NHDF-A with impaired NHEJ-repair, no significant enrichment or depletion of DSB in DHS was observed at any of the analyzed passages (p4, p6, p8) (Figure 26, Supplementary Table 4). Moreover, the signal intensity of the DSB-associated DHS did not influence DSB site distribution (Supplementary Figure 10).
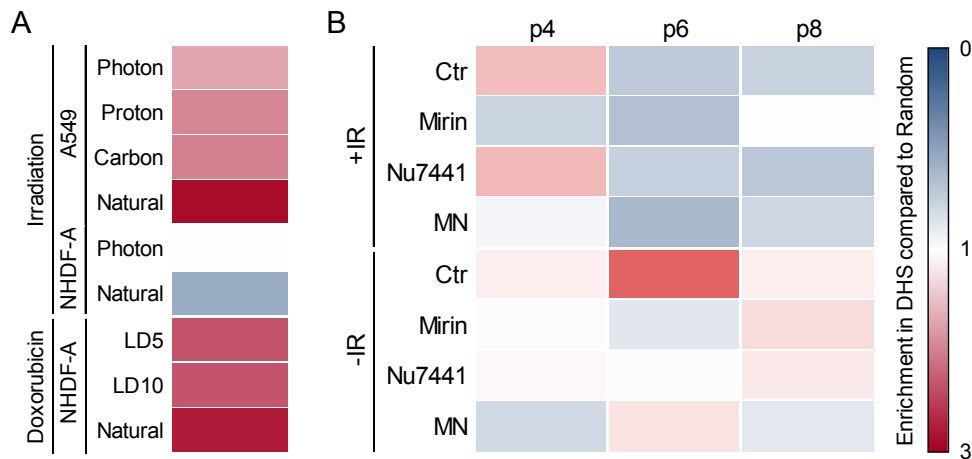
Figure 26: Enrichment of radiation- and doxorubicin-induced DSB sites in DNaseI HS. (A) Radiation and doxorubicin-induced as well as naturally-occurring DSB sites were significantly enriched in DHS. (B) In NHEJ-impaired NHDF-A no significant enrichment of radiation-induced DSB sites in DHS was observed. For statistics, Fisher's exact test was used. MN: Mirin+Nu7441

Interestingly, radiation- induced DSB in NHDF-A were not randomly distributed inside DHS. DSB sites showed a preferential localization downstream of the center of the DHS with an enrichment up to 11-fold as expected by chance (Figure 27). Similarly, doxorubicin-induced DSB sites were also located downstream of the DHS center more frequently.
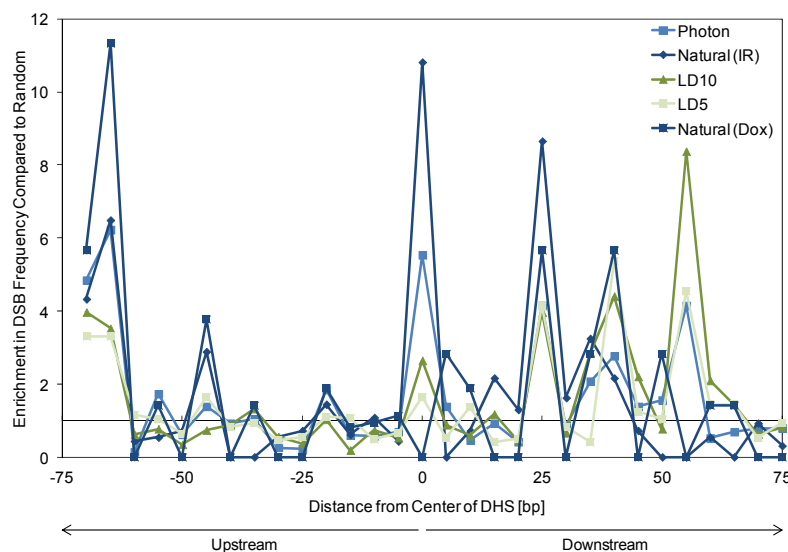
Figure 27: Distribution of DSB sites inside DHS in NHDF-A. Radiation- and doxorubicin-induced as well as naturally-occurring DSB sites were enriched at the beginning and downstream of the DHS center. LD5: Lethal Dose 5 doxorubicin; LD10: Lethal Dose 10 doxorubicin; IR: irradiation; Natural (IR): DSB sites non-irradiated control cells; Natural (Dox): DSB sites in control cells not treated with doxorubicin

In order to analyze whether DSB sites also formed in proximity to DHS, the radius around the DSB sites was expanded up to 10kb. At radii of 2kb, 5kb and 10kb, no differences in the frequency of radiation-, doxorubicin-induced or naturally-occurring DSB sites compared to the random DSB locations were observed (Supplementary Figure 11), indicating that no accumulation of DSB sites occurs in vicinity to open chromatin.

To further elucidate the influence of genomic accessibility on DSB distribution and genomic instability, the association of induced and repaired DSB sites to histone modifications was analyzed. Therefore, histone

modification data were correlated with the A549 and NHDF-A DSB sites. A DSB site was considered to be histone-associated, if the histone modification was located within 2kb around the DSB site. H3K4me1/2/3, H3K9ac, H3K27ac and H4K20me1 are markers for euchromatin, and H3K9me3 and H3K27me3 for heterochromatin. Genomic regions containing both eu- and heterochromatic histone modifications were termed border regions. Using these criteria, radiation- and doxorubicin-induced DSB sites in NHDF-A and A549 were found to be frequently enriched in euchromatin and strongly depleted in heterochromatin. Interestingly, both radiation- and doxorubicin-induced DSB sites were also significantly enriched in border regions of eu- and heterochromatin (Figure 28, Supplementary Table 5). In agreement with this observation, CTCF transcription factor binding sites, which mark genomic border regions of eu- and heterochromatin, were also enriched for induced and repaired DSB sites. The frequency of doxorubicin-induced DSB sites at CTCF binding sites was increased 1.9-fold compared to the random DSB distribution (p-value $< 5 \times 10^{-3}$, Fisher's exact test) (Supplementary Table 6). In NHDF-A with impaired NHEJ-repair activity, the frequency of DSB sites associated to border regions and euchromatin increased during cultivation whereas the frequency of DSB sites in heterochromatic regions slightly decreased over time from passage 4 to passage 8 (Supplementary Figure 12). However, the DSB frequency was below the frequency of the randomly-distributed DSB locations. A significant association of DSB sites in NHEJ-impaired NHDF-A to CTCF-binding sites was only observed in passage 8 of irradiated and Nu7441-treated cells (p-value $< 2.8 \times 10^{-9}$; Fisher's exact test) (Supplementary Table 6).
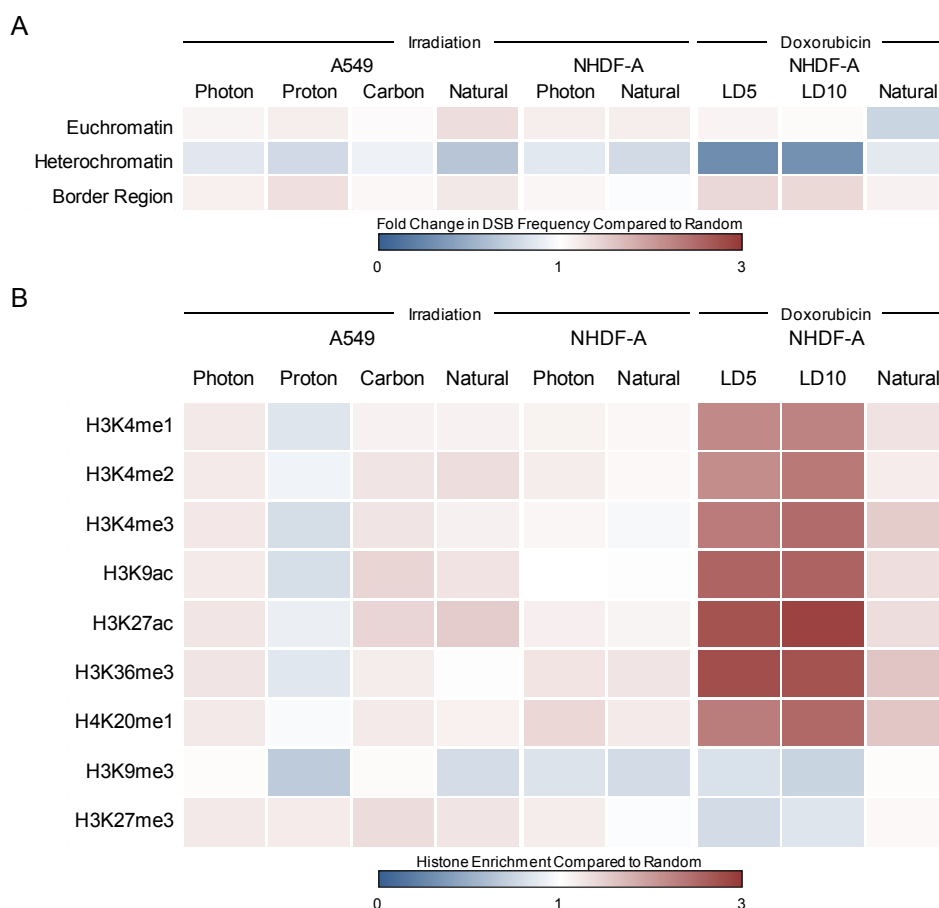
Figure 28: Distribution of DSB sites with respect to chromatin and histone modifications. (A) DSB sites in irradiated and doxorubicin-treated cells were significantly enriched in euchromatin and border regions and depleted in heterochromatin. (B) Histone modifications representing transcriptional silencing (H3K9me3, H3K27me3) were depleted at radiation- and doxorubicin-induced DSB sites in A549 and NHDF-A, whereas histone modifications representing transcriptional activation such as H3K27ac and H3K4me2 were enriched.

Next, the histone modifications associated to DSB sites in radiation- and doxorubicin-surviving cell populations were further analyzed (Figure 28, Supplementary Table 7). In irradiated A549, an enrichment of the histone modifications H3K4me1/2/3, H3K9ac and H3K27ac at carbon- and photon-induced DSB sites was observed (Figure 28). Further, genomic regions marked by H3K9me3, which represents heterochromatin, were depleted for DSB sites in irradiated A549. A similar association pattern of histone modifications was also observed in irradiated and doxorubicin-treated NHDF-A (Figure 28). However, the enrichment of euchromatic histone modifications was more pronounced in doxorubicin-treated NHDF-A than in irradiated cells. In doxorubicin-treated NHDF-A, all euchromatic histone modifications were enriched (Figure 28). In NHDF-A with impaired NHEJ-repair activity, an association of euchromatic histone modifications with radiation-induced DSB sites was observed as well (Supplementary Figure 13). In NHDF-A treated with Nu7441, the frequency of the histone modifications H3K4me2, H3K9ac, and H4K20me1 increased at radiation-induced DSB sites during cultivation, whereas the frequency of heterochromatic histone modifications H3K9me3 and H3K27me3 remained nearly constant (Supplementary Figure 13). The comparison of radiation-induced with naturally-occurring DSB in A549 and NHDF-A revealed that radiation-induced DSB were less frequently depleted in heterochromatin than naturally-occurring DSB (Figure 28). In contrast, the level of DSB depletion is more pronounced in heterochromatin of doxorubicin-treated cells. Moreover, the frequency of doxorubicin-induced DSB was increased in euchromatin compared to naturally-occurring DSB in untreated cells (Figure 28).

In the genome, histone modifications also contribute to fine-tuning of expression levels and genomic structures by combining different histone modifications in the same region. In embryonic stem cells for example, specific promoters at genes involved in the supervision of cell fate decisions and differentiation carry both activating and repressing histone modifications [91] and are termed bivalent promoters. Therefore, the frequency of histone combinations at radiation-induced DSB sites was analyzed and compared to the random and naturally-occurring DSB locations (Figure 29).

A — Irradiation (Natural to the right / Photon below)

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3	H3K79me2
H3K4me1		1,00	1,03	1,01	0,97	1,00	1,03	1,11	0,90	1,06
H3K4me2	0,99		1,03	0,94	0,95	0,96	1,07	0,98	0,92	0,98
H3K4me3	0,99	1,04		1,03	1,02	0,97	0,97	1,07	1,11	0,97
H3K9ac	0,94	0,97	1,10		1,06	0,98	0,96	1,05	0,96	0,94
H3K27ac	0,95	1,00	1,11	1,06		1,00	1,00	1,02	0,92	0,95
H3K36me3	1,07	1,06	1,08	0,99	1,01		0,97	0,95	0,85	0,91
H4K20me1	0,99	1,10	1,12	0,98	1,04	0,99		1,02	1,03	1,01
H3K9me3	1,07	0,98	1,14	1,02	1,05	1,04	1,02		1,10	0,84
H3K27me3	0,86	1,00	1,22	0,87	0,99	0,99	1,07	1,02		1,03
H3K79me2	1,11	1,04	1,08	0,96	0,98	0,92	1,13	0,93	0,99	

B — Doxorubicin (Natural to the right / Doxorubicin below)

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3	H3K79me2
H3K4me1		1,10	1,14	1,17	1,13	1,15	1,23	1,04	1,53	1,23
H3K4me2	1,28		0,94	1,03	0,94	0,99	0,88	0,96	1,10	0,97
H3K4me3	1,31	1,25		0,96	0,96	0,98	0,87	0,91	1,08	0,93
H3K9ac	1,27	1,28	1,27		1,00	1,05	0,95	0,92	1,16	1,01
H3K27ac	1,29	1,28	1,29	1,32		1,16	1,11	1,15	1,25	1,17
H3K36me3	1,28	1,29	1,31	1,32	1,31		1,05	1,12	1,24	1,17
H4K20me1	1,28	1,27	1,32	1,39	1,37	1,25		1,27	1,17	1,11
H3K9me3	1,24	1,33	1,37	1,39	1,33	1,35	1,31		1,04	1,33
H3K27me3	1,40	1,30	1,28	1,42	1,49	1,38	1,30	1,22		0,68
H3K79me2	1,30	1,31	1,31	1,37	1,36	1,29	1,23	1,31	0,71	

Figure 29: Combinations of histone modifications at DSB sites in irradiated and doxorubicin-treated NHDF-A. (A) Specific histone combinations were enriched at DSB sites. In particular, combinations of H3K4me3 at promoters with other histone modifications in irradiated NHDF-A were more frequently identified at DSB sites than expected by chance. (B) In doxorubicin-treated NHDF-A, combinations of H3K4me2, me3 and H3K9ac were enriched compared to random and naturally-occurring DSB. Values in boxes represent the enrichment value for each histone combination compared to the random DSB frequency. Values to the right side of the black boxes are histone enrichments values at naturally-occurring DSB, and values below the black boxes are the values for radiation- (A) and doxorubicin-induced (B) DSB.

At radiation-induced DSB sites in NHDF-A, the combinations of H3K4me3 with H3K9ac, H3K27ac, H3K36me3 and H4K20me1 were exclusively enriched. H3K4me3 and H3K27me3 were enriched at both radiation-induced and naturally-occurring DSB sites and were reported to mark bivalent promoters of developmental genes (Figure 29). In doxorubicin-treated NHDF-A, with the exception of the histone combination H3K79me2 and H3K27me3, all combinations were enriched, and the enrichment values were higher than in untreated control cells (Figure 29). In particular, combinations of H3K4me3 with H3K27me3, H3K9ac with H4K20me1, H3K9me3, H3K27me3, and H3K27ac with H4K20me1, H3K9me3 and H3K27me3 were enriched. Similar to radiation-induced DSB sites, the respective histone modification combinations mark bivalent promoters, active promoters and transcription start sites. In NHEJ-impaired NHDF-A, the combinations of histone marks was stable over the three passages. Enrichment of histone combinations involving H3K27me3 and H3K79me2 with H3K9me3 and H4K20me1 was observed (Supplementary Figure 14). This pattern was similar in all irradiated NHDF-A treated with NHEJ-inhibitors and irradiated cells with active NHEJ-repair (Supplementary Figure 14). In irradiated A549 cancer cells, a strong association of radiation-induced and naturally-occurring DSB to histone combinations of H3K27me3 with H3K4me3, H3K9ac, H3K27ac and H3K36me3 was observed. This pattern was consistent for all radiation sources (Supplementary Figure 15).

Taken together, these results show that the frequency of radiation-induced DSB sites is elevated in accessible genomic regions and in regions regulating gene expression.

3.4.3   Radiation-induced DSB in radiation-survivor cells are enriched in genes and networks regulating cell survival

To date, it remains unknown whether cell survival and growth is influenced by the damage and repair of specific gene classes and/or networks. To address this question, the function of genes which showed DSB labeling was analyzed using the Ingenuity Pathway Analysis (IPA) platform. Only genes with intragenic IDLV labeling were included. In total, 1727 and 2557 genes showed intragenic DSB labeling in photon-irradiated and doxorubicin-treated NHDF-A, respectively. These genes were compared to 1183 and 1566 genes identified in non-treated control cells and in the randomly distributed DSB locations, respectively. The gene sets from all groups were assigned to 275 and 386 categories in irradiated and doxorubicin-treated NHDF-A, respectively. In order to determine the significance of certain functional categories in the list, the enrichment relative to the total number of genes in the random DSB data set in the respective categories was calculated.

Several gene classes were enriched for radiation- and doxorubicin-induced DSB. In particular, genes involved in the regulation of cell cycle, cell death and gene expression were significantly enriched for both radiation- and doxorubicin-induced DSB sites. Moreover, doxorubicin-induced DSB also showed a preference to be located in genes associated with DNA replication, RNA post-transcriptional modification, protein degradation and protein synthesis. By analyzing the subcategories of each functional group, differences in the distribution of radiation-induced, doxorubicin-induced, naturally-occurring and randomly-distributed DSB sites became apparent. A strong preference for radiation and docxorubicin-induced DSB sites to be located in genes involved in cell cycle regulation was observed (p-value < 0.05; Student's *t*-test). In particular, genes associated to cell cycle progression showed an increased frequency of both radiation-induced and naturally-occurring DSB sites. Furthermore, genes regulating G1-, M- and interphase were exclusively enriched for radiation- and doxorubicin-induced DSB. In irradiated and doxorubicin-treated NHDF-A, genes with molecular functions involved in the regulation of cell death were significantly enriched (p-value < 0.05; Student's *t*-test). In particular, genes involved

in regulating apoptosis showed a strong enrichment. This enrichment was exclusive for radiation- and doxorubicin-induced DSB, and did not occur in non-irradiated cells and the random DSB data set (Figure 30). Compared to the random DSB locations, a significant enrichment of radiation-induced DSB in genes that function in the regulation of gene expression was observed as well (p-value $< 5\times10^{-3}$; Student's t-test). Radiation-induced DSB sites were exclusively enriched in genes that activate expression of DNA and RNA (Figure 30).
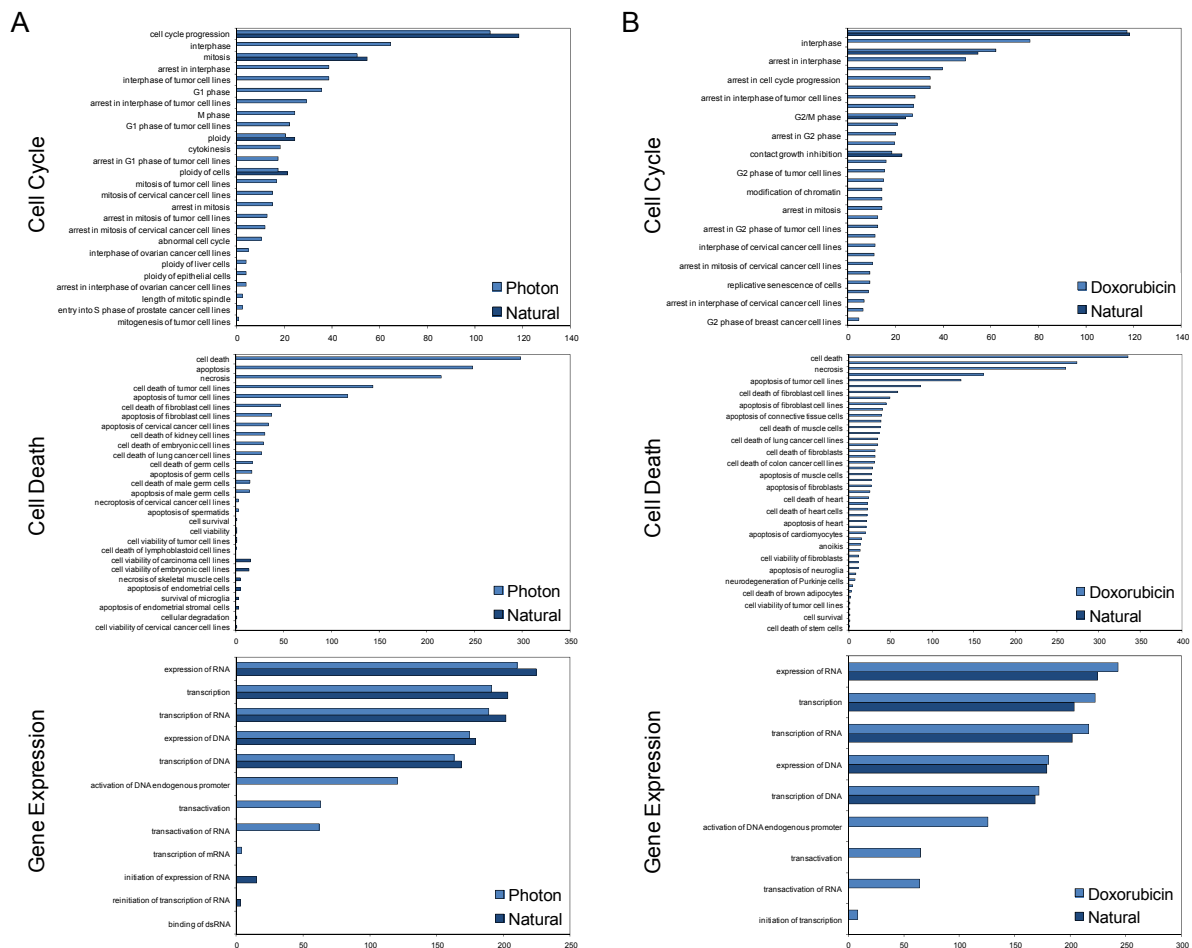
Figure 30: Ingenuity Pathway Analysis (IPA) of radiation- and doxorubicin-induced DSB sites. Genes in irradiated (A) and doxorubicin-treated (B) NHDF-A cells with IDLV integration within the gene-coding region were classified according to physiological function. Enriched genes were frequently involved in processes regulating cell cycle, cell death and gene expression. Within each category, specific disease and function annotations were enriched compared to the random DSB locations. The x-axis represents the enrichment of DSB sites in the respective gene function compared to the random DSB sites.

These results indicate that the specific gene classes are repaired more frequently upon genotoxic stress, and that irradiation and doxorubicin treatment affect the same gene categories. Therefore, the distribution of induced and repaired DSB sites might also be non-random with respect to upstream regulators. Thus, genes with intragenic DSB labeling were also analyzed with respect to upstream regulators and miRNA seed sequences. An association of radiation- and doxorubicin-induced and repaired genes to upstream regulators, including transcriptional regulators, growth factors and microRNA seed sequences in NHDF-A was observed. Within each type of upstream regulator, specific molecules were identified that regulate the expression of genes marked by radiation- and doxorubicin-induced DSB. In the group of transcriptional regulators, DSB sites were located in genes regulated by the stress- and apoptosis-regulators FOS, TP53 and ERG. MicroRNAs regulate the expression of several target mRNAs by binding to specific sequences and inducing their degradation before translation.

IDLV-marked genes in irradiated and doxorubicin-treated NHDF-A were regulated by several microRNA including miR182-5p, miR185-5p, miR181, and miR145 (Supplementary Table 8, Supplementary Table 9), which have been reported to be involved in the DNA damage response following irradiation [92]. These results indicate that genes involved in maintaining genomic integrity and cell viability and are regulated by specific molecules are preferentially repaired and/or are selected for during expansion.

3.4.4   Identification and analysis of radiation-induced DSB sites over time

The clonal distribution of modified cells in clinical gene therapy trials is frequently studied by analysing lentiviral insertion sites (IS) [88]. Since every IS is located at a different position in the genome of each cell, IS represent unique marks for every cell and allow to determine the clonality and fate of genetically modified cells [79]. Thus, the identification of radiation-induced DSB sites in different passages after irradiation may also enable the tracking of DSB sites over time, reflecting the dynamics in the clonality of irradiated cells.

In total, 975 DSB sites were identified in irradiated NHDF-A in at least two of the three passages (p4, p6, p8). The majority of these DSB sites (726) were identified in both p6 and p8, and 43 DSB were identified in all three passages. The relative frequency of eight of the DSB sites present in all three passages increased from p4 to p8, one remained unchanged and 34 DSB sites decreased. In irradiated NHDF-A with active NHEJ-repair, six DSB sites were identified, of which five decreased. In irradiated NHDF-A with impaired NHEJ-repair, the majority of DSB sites also decreased over time (Supplementary Table 10). The strongest increase of a single DSB was observed in mirin-treated and irradiated NHDF-A. The relative frequency increased more than 21-fold from p4 (0.019%) to p6 (0.411%) and was located 14kb upstream of the transcription start site of the *MED14* gene. This genomic region is characterized by the presence of H3K4me1, H3K4me2, H3K4me3 and H3K27me3 histone modifications. The combination of H3K4me2 and H3K27me3 is associated with poised promoters at embryonic stem cell genes [91]. The strongest decrease in the relative DSB frequency was observed for a naturally-occurring DSB site in non-irradiated NHDF-A with active NHEJ-repair activity, located more than 615kb downstream of the *DIAPH3* gene (0.12% in p4 and 0.001% in p8) and was associated with the heterochromatin histone mark H3K9me3 (Supplementary Table 10). However, no general association of specific histone modifications with an increase or decrease in DSB sites over time was observable.

Comparison of the ten most prominent DSB sites in irradiated NHDF-A in passage 4, 6 and 8 is presented in figure 31. In irradiated NHDF-A with active NHEJ-repair activity, only a single DSB site which was located in intron 7 of *FANCD2* gene was present in two subsequent passages with a low retrieval frequency (2.4% and 1.5%). In passage 8 in irradiated NHDF-A with active NHEJ-repair, a single DSB site located in the *C12orf63* open reading frame with a relative frequency of 23% was present. In NHDF-A with impaired NHEJ-repair activity, no dominance of a DSB site was observed in any passage. The total contribution of the top ten DSB sites in each passage did not exceed 30% of the total DSB sites. In mirin+Nu7441-treated and irradiated as well as Nu7441-treated and irradiated NHDF-A, one and two DSB sites occurred in two passages within the top ten DSB sites, respectively (Figure 31). The two DSB sites in Nu7441-treated and irradiated NHDF-A were associated to a gene involved in calcium ion transmembrane transporter activity (*ITPR2*) and one with unknown function (*TMEM90A*). In mirin+Nu7441-treated and irradiated NHDF-A, the single DSB site was associated to the gene *AMOT*, whose function is related to angiostatin binding and receptor activity.

+ IR Control

Rank	p4		p6		p8	
1.	KIAA0564	4.49	IMMP2L	6.59	C12orf63	23.34
2.	GATA3	3.12	HIST1H2AH	6.30	ADRBK2	2.67
3.	MMP13	3.05	FAM84A	4.66	C2orf86	1.77
4.	YWHAZ	2.12	SLITRK1	4.19	FANCD2	1.48
5.	AUTS2	1.75	GSDMC	3.84	UBE2D3	1.45
6.	ADARB2	1.56	PCGF3	3.10	PTPLA	1.34
7.	MBNL1	1.25	FANCD2	2.40	INTU	1.32
8.	NTRK3	1.06	KCNMB2	2.20	APTX	1.30
9.	SUFU	1.00	MYH13	2.04	C12orf50	1.17
10.	BAI3	1.00	ABR	1.50	SPIN4	1.03
All other IS	793	79.61	1184	63.18	1749	63.14

+ IR +Mirin

Rank	p4		p6		p8	
1.	PLXDC2	1.83	NFAT5	7.69	FANCL	3.26
2.	DPP10	1.47	SH3BP4	5.46	PXMP3	2.61
3.	PFTK1	1.41	PTPN12	5.39	RARB	2.34
4.	ATAD1	1.28	FNDC3A	4.93	BARD1	1.85
5.	CNTNAP5	1.16	PDGFD	4.23	KLHDC6	1.85
6.	LPHN2	1.10	SLC4A4	3.61	MED14	1.74
7.	KCNH7	1.10	TBC1D1	2.61	JAG1	1.59
8.	ARHGAP6	1.10	MDM1	2.33	ZNF33B	1.58
9.	EFCAB1	1.04	C15orf41	1.96	STXBP6	1.57
10.	FAM19A1	0.98	POU2F1	1.43	EPHA4	1.53
All other IS	754	87.54	1449	60.36	3919	80.08

+ IR +Nu7441

Rank	p4		p6		p8	
1.	CNTN3	2.07	BCOR	5.74	TMEM90A	7.44
2.	ZNF33B	1.97	SPTLC3	4.17	ABCA13	2.75
3.	CDH12	1.97	ITPR2	2.82	ZP4	1.83
4.	WNT7A	1.68	SYT17	2.71	GRIA2	1.73
5.	TRIM2	1.58	OR2AT4	2.05	RABGAP1L	1.64
6.	KCNJ3	1.38	FOXR2	2.00	ITPR2	1.54
7.	ANO4	1.28	HACE1	1.88	KIAA1377	1.40
8.	SMYD2	0.99	TMEM90A	1.81	FLRT2	1.39
9.	GPR123	0.99	OLFM4	1.56	CASK	1.31
10.	POU5F1B	0.79	IMMT	1.48	BMP6	1.26
All other IS	606	85.31	1917	73.79	3121	77.72

+ IR +Mirin +Nu7441

Rank	p4		p6		p8	
1.	ITGA10	1.18	RFNG	4.07	OR1F1	2.09
2.	CPEB3	1.18	SRrp35	3.78	CXCR4	1.46
3.	KLHL2	1.18	SPOCK1	3.69	ZNF706	1.37
4.	KIF2B	1.11	AMOT	3.57	CDKAL1	1.07
5.	CMBL	1.11	PLB1	3.36	AMOT	1.04
6.	STAG1	0.96	SLC37A3	3.20	BCHE	1.02
7.	FAM47A	0.89	SORL1	1.72	P4HA1	1.00
8.	CDH8	0.81	ZNF25	1.46	DPYD	0.96
9.	AP1S3	0.81	MDM4	1.44	CCDC91	0.95
10.	CBLN2	0.74	PXK	1.35	SAFB	0.95
All other IS	777	90.01	2141	72.36	4233	88.09

Figure 31: Top ten DSB sites in irradiated NHDF-A with active (Control) and impaired NHEJ-repair activity in passage 4, 6 and 8. The names of the DSB-associated RefSeq genes and the relative retrival frequencies are shown in the tables for p4, p6 and p8. IR: irradiated

In order to delineate genomic factors that potentially influence the changes within the top ten DSB sites, the DSB positions relative to the DSB-associated gene were investigated. In irradiated NHDF-A with active NHEJ-repair, four, three and three of the top ten DSB sites in passage 4, 6 and 8, respectively, were located upstream of the TSS. In NHEJ-impaired and irradiated NHDF-A, the number of DSB sites located upstream of TSS either remained constant (mirin), decreased (Nu7441) or increased (mirin and Nu7441) during cellular expansion. Similar results were observed for DSB sites located in genes and downstream of the transcription termination site (Supplementary Table 11). Thus, these results indicate that the changes observed in the top ten DSB site in p4, p6 and p8 cannot simply be explained by the location of the DSB sites in gene-coding regions.

Therefore, the histone modifications in the vicinity of DSB sites might explain the changes observed in the most prominent DSB sites during cultivation. Therefore, the DSB sites from each passage were correlated with histone modification data. Radiation-induced DSB sites in NHEJ-proficient cells increasingly accumulated in border regions and disappeared in heterochromatin. In irradiated NHDF-A with impaired NHEJ-repair activity, an increase of the top ten DSB sites in border regions was observed as well (Figure 32). Analysis of the histone modifications in vicinity of DSB sites did not show an association of specific histone modifications with DSB sites.
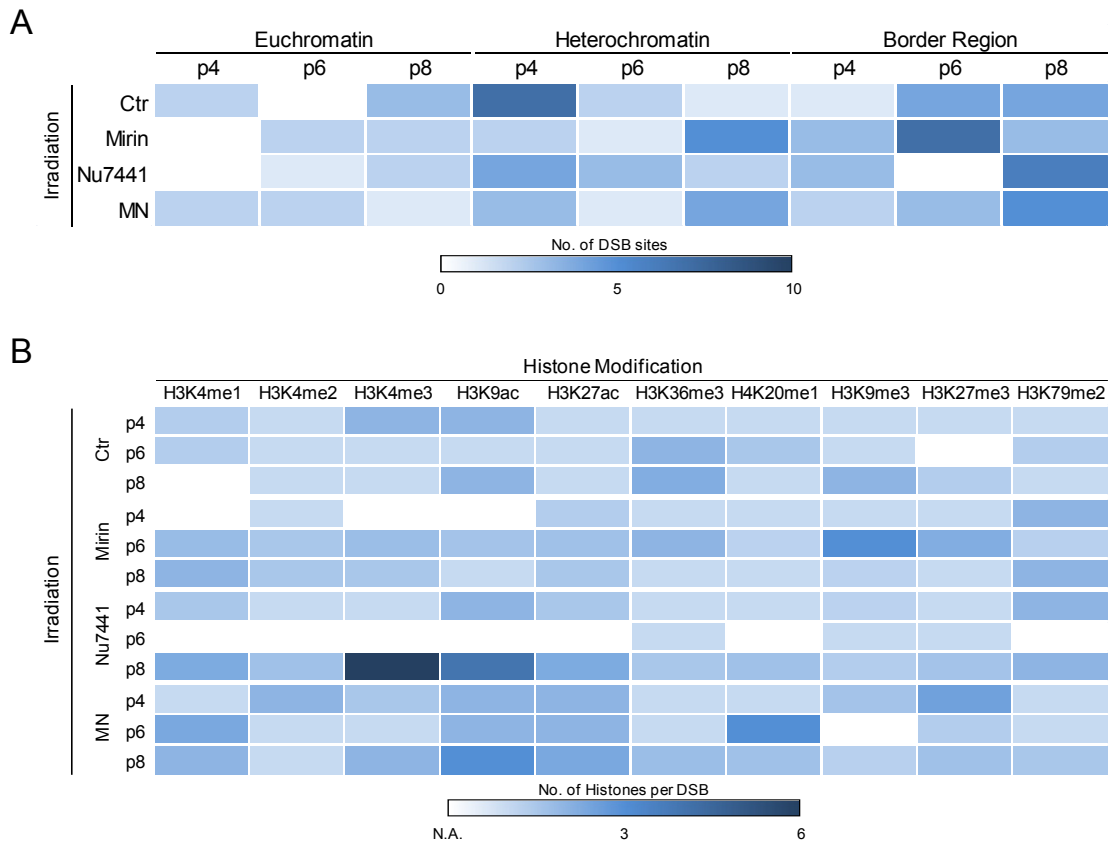
Figure 32: Histone modifications associated with the top ten DSB sites in irradiated NHDF-A in passage 4, 6 and 8. (A) Radiation-induced DSB sites in NHDF-A increasingly accumulate in genomic border regions that marked by both eu- and heterochromatin histone modifications. (B) The number of histone modifications at radiation-induced DSB sites in indicated passages. N.A.: No association of DSB site with histone modification

3.5   Identification of frequently damaged and repaired genomic regions

3.5.1   DSB Trapping in irradiated and passaged cells reveals common regions of radiation-induced and repaired damage

To date, it remains unknown whether genomic regions with increased vulnerability exist and whether the accumulation of DNA damage in specific genomic regions influences cell survival. Such genomic regions would be expected to show a non-random accumulation of DNA damage and genomic instabilities. Since immunostaining does not allow stable genetic tagging of induced and repaired DNA damage in surviving cell populations over time, IDLV-based DSB trapping might identify genomic areas in which DNA damage builds up. To analyze whether clustering of induced and repaired DSB sites occurs, a statistical method was applied which was previously used to analyze the distribution pattern of retroviral integration sites [84]. This methodology identifies the number of DNA damages marked by IDLV integrations in genomic regions with a specific size (Figure 33). Generally, two, three, four and five IDLV copies in a genomic region of 30, 50, 100 and 200kb were considered to be non-randomly distributed and defined as DSB hotspots of second, third, fourth and higher order, respectively (Figure 33). Photon-, proton-, and carbon-induced DSB sites in a cancer cell line were combined as radiation-induced DSB sites. Moreover, genomic regions which show accumulation of DSB sites were termed DSB hotspots. If a DSB hotspot only contained radiation- or doxorubicin-induced DSB sites, it was termed radiation-or doxorubicin-related DSB hotspot, respectively. Similarly, a genomic region showing

accumulation of naturally-occurring DSB sites only was termed a naturally-occurring DSB hotspot. In case, both naturally-occurring and radiation-/doxorubicin-induced DSB sites occurred in the same hotspot, the hotspot was called to be overlapping.

By applying these criteria to the DSB data sets from irradiated or doxorubicin-treated cells, a non-random distribution of DSB sites became apparent (Figure 33, Table 7). In all three irradiated cancer cell lines, the majority of DSB sites existed as single DSB. In average, only 30% of the DSB sites were clustered. Of these, 17% were located in hotspots with an order of four or higher. In contrast to the cancer cell lines, the clustering of DSB sites upon irradiation and doxorubicin-treatment was increased in primary fibroblasts. In irradiated or doxorubicin-treated NHDF-A, 55.7% and 34.6% of the DSB sites were located in DSB hotspots, respectively. Furthermore, in irradiated NHDF-A, 38.4% of all DSB sites were clustered in genomic hotspots with orders of four and higher. In NHDF-A cells that were treated with doxorubicin, 17.0% of the DSB sites were clustered in DSB hotspots of fourth or higher order. For both irradiated and doxorubicin-treated cancer cells and fibroblasts, no enrichment of DSB clusters on specific chromosomes was observed (Figure 33). In NHDF-A with impaired NHEJ-repair activity, the fraction of DSB sites in hotspots was lower than in irradiated and doxorubicin-treated cells, but increased from passage 4 to 8 from 4.0% up to 27.5%. Similarly, the fraction of DSB sites in hotspots of fourth or higher order increased over time from 1.8% in passage 4 to 22.1% in passage 8 (Table 7).

Table 7: Overview of radiation- and doxorubicin-induced DSB in hotspots in expanded cell populations. Absolute numbers are the sum of both induced and naturally-occurring DSB sites in the respective cell type. +NHEJi: cells treated with NHEJ-repair inhibitors

Cell Type	DSB Induction	#DSB	#DSB in Hotspots	Fraction of DSB in Hotspots [%]	#DSB in Hotspots (Order ≥4)	Fraction of DSB in Hotspots (Order ≥4)
A549	Radiation	4655	1508	32.4	652	14.0
PC3	Radiation	2261	654	28.7	210	9.3
U87	Radiation	2873	871	30.6	304	10.6
NHDF-A	Radiation	10846	6046	55.7	4169	38.4
	Doxorubicin	7991	2767	34.6	1360	17.0
NHDF-A +NHEJi	Radiation (p4)	6187	248	4.0	113	1.8
	Radiation (p6)	8391	1126	13.4	736	8.8
	Radiation (p8)	16804	4619	27.5	3713	22.1
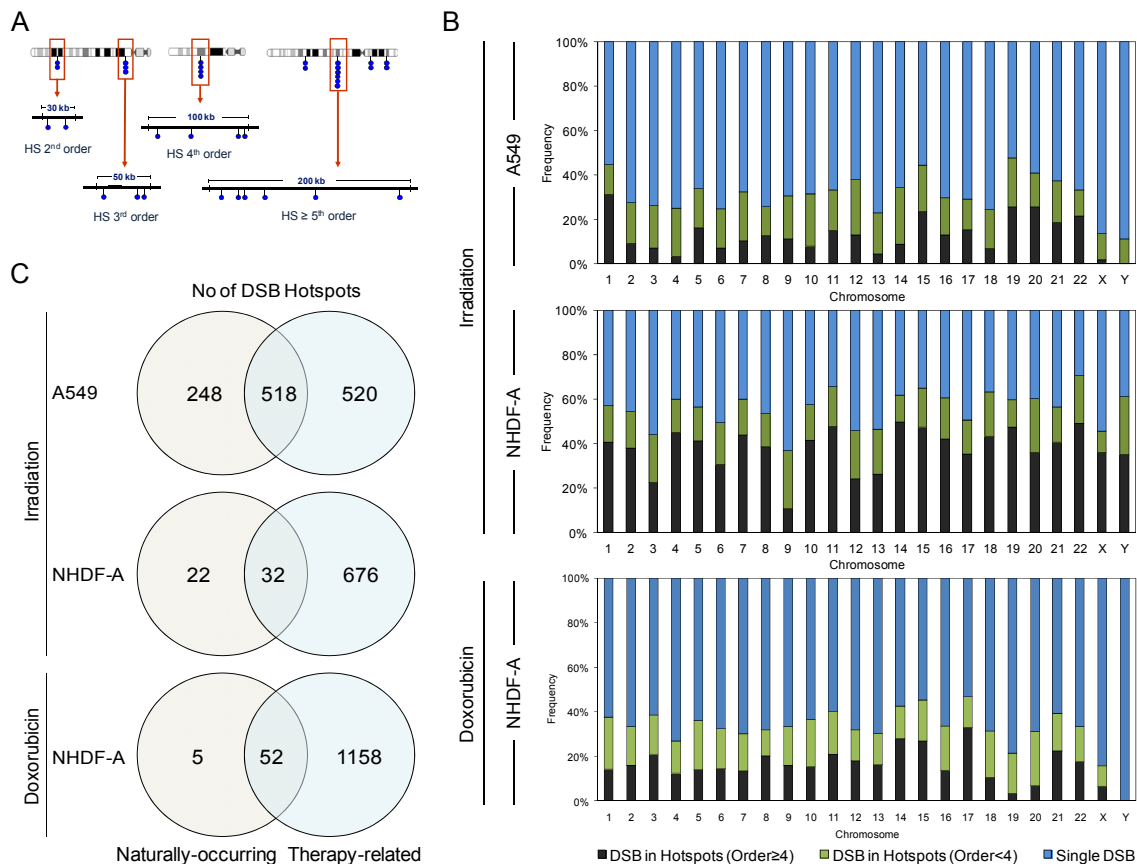
Figure 33: Distribution of DSB hotspots in A549 and NHDF-A. (A) Definition of DSB hotspots. DSB sites are indicated by the blue circles. (B) Chromosomal distribution of DSB hotspots in irradiated A549, irradiated NHDF-A and doxorubicin-treated NHDF-A. In A549 cells about one third (32.4%) of the DSB sites were clustered, and 18.4% of the A549 DSB sites accumulated in hotspots of 2nd and 3rd order. In irradiated NHDF-A, 55.7% of the DSB sites were clustered in genomic hotspots of DSB induction and repair, whereas 17.3% of the DSB were located in DSB hotspot of 2nd and 3rd order. In doxorubicin-treated cells, 34.6% of the DSB were located in hotspots of DSB induction and repair, and 17.0% existed in DSB hotspots of 2nd and 3rd order. Enrichment of DSB clusters on specific chromosomes was not observed. (C) The majority of DSB hotspots in A549 (90%) was exclusively identified in irradiated cells, whereas the frequency of radiation-related DSB hotspots in NHDF-A was 40%. Also, more than 90% of the DSB hotspots identified in NHDF-A were specific for doxorubicin-treated cells. HS: Hotspots, No: Number

The majority of DSB hotspots in cancer cell lines (90%) was exclusive for irradiated cells and not identified in untreated control cells. In NHDF-A treated with doxorubicin, the frequency was increased up to 94%, resembling doxorubicin-related DSB hotspots. In contrast to irradiated cancer cell lines and doxorubicin-treated fibroblasts, the frequency of radiation-related DSB hotspots in NHDF-A was decreased (40%) (Figure 33). In NHDF-A treated with NHEJ inhibitors, the frequency of radiation-related DSB hotspots increased from 25% in passage 4 to 66% in passage 6 and 56% in passage 8 (Supplementary Figure 16). Interestingly, within the radiation-related DSB hotspots the percentage of DSB hotspots containing solely DSB sites from NHEJ-impaired NHDF-A increased from 22% to 62% during passaging (Supplementary Figure 16).

In cancer cell lines, 4.3% of all identified DSB hotspots were also identified in non-irradiated cells. This frequency was further increased in NHDF-A, where close to 50% of the DSB hotspots contain both naturally-occuring as well as radiation-induced DSB sites. In doxorubicin-treated NHDF-A, the frequency of these overlapping DSB hotspots was 1%. These results indicate that radiation- and doxorubicin-induced DNA damage hotspots overlap with naturally-instable genomic regions (Figure 33). The frequency of overlapping DSB hotspots in irradiated NHDF-A with impaired NHEJ-repair remained nearly unchanged over the three passages at about 40%.

The 20 highest-ranked DSB hotspots in radiation-surviving NHDF-A were located inside or in close proximity to known tumor suppressor genes (*FAT1, CUX1, NEGR1*)[93-95], genes driving tumor progression (*TRIM36, ARRB1, TCL6*)[96, 97] and genes that are frequently mutated in different types of cancer (*KIAA0574, TCL6, TRIM29, MOSPD1*)[96-98]. Similarly, an association of 20 highest-ranked DSB hotspots in irradiated cancer cell lines with cancer-relevant genes was observed. Identified genes were associated to cancer (*ATP2B4, SKI, TMED3, TLR4, CDK10, PIAS1*) [99-105], show tumor suppressor activity (*PPP2A, SDCCAG1*) [103, 106] and are involved in the regulation of apoptosis (*BRE*) [107].

In NHDF-A, a total of 504 and 1261 DSB sites was associated to radiation- and doxorubicin-related DSB clusters with an order of four or higher, respectively. Of these, 213 (42.3%) and 1093 (86.7%) DSB were intragenic, with the majority being located in introns. In contrast to doxorubicin-specific DSB clusters in NHDF-A, a preference for radiation-related DSB hotspots to be located within 10kb upstream of the TSS and at the 3` end of genes (Figure 34). With respect to the affected gene transcripts, the orientation of the IDLV vector located in radio- and doxorubicin-related hotspots did not show any preference to be located in sense or antisense orientation (Figure 34).
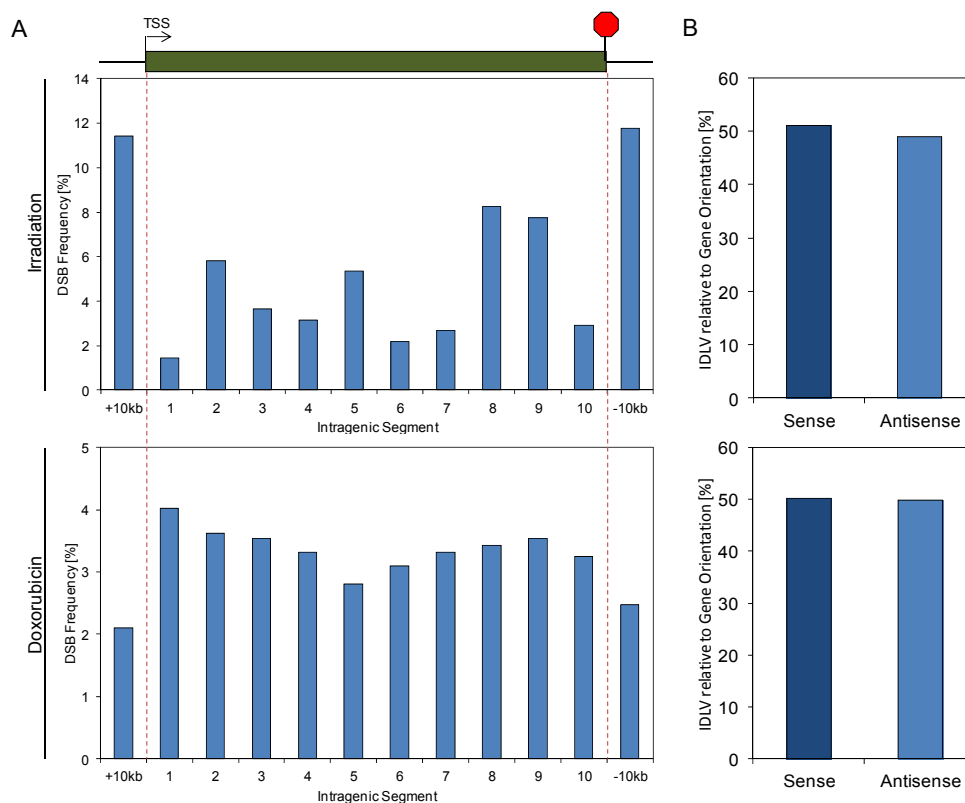
Figure 34: The genomic distribution of radio- and doxorubicin-related DSB hotspots. (A) Radiation-related DSB hotspots (top) preferentially occurred within 10kb up- and downstream of the gene coding region. (B). With respect to the gene orientation, IDLV in radiation- (top) and doxorubicin-related (bottom) DSB hotspots did not show any preference to be located in sense or antisense orientation of the gene.

3.5.2 Radiation-related DSB hotspots overlap with genes involved in maintaining genome stability and DNA repair

In irradiated NHDF-A, 54 radiation-related hotspots with more than four independent DSB per cluster (order >3) were identified, consisting of 88 genes and 412 DSB. More than 50% of these radiation-related DSB hotspots

were intragenic (34 of 54) and predominantely located in introns. Moreover, IPA analysis of the radiation-related DSB hotspots revealed that the majority of the 88 genes in these hotspots are linked to epithelial neoplasias (76%) and cancer (88%) (Supplementary Table 12). The top five radiation-related DSB hotspots in NHDF-A were located inside or in proximity to the genes *ERI1*, *ELF4*, *TOP1*, *ZNF33B* and *ZYX*. Interestingly, with the exception of *ZNF33B*, genetic screens showed that these genes are involved in the maintenance of genome stability, radiation-sensitivity and DNA repair. In an RNAi-based screen, the *TOP1* gene that encodes the enzyme topoisomerase 1 has been identified to sensitize cells to irradiation [108]. Further, *ERI1* is a 3' exonuclease that degrades histone mRNA at the end of DNA replication, which is important for maintaining genome stability. Hence, *ERI1*-deficient cells show increased histone levels that cause chromosome loss and genomic instability [109, 110]. Another RNAi-based screen revealed that *ELF4* and *ZYX* are involved in the regulation of homologous recombination repair [111].

In NHDF-A treated with doxorubicin, 171 doxorubicin-related DSB hotspots (order >4) were identified, consisting of 389 genes and 1053 DSB. Similar to the radiation-related DSB hotspots in NHDF-A, the majority (168 of 171) of the doxorubicin-related DSB hotspots were intragenic and also predominantly located in introns. The majority of DSB hotspots were located in genes associated to neoplasias (68.4%) and cancer (80.7%) (Supplementary Table 13). The top five doxorubicin-related DSB hotspots were located inside or in proximity to the genes *MBTD1/UTP18*, *BIRC6*, *SNX1/PPIB/CSNK1G1*, *RAC1/DAGLB/KDELR2* and *XRCC5*. These genes have been reported to play an important role in the regulation of DNA repair and genomic stability. UTP18 is an important regulator of radiation sensitivity and DNA repair. Upon depletion by RNAi, the frequency of HR is decreased [111] and cells show increased ionizing radiation sensitivity [108]. Moreover, BIRC6, frequently overexpressed in several human cancers, interacts with p53, thereby facilitating its degradation and frequently leading to carcinogenesis [112]. The fourth largest doxorubicin-related DSB hotspot was located in intron 1 of the *RAC1* gene. RAC1 belongs to the Ras kinase superfamily, and RNAi-mediated downregulation of RAC1 leads to reduced cell viability [113]. Furthermore, XRCC5 is the 80kDa subunit of the Ku-heterodimer that recognizes and binds to DSB sites. In complex with Ku70, XRCC5 is required for efficient interaction of Ku heterodimer with DNA-PK and the initiation of NHEJ-repair activity [114]. Interestingly, no doxorubicin-related DSB hotspot overlapped with radiation-related DSB hotspots, indicating that the distribution of DSB hotspots is dependent on the DNA damaging agent.

Impairment of NHEJ-repair activity in NHDF-A prior to irradiation did not influence the distribution of radiation-related DSB sites (Supplementary Table 14). In total, 123 DSB hotspots (order>4) containing 685 DSB sites and 219 genes were identified. The top five radiation-related DSB hotspots in passage 6 were located in proximity or inside *LOC729355*, *RFNG/ CCDC57*, *SHCBP1*, *MID1IP1/ BCOR* and *SORL1*. Even though the function of SHCBP1 is still unknown, it may play a role in signaling pathways governing cellular proliferation, cell growth and differentiation (function by UniProtKB/Swiss-Prot). BCOR is a transcriptional repressor that inhibits gene expression when recruited to promoter regions and may influence apoptosis [115]. The top five radiation-related DSB hotspots in passage 8 were located in proximity or inside the genes *PPIAL4G*, *HERPUD2*, *ZNF804A*, *EFCAB3/METTL2A* and *ST6GALNAC5/PIGK*. Surprisingly, with the exception of the *PIGK* gene, the function of these genes are unknown, and none of these genes has been reported to be involved in maintaining genome stability or regulating cell death. Moreover, only a minority of the radiation-related DSB hotspots in NHEJ-impaired cells in passage 6 and 8 were associated to cancer and neoplasia (Supplementary Table 14).

Taken together, these results indicate that the DSB hotspot analysis identifies known and currently unknown genomic regions that are important for the maintenance of genome stability and cell survival after irradiation and chemotherapy.

3.5.3   DSB hotspots overlap with eu- to heterochromatin border regions

By analyzing the DSB hotspot distribution with respect to chromatin topology, it became apparent that the distribution is influenced by the histone environment. In the irradiated or doxorubicin-treated cells, the majority of naturally-occuring, overlapping and radiation-related DSB hotspots were preferentially located in border regions of the genome (Figure 35). In contrast, all three types of DSB hotspots were not frequently detected in heterochromatic regions. Moreover, both radiation- and doxorubicin-related DSB hotspots in NHDF-A showed association to H3K27ac histones located within 10kb of the DSB hotspots (Figure 35). This histone modification frequently marks active genomic regulatory elements as annotated by the ENCODE project.
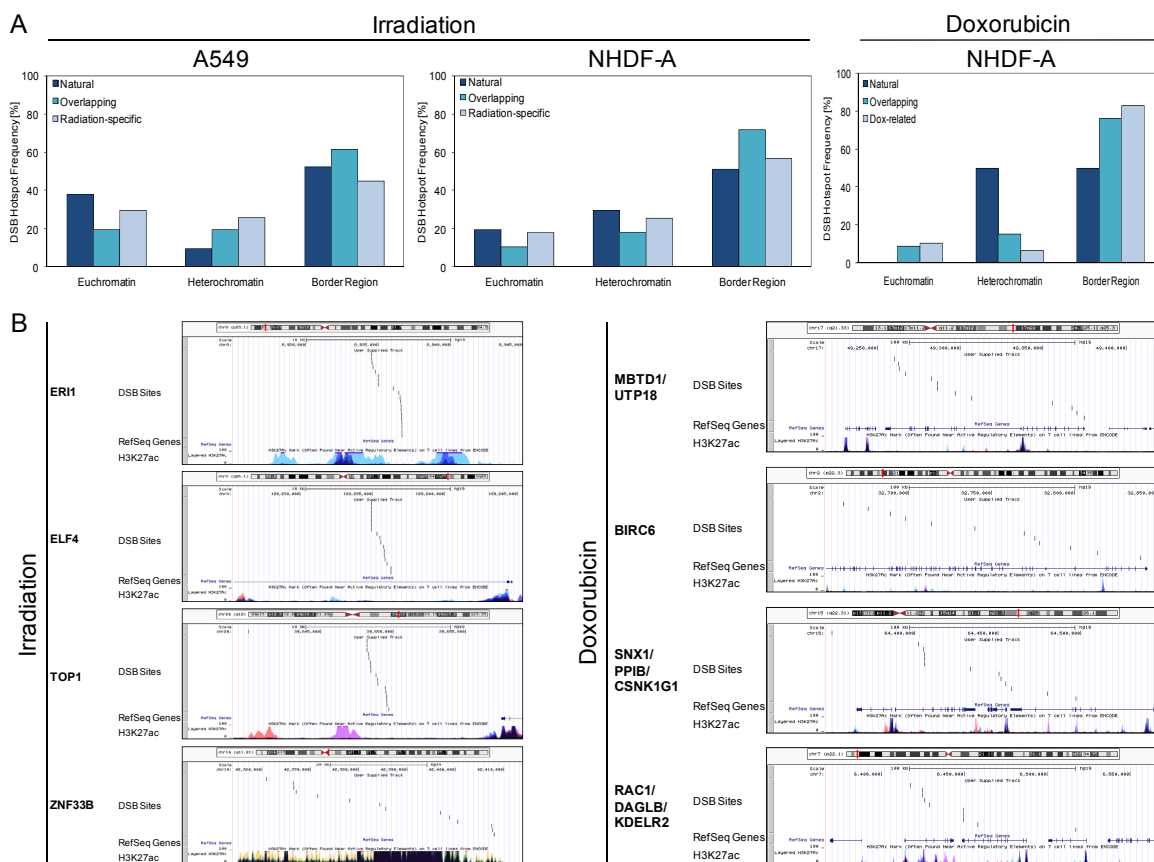
Figure 35: Chromatin topology at radiation- and doxorubicin-related DSB hotspots in A549 and NHDF-A. (A) The histone environment present in radiation- and doxorubicin-related, overlapping and naturally-occurring DSB hotspots were composed of chromatin modifications favoring euchromatin and border regions which contain both of eu- to heterochromatin marks. More than 50% of the radiation-related, doxorubicin-related and naturally-occurring DSB hotspots in A549 and NHDF-A were located in chromatin regions containing histone modifications present in eu- and heterochromatin boundaries. (B) The most intense radiation- and doxorubicin-related DSB hotspots were in direct proximity (<10kb) to gene regulatory elements marked by the presence of H3K27ac. The genes in vicinity to DSB hotspots are depicted. DSB sites in hotspots are displayed in UCSC genome browser.

The frequency of radiation-induced DSB hotspots in NHDF-A increased during expansion in cell culture, whereas the frequency of naturally-occurring hotspots decreased. Further, the proportion of overlapping hotspots in each passage remained constant (Supplementary Figure 15). This change in the relative frequency of the different hotspot types was consistent throughout the different chromatin regions. The frequency of radiation-related DSB hotspots in eu- and heterochromatin as well as border regions increased over time, whereas the

frequency of overlapping DSB hotspots remained constant (Supplementary Figure 15). Interestingly, in passage 4, 56% of the radiation-related DSB hotspots were uniquely identified in NHEJ-impaired NHDF-A and the relative proportion of these hotspots increased to 73% in passage 8. Finally, the association of DSB hotspots with specific histone modifications was analyzed. This analysis revealed that the frequency of specific histone modifications changed during expansion of cells and that these changes were mostly pronounced at radiation-related DSB hotspots. The frequency of H3K4me1/2, H3K27ac and H3K27me3 increased more than 1.5-fold from passage 4 to passage 8. In overlapping hotspots, the histone frequency did not change significantly, whereas in natural DSB hotspots H3K4me3 and H3K27me3 increased more than 1.5-fold. In radiation-related DSB hotspots identified solely in NHEJ-impaired NHDF-A, the frequency of all histone modifications with the exception of H3K27me3 increased more 1.7-fold from passage 4 to 8. The strongest histone enrichment was oberved for H3K79me2 (Supplementary Table 16).

4.   DISCUSSION

For determining the distribution of radiotherapy-induced DSB sites and genomic instabilities in the genome of surviving cancer cell lines and primary human dermal fibroblasts, a total of more than 20,000 unique DSB sites was analyzed and compared to doxorubicin-induced and naturally-occurring DSB sites. The results demonstrate that IDLV trapping and LAM-PCR are suitable methods for the detection of induced and repaired DSB sites and to analyze the genomic factors associated with the DSB sites.

The results presented in this thesis point to a non-random distribution of radiation-induced and repaired DSB sites in the genome of radiation-survivor cells. Radiation- as well as doxorubicin-induced DSB sites were enriched in regions marked by euchromatic histone marks and depleted in heterochromatin. Moreover, radiation- and doxorubicin-induced DSB sites frequently accumulated in small genomic regions which preferentially form at the border of eu- to heterochromatin. Genes located in proximity to these hotspots of DSB induction and repair were associated with cancer formation and other genetic disorders, indicating that IDLV-mediated DSB trapping is a superior method to identify genomic regions which are functionally important for the maintenance of genome stability and possibly for the development of radioresistance. The results and the development of new approaches for direct DSB detection and sequencing will be further explained and discussed.

4.1   Immunostaining of H2AX is not suitable to detect DSB and genomic instabilities in cancer therapy surviving cell populations

To date, the most frequently used method to study DNA repair activity is immunostaining of DNA repair proteins, such as γH2AX. As shown in chapter 3.1.1, the microscopically-visible foci disassemble from the DSB site within 8h after irradiation and chemotherapy, when DSB repair is complete. This is in good agreement with studies on DSB repair kinetics in different cell types [116, 117]. Even though, a few foci may persist as a consequence of complex DNA damages that are difficult to repair [118]. These rare events cannot be efficiently detected, since the number of foci is close to the background level. Moreover, induced DNA damage is not always detectable by immunostaining as observed in the case of etoposide-induced DNA damage. No increase in γH2AX foci in cells treated with etoposide compared to untreated cells was detectable, indicating that TOP2-linked DSB remain undetected by the DDR unless TOP2 is removed enzymatically and the free DNA ends are exposed. It has been reported that only a small subset of all DNA breaks produced by etoposide (0.3%) are sensed by the DNA damage response and activate DNA damage repair [119]. Furthermore, γH2AX and other repair factors accumulating at DSB sites do not always colocalize acurately with the DSB site, but can form foci that span for megabases around the DSB site, making it difficult to identify the exact position of the DSB site [120]. In addition, immunostaining of DNA repair proteins does not enable tracking of sub-lethal DSB sites over time in surviving cells and cannot give any insights into a possible link between DSB induction, repair and cell fate decision [121]. Due to these limitations, there is little information available on whether repaired DSB sites in the genome of surviving cells are distributed randomly in the genome, whether certain genomic or cellular factors influence DSB distribution, and how cancer therapy induces therapy resistance, cellular transformation and carcinogenesis. The analysis of induced and repaired DSB sites in therapy surviving cell populations by placing a stable genetic mark at the DSB site could therefore lead to a better understanding of the mechanisms that initiate and drive DSB-induced therapy resistance and may lead to the development of improved chemo- and radiotherapies.

4.2   NHEJ-mediated IDLV integration at DSB sites stably marks DNA damage and repair sites in living cells

In this thesis, radiation- and doxorubicin-induced and NHEJ-repaired DSB were efficiently marked using an integrase-deficient lentiviral vector (IDLV). The IDLV carries a point mutation in its integrase gene, which blocks its integrase activity, resulting in the generation of increased levels of episomal vector copies in transduced cells. The lentivirus can only become integrated into the DNA in the presence of a DSB by NHEJ-repair activity [122]. Following irradiation or doxorubicin treatment of different cell types transduced with IDLV, integration of IDLV can be followed by PGK-driven EGFP expression using flow cytometry. Irradiated and doxorubicin-treated cells showed significant increase in both γH2AX foci numbers as well as in the frequency of EGFP$^+$ cells compared to non-treated control cells, demonstrating that the different cancer therapy strategies induce high levels of transient DSB, at which IDLV becomes stably incorporated into the genome. By comparing the frequency of EGFP$^+$ cells amongst various irradiated cancer cell lines, it became apparent that IDLV-mediated trapping at radiation-induced DSB is not restricted to a specific cell type, and that the different cell types show diverse IDLV integration frequencies, probably reflecting different NHEJ-repair activities.

IDLV integration was reported to be mediated by NHEJ-repair activity [122]. In order to show NHEJ-mediated IDLV integration, the LTR deletion frequency was analyzed, since NHEJ-repair results in deletion of DNA bases from free DNA ends. This analysis revealed that up to 53% of the integrated vectors carried single bases to large (20bp) deletions. Compared to the LTR deletion frequency after integrase-mediated DNA integration (2.2%), this demonstrates that IDLV incorporation is mediated by error-prone Non Homologous End Joining (NHEJ) repair activity. Which of the two sub-pathways of NHEJ (classical or alternative) is responsible for integration or what percentage each pathway contributes to IDLV integration could not be determined from the LTR deletion frequencies, since both repair pathways frequently result in DNA deletions [19]. Moreover, it is likely that other, homology-directed DNA repair pathways contribute to the integration of IDLV into the genome, since impairing NHEJ-repair significantly increased the frequency of irradiated EGFP$^+$ cells and thus IDLV integration events. A possible explanation is that the inhibition of the NHEJ-repair pathways shifted the ratio between the different repair mechanisms towards repair via homologous recombination. In support of this theory are several studies showing that cells deficient for classical NHEJ-repair show increased frequencies of alternative NHEJ [123] and homology-directed repair activities [124]. Moreover, studies conducted in flies where DNA ligase IV, a gene that is required for ligation of free DNA ends in NHEJ [125], was knocked out significantly increased homologous recombination rates [126]. In order to determine which pathway compensates for NHEJ deficiency, the LTR deletion frequency in NHDF-A with impaired classical or alternative NHEJ-repair was analyzed. Since the LTR deletion frequency was comparable to the deletion frequency in NHDF-A with active NHEJ-repair, it is likely that NHEJ inhibition was incomplete at the chosen NHEJ inhibitor concentrations or that inhibition of NHEJ could have delayed integration until NHEJ-repair is fully active again.

IDLV-based DSB trapping is quantitative and enables comparison of the number of DSB sites between different radiation sources. Quantitative-RT-PCR analysis revealed that irradiation of A549 cells with a single 4Gy photon beam led to the incorporation of one IDLV copy in every two cells. High LET irradiation with a single 4Gy proton beam induced the same level of DSB sites. The background IDLV integration frequency was calculated to be 0.034 per cell, i.e. about 1 IDLV integration event in 30 cells. These numbers may be an underestimation of the actual number of IDLV-labeled DSB sites, since only a fraction of the lentiviral vector was amplified and quantified. Moreover, IDLV episomes may also be subject to radiation damage and non-LTR mediated integration [122]. Indeed, non-LTR-mediated integration events have been reported for integrase-deficient lentiviral vectors [127], and integrated IDLV showed gross reorganizations including deletions, duplications and

reorganizations of the vector sequence in human colon carcinoma [128]. LAM-PCR analysis of integrated HIV-1 IDLV in Hela and fetal mouse embryo SC-1 cells has revealed both complete and deleted LTR-end sequences [129]. Thus, in order to analyze the integrity of IDLV and to check for non-LTR-mediated IDLV integration, vector target enrichment (SureSelect, Agilent) for HIV-1-derived vectors was performed. In this approach, lentivirus-specific RNA probes bind to the integration vector, are enriched on streptavidin-coated beads and prepared for pyrosequencing. The obtained results indicate that the majority of IDLV was intact, and only a few reads showed that the vector continuity was disrupted. However, the number of reads mapping to the lentiviral sequence was low (1%) so that additional optimizations in the protocol to increase the yield are required. Since implementation of the bioinformatical analysis platform in our group is still ongoing, the target enrichment system will be useful to identify both LTR- and non LTR-mediated integration events and possibly detect structural genomic variations at DSB sites.

Taken together, these results demonstrate that IDLV is a versatile tool to capture and quantify cancer therapy-induced and naturally-occurring DSB genome-wide in various cell types, which enables the characterization of DSB at previously uncharted resolution and detail in living cells.

4.3   Identification of radiation-induced DSB sites

In order to better understand, how DSB sites are distributed in radiation-surviving cell populations and how differential DSB repair activity affects cell fate decision, the radiation-induced DSB sites marked by IDLV were amplified by LAM-PCR. In total, more than 20,000 unique DSB sites in cancer cell lines and primary fibroblasts that have survived irradiation were obtained. The analysis revealed that the distribution of DSB and NHEJ-repair activity sites is not biased towards specific chromosomes or chromosomal compartments. Moreover, radiation-as well as naturally-occurring DSB sites were not enriched in gene coding regions, in transcriptional control regions located within 10kb upstream of the transcription start site and in CpG rich regions. Similar results were obtained in doxorubicin-treated NHDF-A. Thus, DSB sites are labeled in all genomic regions, can be efficiently amplified and are not enriched in gene-coding regions.

The obtained results also indicate that IDLV integration into the genome is unlikely to be integrase (IN)-mediated, since the IDLV integration pattern does not correlate to the integration pattern of integrase-competent lentiviruses (ICLV). ICLV preferentially integrate into gene-coding regions, at CpG islands and on gene rich chromosomes 16, 17, 19 and 22 [61, 62]. Besides, it was shown that IDLV becomes integrated in pre-existing DSB as observed in U2OS cells [130]. The 'CA' dinucleotide that is located at both LTR ends and that is removed during IN-mediated integration was frequently detected in the IDLV vector-genome amplicons. These results demonstrate that lentiviral integration preferences mediated by integrase play no quantitative role in IDLV capture and that IDLV integration events underly mechanism other than integration by ICLV. Hence, the genomic mapping of DSB reflects the repertoire and spatial distribution of acquired genomic damage in irradiated or doxorubicin-treated cell populations.

4.3.1   Transcriptional activity before irradiation does not influence DSB site distribution

Actively transcribed genomic regions have been described to be prone to DSB induction by exogenous agents [120, 131, 132], since transcription induces an open chromatin state that exposes the genomic DNA. Moreover, actively transcribed genomic regions colocalize with sites of recurrent, early replicating DNA lesions, termed

early replication fragile sites (ERFS) [133]. During simultaneous DNA replication and transcription, conflicts frequently arise between the DNA replication and transcriptional machinery, which can result in stalling and destabilization of the transcriptional and replication machinery as well as in DSB induction. Such DNA damages arising in ERFS are a potential source for genomic instability and tumorigenesis [133]. Therefore, the mRNA transcriptional activity in the genome of A549 cancer cells around the DSB site was analyzed. More than 200,000 mRNA microarray transcripts were correlated with 4,655 DSB sites in 500kb genomic windows around the DSB site. No enrichment of radiation-induced or naturally-occurring DSB sites with expressed or not-expressed mRNA signatures was observed. DSB sites were not enriched in highly active mRNA transcriptional signatures as well as in transcribed genomic regions that contain more than four active mRNA signatures. These results indicate that the transcriptional activity at the time point of irradiation does not seem to influence DSB site distribution.

However, it cannot be excluded that transcriptional changes that occur in response to irradiation may influence DSB site distribution. In particular, genomic regions encoding genes that are involved in DDR and become actively transcribed upon DSB induction may be more susceptible to DSB induction by reactive oxygen species or delayed genomic instability [44]. In order to further delineate the influence of transcription on DSB distribution and NHEJ-repair activity, whole transcriptome sequencing and LAM-PCR during expansion of IDLV-transduced, radiation-surviving cells would need to be performed. This analysis would help to unravel the role of both coding and non-coding RNA on DSB repair activity and induction of delayed genomic instability.

4.3.2 DSB site distribution is non-random with respect to the genome accessibility

The basic component of the chromatin is the nucleosome, which is composed of histone octamers, surrounded by 147bp genomic DNA and separated by 10-15bp nucleosome-free DNA. The accessibility of the DNA and the structure of the genome in the nucleus are regulated by modifications of the histones at their lysine residues. Each histone modification is associated with activation or repression of transcription and with various genomic features such as promoters, transcribed regions, enhancer and insulators [134]. Large-scale mapping of histone variations such as by the ENCODE project have emerged as powerful tools to characterize the structure of the human genome in different cell types [134]. At promoters, histone modifications contribute to fine-tuning of gene expression levels. Inactive promoters are characterized by tri-methylation of lysine 27 at histone 3 (H3K27me3), H3K9me3 and methylated DNA. Prior to transcription, the DNA promoter sequence becomes more accessible for transcription factor binding due to methylation of lysine 4 at histone 3 (H3K4me2, H3K4me3), acetylation and incorporation of histone H2A.Z into the nucleosomes at the transcription start site (TSS). At gene bodies, histones discriminate between active and inactive conformations [134]. Actively transcribed genes are marked by H3K36me3, whereas inactive genes are associated with H3K9me2/3. Moreover, histone marks correlating with active enhancer regions are associated with H3K4me1/2, H3K27ac and H2A.Z, whereas inactive enhancers are enriched for H3K9me2/3.

The DSB repair activity was reported to be unevenly distributed in the genome with regions devoid of repair proteins [14, 131]. Phosphorylated H2AX occurred preferentially in euchromatic and actively transcribed genomic regions after X-ray irradiation, and in both eu- and heterochromatin at times when these regions were undergoing replication. DSB of endogenous origin were reported to be frequently enriched in sub-telomers, reflecting replication- and transcription-mediated stress during cell division. These results have led to the hypothesis that heterochromatin blocks H2AX phosphorylation due to its high compaction and limited accessibility. However, it was shown that radiation-induced DSB in heterochromatin can be efficiently detected

by the DNA repair machinery. Upon recognition, the local chromatin structure is decondensed at the DSB site in order to enable DSB repair to occur [16]. These findings demonstrate that the initially randomly-induced DSB sites are processed differently in heterochromatin and euchromatin. Since these studies were performed at early time points after irradiation, no information about the distribution of DSB sites with respect to chromatin status in radiation-surviving cells is available. In particular, the genomic structure and histone modifications at sites of radiation-induced DSB and genomic instability have not been characterized yet. Thus, in order to analyze the mechanisms of genomic instability in radiation-surviving cell populations, the DSB sites were correlated with chromatin accessibility and histone modification data available from the ENCODE project.

Open chromatin regions are hypersensitive to DNaseI digestion, because the chromatin has lost its condensed confirmation. These regions are generally associated with transcriptional activity, since an open chromatin state is required for transcription factor binding. Radiation-induced DSB in A549, but not radiation-induced DSB in NHDF-A were significantly enriched in open chromatin. DSB in doxorubicin-treated cells also showed a significant enrichment in DHS. The enrichment of induced DSB sites was further supported by analyzing the histone modifications at the DSB sites. Radiation-induced DSB sites in the two cell types A549 and NHDF-A were significantly enriched in euchromatin, at border regions of eu- to heterochromatin and depleted in heterochromatin. The same distribution pattern was also observed for doxorubicin-induced DSB sites. By further analyzing the histone modifications associated to radiation- and doxorubicin-induced DSB sites, enrichment and depletion of specific histone modifications became apparent. The heterochromatic histone modification H3K9me3 was depleted at radiation- and doxorubicin-induced as well as naturally-occurring DSB sites. H3K9me3 marks constitutive heterochromatin present at centromers and telomers, which are important functional elements for the maintenance of genome integrity. DSB in heterochromatin were shown to be more detrimental to cell viability than DSB located in euchromatin, since heterochromatin has functions in chromosome segregation, nuclear organization and regulation of gene expression. If DSB accumulate in heterochromatin and cannot be efficiently repaired, genomic instability and loss-of-function can occur. Thus, cells that display increased levels of DSB in heterochromatin (HC) are more likely to undergo apoptosis [135], which is reflected in the reduced DSB frequency in HC of radiation- and doxorubicin-surviving cell populations. The DSB frequency in heterochromatic regions marked by H3K9me3 also gives some insight into the spatial nuclear localization of the DSB sites. Since laminar-associated domains generally interact with genomic regions marked H3K9me3 [136, 137], it is likely that both radiation- and doxorubicin-induced DSB sites are depleted at the nuclear surface and accumulate more frequently close to the center of the nucleus.

In irradiated and doxorubicin-treated as well as untreated NHDF-A the histone modifications H3K27ac, H3K36me3 and H4K20me1 were enriched at identified DSB sites. H3K36me3 is associated with actively transcribed genes, and H4K20 becomes methylated at active gene promoters [138]. In addition, H3K4me1 and me2 were enriched exclusively at DSB sites identified in doxorubicin-treated NHDF-A, representing active promoter and enhancer regions. Interestingly, some DSB were marked by specific combinations of two histone modifications. In particular, the frequency of H3K4me3 and H3K27me3 was enriched at radiation-induced as well as naturally-occurring DSB. Developmental genes that are silenced in embryonic stem cells and become transcribed during differentiation are marked by H3K27me3 and H3K4me3 at their promoters [91]. Their exact role in differentiation and development remains unknown. However, a recent study showed that these bivalent histone marks define gene sets in ovarian cancer that distinguish malignant, tumour-sustaining and chemo-resistant ovarian tumor cells, indicating that upon damage deregulated promoter activity can induce cancer formation [139]. Furthermore, exclusively in irradiated- and doxorubicin-treated cells, the histone modification H3K4me3 was enriched in combination with several other histone marks including H4K20me1, H3K36me3 and

H3K27ac at DSB sites. These histone combinations are only present at active promoters and transcription start sites in the genome, indicating that cancer-therapy induced DSB in surviving cell populations preferentially occur or are select for open chromatin associated with gene regulatory regions. Thus, IDLV-based DSB trapping identifies functional genomic sites in which error-prone DSB repair in promoters, genes and at TSS may have consequences for gene regulation and expression activity by directly affecting inter- and intrachromosomal interactions.

4.3.3   Genes involved in specific cellular processes are enriched for induced DSB

To date, it has not been evaluated, whether cell survival after therapy is dependent on the repair of specific gene classes. Thus, Ingenuity Pathway Analysis (IPA) was performed to investigate, whether specific gene classes that have been damaged and repaired were enriched in survivor cells. Captured DSB in irradiated cells were frequently and significantly enriched in genes that have specific functions related to the control and regulation of gene expression, cell cycle, cell death and survival, which are relevant factors in oncogenesis. Similarly, doxorubicin-induced DSB sites also preferentially occurred in genes with the same functions. This indicates that error-prone repair of cancer-therapy induced DSB could deregulate the expression of these genes. To which extent these genes were inactivated or activated could not be determined. However, mutations in these gene classes could potentially lead to increased DNA damage induction as well as lack of cell cycle control and defects in apoptosis induction, which in turn could cause further destabilization of the genome [140]. Indeed, several human disorders associated with defects in genome maintenance show enhanced cancer susceptibility (Table 1).

Moreover, both radiation- and doxorubicin-damaged and repaired genes share common upstream regulators and target sequences of specific microRNA. The down- and up-regulation of microRNAs in response to radiotherapy has been shown to be associated with the DNA damage response following the exposure to radiation. Particular micro (mi)RNAs have been identified to regulate DSB-labeled genes in irradiated NHDF-A. Some of these miRNA were reported to increase cellular survival following irradiation and altered radiosensitivity of cell lines and primary cells [141, 142]. The miRNA-185 which was exclusively identified as an upstream regulator in irradiated NHDF-A is a known inhibitor of ATR transcription, and repression of this miRNA conferred radio-resistance upon X-ray exposure [143]. Moreover, miRNA-145 promotes apoptosis and modulates radiation sensitivity of cancer cells. Among its various target mRNA is the p53-negative regulator MDM2, which is upregulated in epithelial cancers, thereby inhibiting p53-induced apoptosis [144]. Taken together, these results indicate that cell survival may be dependent on the repair of specific gene classes and genes that are regulated by stress-response regulating proteins and miRNA.

4.3.4   Clonality of DSB site distribution over time

The identification of radiation-induced DSB sites in different passages after irradiation enables the tracking of DSB site kinetics. Because IDLV-tagged DSB sites uniquely mark individual irradiated cells, the dynamics in DSB distribution reflect the changes in the clonality of irradiated cells [79].

Analysis of the ten strongest DSB sites in irradiated NHDF-A at three different passages revealed that no dominant DSB clone arose in the course of cultivation. Instead, the results point to a more polyclonal cell population in which a few clones rise, disappear and are quickly followed by other clones. In order to find out what factors influence appearance and disappearance of radiation-induced DSB sites, the genomic location and

the histone modifications at radiation-induced DSB sites detected in all three passages were analyzed. Neither the position of the DSB sites around the transcription start site nor a specific histone modification was associated with rise and decline of top ten DSB sites. This was also true for NHDF-A cells with impaired NHEJ-repair activity. Thus, it is more likely that these cells carry additional, non-labeled DSB sites and genomic aberrations that could induce changes in gene regulation and expression [140]. To determine which structural genomic changes occur and how these influence gene expression more deeply, single cell clones need to be isolated and the genome structure as well as transcriptional activity analyzed. In particular, analyzing the frequency of translocations is highly important since these events are threatening genomic integrity which contributes to both haematopoietic and solid tumors. Recent studies indicate that these translocations are not randomly distributed, but appear to be influenced by the nuclear architecture and the transcriptional activity at the translocation site. However, these studies are non-conclusive for the whole genome, since these data stem from studies conducted in cells at a single locus. Thus, analyzing the mechanisms underlying the occurrence of genomic rearrangements would help to understand their functional consequences.

4.3.5   Radiation-induced and repaired DSB sites cluster in hotspots in specific genomic regions

Analyzing the distribution of DSB sites in the genome revealed that radiation-induced and naturally-occurring DSB sites clustered in hotspots. The most intense genomic hotspot of DSB in irradiated and expanded cell populations harbored up to 131 DSB in a 200kb genomic region and overlapped with genomically unstable regions in non-irradiated cells. These regions, referred to as overlapping DSB hotspots, were located inside or in close proximity to genes which have been frequently associated with tumor suppressor activity (*FAT1*, *CUX1*, *NEGR1*), have anti-angiogenic function (*OPTC*), are frequently mutated in various types of cancers (*TRIM36*, *FCGR1B*, *TCL6, SLCO2B1*) and other genetic disorders (*ARRB1, NFASC, TRIM29, MOCS2*). Since the identified overlapping hotspots were associated with cancer-relevant genes, this suggests that these regions are intrinsically unstable and that cancer-therapy induced genomic instability increases the mutation level in these regions. Damage and error-prone repair may result in genome defects initiating or promoting radioresistance and carcinogenesis. This becomes apparent by the findings affecting the *FAT1* locus. The tumor suppressor gene is located in a genomic region frequently lost in numerous types of human cancer and is mutated at high prevalence, leading to aberrant Wnt activation [93].

The majority of the identified DSB hotspots in irradiated and doxorubicin-treated NHDF-A, however, were specific for radiation- or doxorubicin-related DNA damage, respectively, and not detectable in untreated control cells. The top five radiation- and doxorubicin-related DSB hotspots in NHDF-A were located inside or in proximity to genes shown to be involved in the maintenance of genome stability (*ERI1*, *BIRC6*), regulation of radiosensitivity (*RAC1*, *UTP18*, *TOP1*) and DNA repair (*ELF4*, *ZYX*, *UTP18*, *XRCC5*). Since these therapy-related hotspots do not contain naturally-occurring DSB, this indicates that these genomic regions are not naturally unstable, but that induced genomic stress destabilizes the genome in these locations and causes accumulation of DSB. Similar to the overlapping DSB hotspots, the identified genes may potentially be involved in cancer therapy-induced carcinogenesis and tumor-resistance mechanisms which influence survival of sub-lethally treated cells in clinical therapy as exemplified by the *ERI1* and *RAC1* genes. ERI1 is an exonuclease responsible for the degradation of histone mRNA upon completion of DNA replication. Failure to degrade histone mRNA was shown to result in loss of genomic fragments and whole chromosomes [109, 110]. However, no functional study has reported that mutations or loss of *ERI1* can drive cancer initiation or progression. In contrast, the Rho GTPase Rac1 has been implicated in head and neck squamous cell carcinomas (HNSCC) insensitivity to

radiotherapy, which frequently results in tumor recurrence. In radiation-resistant cells increased expression, activity and nuclear translocation of the GTPase protein was observed, and chemical inhibition of Rac1 expression and activity resulted in significant improvement of HNSCC sensitivity to ionizing radiation [145]. In addition to this association, strong and weak genomic enhancers annotated by the ENCODE project in direct proximity (<10kb) to the radiation- and doxorubicin-related DSB hotspots were identified. This may indicate that a potential alteration of long-range inter- and intra-chromosomal enhancer function and of the nuclear architecture can be expected upon error-prone repair by disrupting the base composition at genomic enhancer regions. Interestingly, radiation- and doxorubicin-related DSB hotspots occurred in different genomic regions, which did not overlap. This indicates that the susceptibility of genomic regions and the pattern of instabilities may depend on the mechanism of DSB induction, i.e. irradiation or interference with DNA replication. Hence, analyzing the DSB hotspot distribution might therefore be used to predict tumor response to specific cancer treatment. Furthermore, it could be speculated that not only the location of DSB hotspots, but also the frequency and type of mutations induced by irradiation is different from doxorubicin. Indeed, it has been shown that cells that have been irradiated or treated with doxorubicin at the equivalent cytotoxic concentration show more translocations after irradiation than doxorubicin-treated cells. Accordingly, doxorubicin-treated cells would be expected to show lower translocations frequency [146].

These results clearly indicate that IDLV-based DSB trapping identifies known and new genomic regions that are destabilized in radiation-surviving cell populations and may lead to induction of genome destabilization and radiation-resistance. Hence, in comparison to previous DSB detection technologies, IDLV-based DSB trapping enables the identification of functional DSB sites that influence genome stability and cell survival. In order to evaluate experimentally whether the identified genomic regions and genes influence genome stability and cell survival, the rapid developments in the field of genome editing can accomplish this task. Endonucleases such as TALENs or CRISPR/Cas9 systems can be designed to target basically every sequence in the genome and induce DSB, which are subsequently repaired by the cellular DNA repair machinery. These site-specific nucleases can be used to disrupt the integrity of the DNA at the identified DSB hotspots. The effect of these deletions on genome integrity, apoptosis induction, cell survival and cellular growth under radiation- or doxorubicin-induced stress can be analyzed.

Overlapping as well as radiation- and doxorubicin-related DSB hotspots were associated to euchromatic and border regions of eu- and heterochromatin. These genomic regions are not located at the periphery of the nucleus, but closer to the center. In genomic border regions, an increased frequency of persisting DNA damage sites was reported [147]. Border regions of eu- and heterochromatin are at high risk for non-allelic homologous recombination during meiosis when numerous programmed DSB are induced for recombination. Furthermore, heterochromatin boundary regions are hotspots for *de novo* kinetochore formation and contribute to centromere identity [148]. The kinetochore is a large protein complex associated with centromeric DNA that is responsible for mediating the segregation of sister chromatids during mitosis via interactions with the mitotic spindle. Thus, the protection of border regions appears to be a general requirement to prevent formation of dysfunctional centromeres and self-induced genome rearrangements that are associated with several genomic diseases [149, 150].

4.4   Methods for *in situ* labeling of induced DSB sites

Since IDLV-based DSB labeling is feasible to study the distribution of induced and repaired DSB sites in cancer-therapy surviving cell populations, it is reasonable to assume that it might also enable the analysis of DNA repair kinetics within the first few hours after irradiation. LAM-PCR of IDLV-transduced and irradiated NHDF-A at 1h, 4h, 8h, and 24h after irradiation did not identify more DSB sites than in non-irradiated control cells. A possible reason for the low number of induced DSB sites is the presence of high numbers of non-integrated, episomal IDLV vector copies. Thereby, an increased amplification of LTR-LTR and LTR-vector backbone sequences during LAM-PCR that overlie the amplicons of the vector-genome junctions is the result. Thus, the number of labeled DSB sites is assumed to be higher. In order to reduce the number of episomal amplicons, several strategies can be applied. One possibility would be to cut the isolated genomic DNA with restriction enzymes that specifically target the IDLV vector sequence. The resulting smaller vector fragments could subsequently be separated from the genomic DNA, which is then subjected to LAM-PCR and deep sequencing. Furthermore, to increase the sensitivity during pyrosequencing, the number of reads for each sample can be increased. Here, each sample was sequenced at a depth of 800,000 reads. With the continued improvements in next-generation sequencing technologies, additional increases in the sequencing depth should be possible, allowing deeper sequencing of vector-genome junctions in the presence of vector amplicons.

In order to study the kinetics of DSB induction and repair after irradiation, two new approaches were developed and evaluated concerning their efficiency to label DSB sites *in situ*. The first approach was based on genome-wide labeling of DSB sites by the template-independent DNA polymerase terminal deoxynucleotidyl transferase (Tdt), which adds nucleotides to free 3' OH DNA termini in a random fashion [151]. In the second approach, free DNA ends were labeled by DNA ligases, which attach a specific DNA linker sequence to the DSB site. This method was termed LAM-DST. In order to show that direct DSB labeling by Tdt and DNA ligases combined with LAM-PCR can identify DSB, DSB sites were introduced into the genome by a CCR5-specific ZFN. LAM-PCR of Tdt-labeled samples did not identify any of the reported ZFN on- or off-target sites. Instead, unrelated genomic DNA sequences with an increased content of adenine and thymidine were amplified during LAM-PCR. Since the PCR primer used during exponential amplification has a polyA tail, this implies that the primer binds unspecifically to DNA sequences that have a similar sequence to the Tdt-mediated polyU tail. It may be possible that the transfection efficiency or ZFN activity was too low to induce efficient DSB levels for labeling and amplification. Therefore, Tdt-based DSB labeling might still be useful to study on-/off-target ZFN activity and DSB repair kinetics despite unspecific DNA amplification. However, in order to use Tdt-mediated DSB labeling for genome-wide DSB detection, the rate of unspecific DNA amplification needs to be greatly reduced and parameters for separating "true" DSB sites from unspecifically amplified sequences selected.

PCR on the CCR5 locus in samples treated with CCR ZFN and labeled by T4 DNA ligase or circLigase demonstrated that the ZFN was active and that the DNA ligases efficiently labeled the free DNA ends. However, subsequent LAM-PCR on circLigase-labeled samples did not identify the CCR5 ZFN on- and off-target loci. This suggests that LAM-DST can detect induced DSB, but that additional adaptations for the reduction of artifically induced DSB during processing and random amplification are required to enable genome-wide DSB site identification. A potential source for artifically-induced, non-ZFN-mediated DSB is the utilization of excessive concentrations of paraformaldehyde (PFA). PFA is used for the stabilization of the chromatin-DNA complexes, and, if used at too high concentrations and for too long, it can result in the hydrolysis of the phosphodiester bonds in the DNA [152]. This creates free 3' OH termini that are labeled by the DNA ligases. Secondly, linear double-stranded DNA molecules are susceptible to destabilization, which can create free 3' OH sites for ligation of 5' phosphorylated DNA. This could be a potential source for the formation of DNA linker concatemers. Thus, in order to increase

the stability of DNA linker molecules against nuclease degradation, hairpin forming DNA sequences and phosphothioate residues at the 3' DNA termini can be introduced [153].

Even though CCR5 ZFN-induced DSB sites could not be efficiently identified by LAM-DST and LAM-PCR, a direct *in situ* ligation approach is a valuable tool to study DSB site distribution as well as designer nuclease activity sites. In contrast to classical methods used for studying DNA repair, *in situ* ligation offers several advantages. First, it enables direct marking of DSB sites at single-nucleotide resolution, the analysis of the kinetics of DSB repair in different cell cycle phases and the influence of the genomic context on DSB distribution. Moreover, *in situ* DSB labeling is organism independent, thereby allowing the identification of DSB sites in multiple cell types and under various conditions. Further, it can help to explore the basis of genomic instability by combining data from *in situ* labeling with histone and transcriptional data sets [154]. The results gained from such studies could prove highly valuable for analyzing the mechanisms underlying cancer therapy resistance. Furthermore, analyzing the global distribution of endonuclease-induced DSB sites such as CRISPRS/Cas9 and TALEN is important in order to assess and improve the safety of novel gene therapy treatments. Recent designer nuclease off-target studies have shown the enormous potential of DSB detection sequencing methods, in line with the developments described here in this work [155-157].

5. SUPPLEMENT

5.1 Supplementary Figures
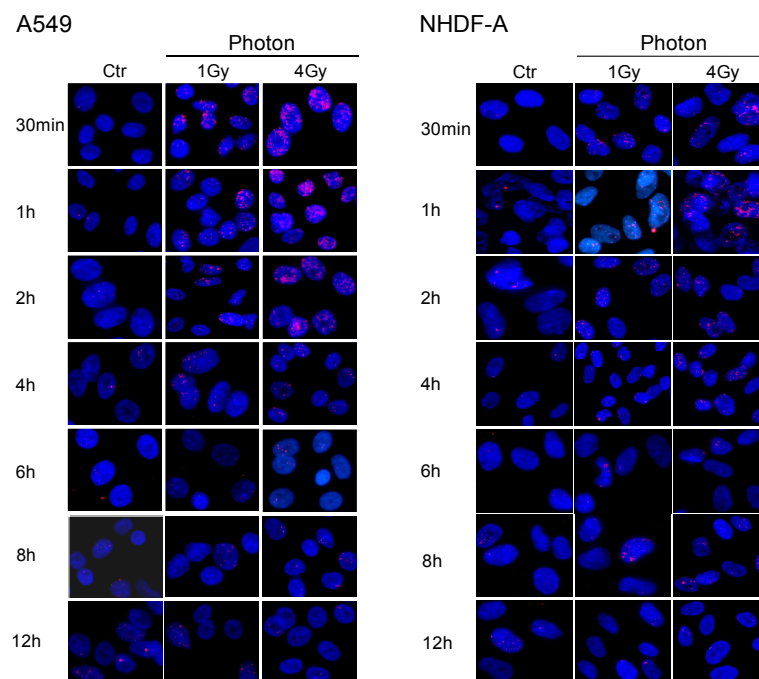
Supplementary Figure 1: Immunostaining of γH2AX foci in irradiated A549 and NHDF-A
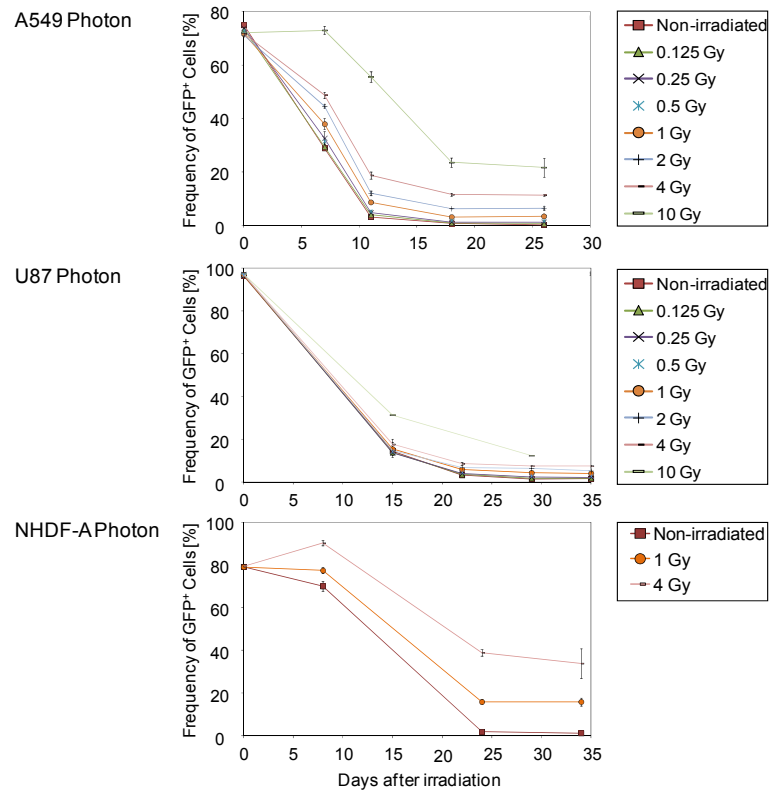
Supplementary Figure 2: Frequency of EGFP[+] cells transduced with IDLV during cultivation after irradiation.
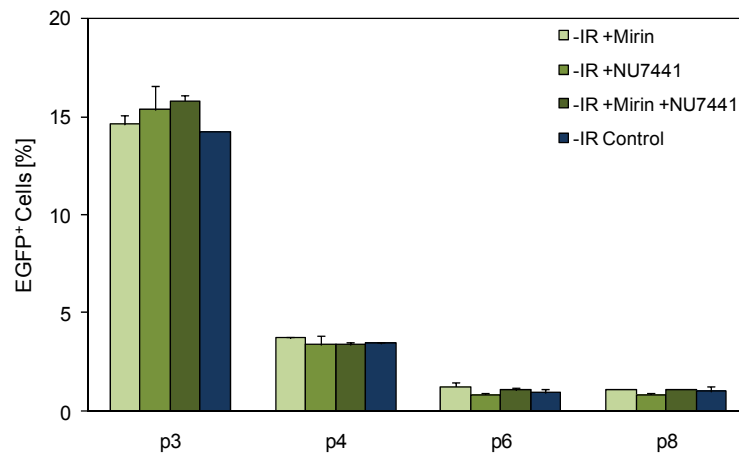
Supplementary Figure 3: Frequency of EGFP[+] and IDLV-transduced NHDF-A with impaired NHEJ-repair activity during expansion in cell culture. These cells were not exposed to irradiation. EGFP values were determined by flow cytometry.Control: active NHEJ-repair activity; p: passage
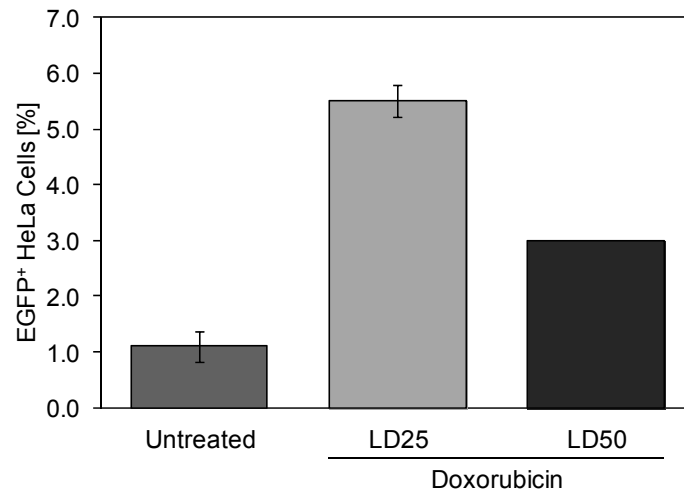
Supplementary Figure 4: Frequency of EGFP+, IDLV-transduced HeLa cells treated with LD25 and LD50 doxorubicin after 21 days in culture. EGFP values were determined by flow cytometry.
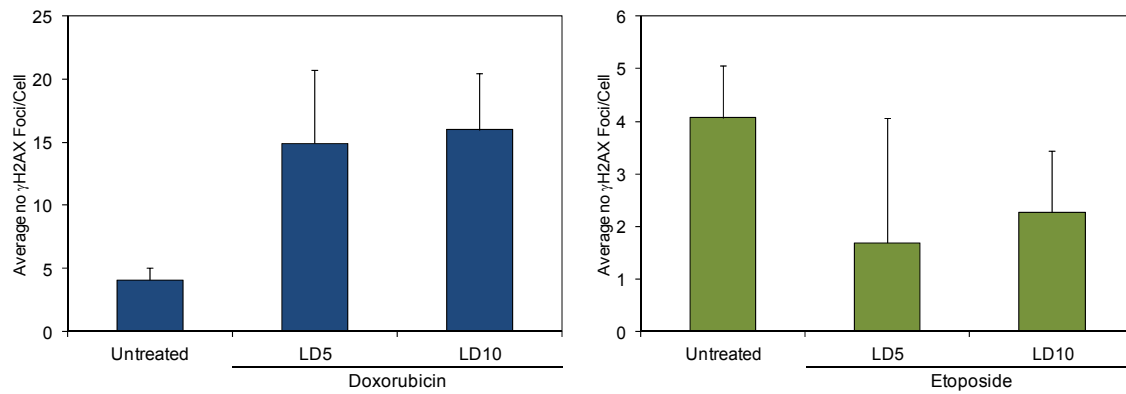
Supplementary Figure 5: Average γH2AX foci number in NHDF-A treated continusly with doxorubicin and etoposide for 43h. Error bars represent the standard deviation of the mean of the γH2AX foci number per cell.
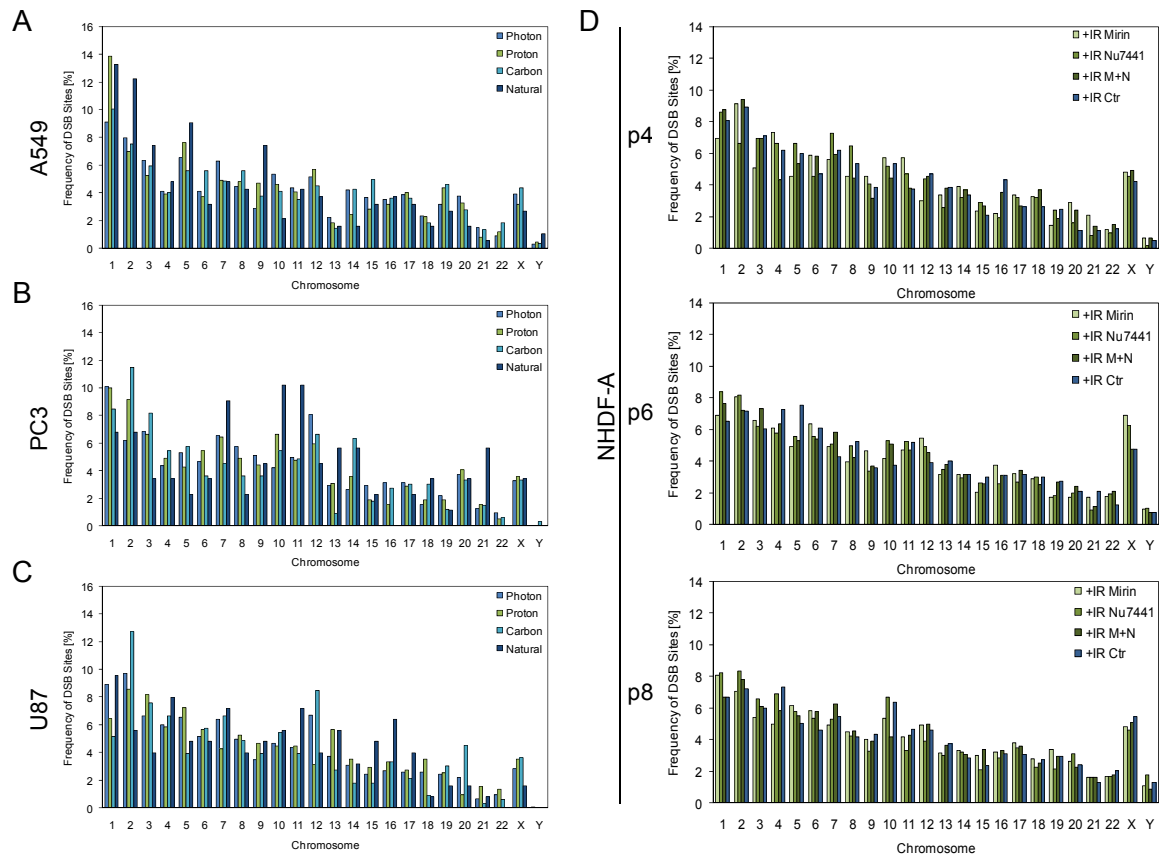
Supplementary Figure 6: Chromosomal distribution of radiation-induced DSB sites in A549 (A), PC3 (B), U87 (C) and NHDF-A with impaired NHEJ-repair activity (D).
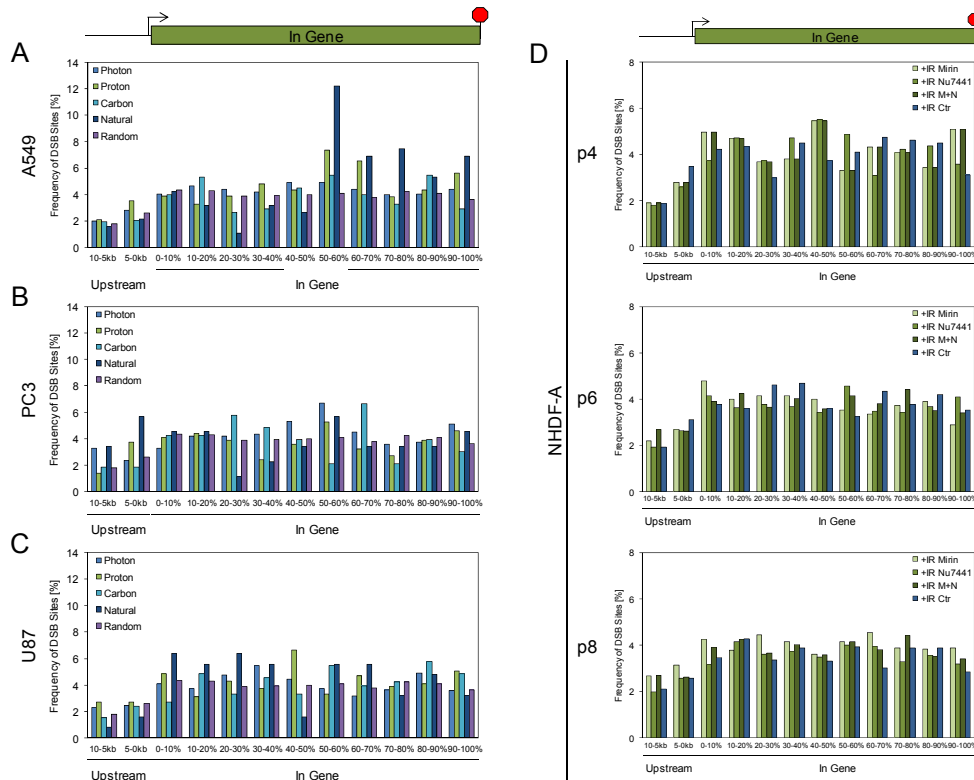
Supplementary Figure 7: Distribution of radiation-induced DSB sites in genes and at the transcription start site (TSS). A549 (A), PC3 (B), U87 (C), NHDF-A with impaired NHEJ-repair activity (D).

Supplementary Figure 8: Analysis of the genome sequence at doxorubicin-induced DSB sites. Since the first two nucleotides frequently map to the LTR sequence of the IDLV, these nucleotides were not included for sequence analysis. In the random DSB data set (bottom), the first two nucleotides were included, since these sequences were generated *in silico* and do not contain LTR sequences.
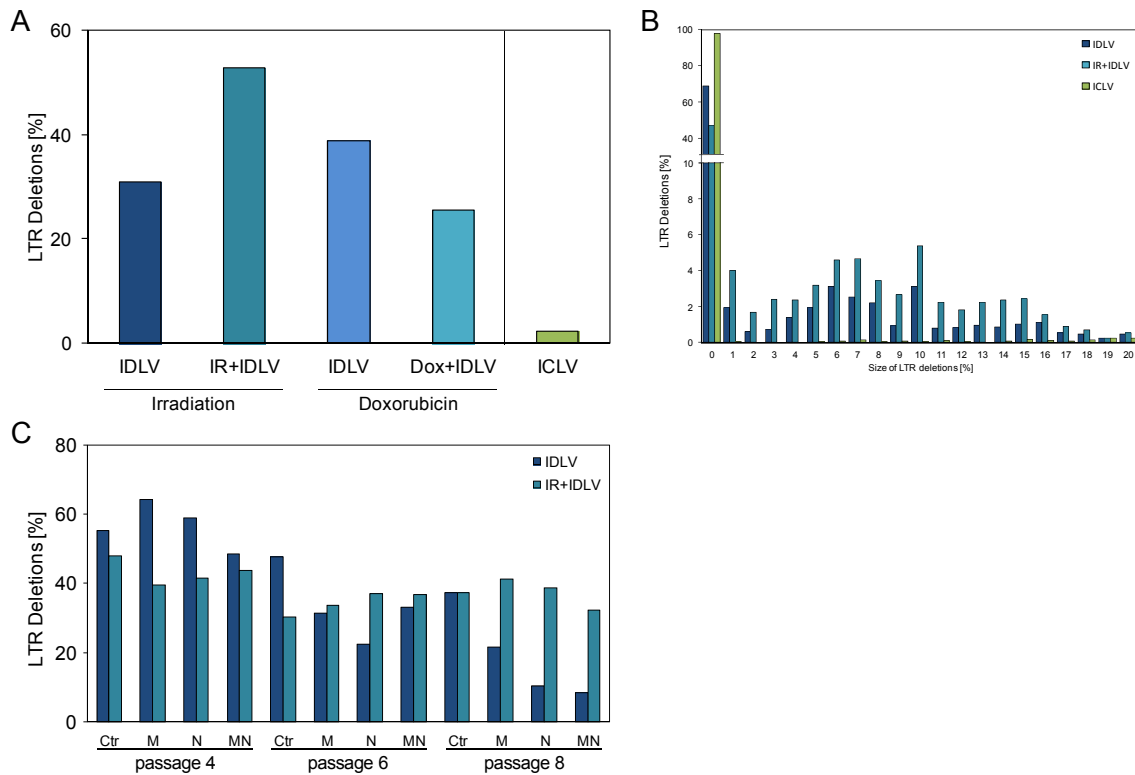
Supplementary Figure 9: Frequency of LTR deletions in integrated IDLV vector copies compared to an integrase-competent lentivirus (ICLV).(A) Total frequency of IDLV and ICLV with deleted LTR sequences. (B) Size of LTR deletions in IDLV and ICLV. (C) LTR deletion frequency in NHDF-A with impaired NHEJ-repair activity
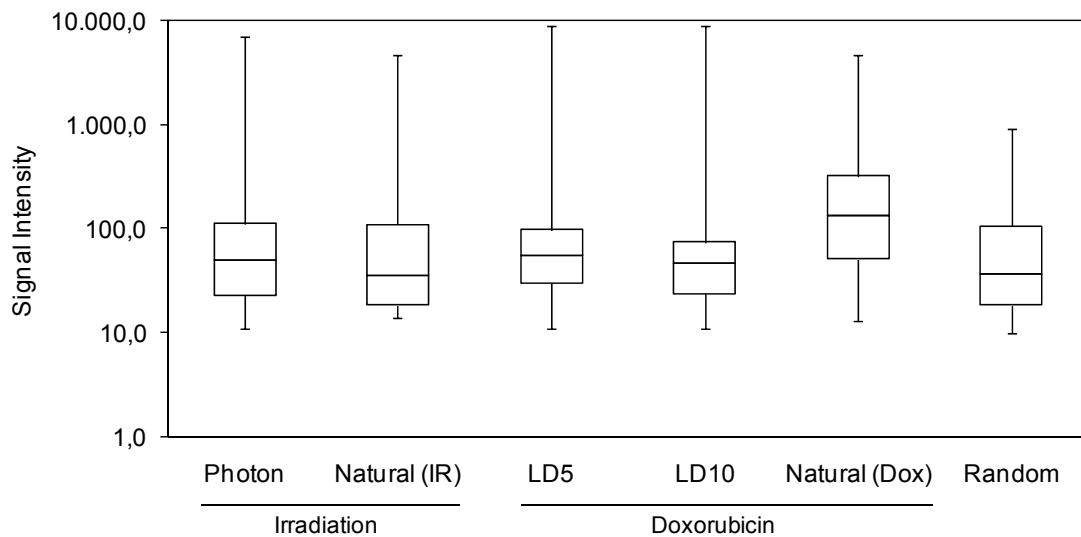
Supplementary Figure 10: The signal intensity of the DNaseI HS does not influence distribution of radiation- and doxorubicin-induced DSB sites. Boxes represent upper and lower quantiles, and middle line indicates the mean.

A549

NHDF-A
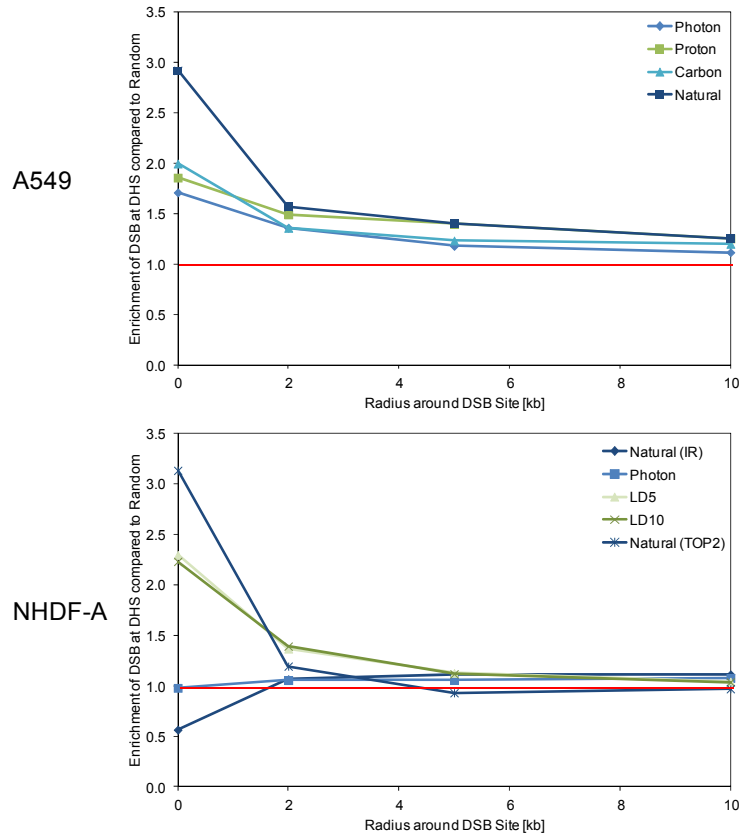
Supplementary Figure 11: DSB sites do not cluster in proximity to DnaseI HS, but are enriched inside DnaseI HS. The red line indicates the frequency for the random DSB data set.
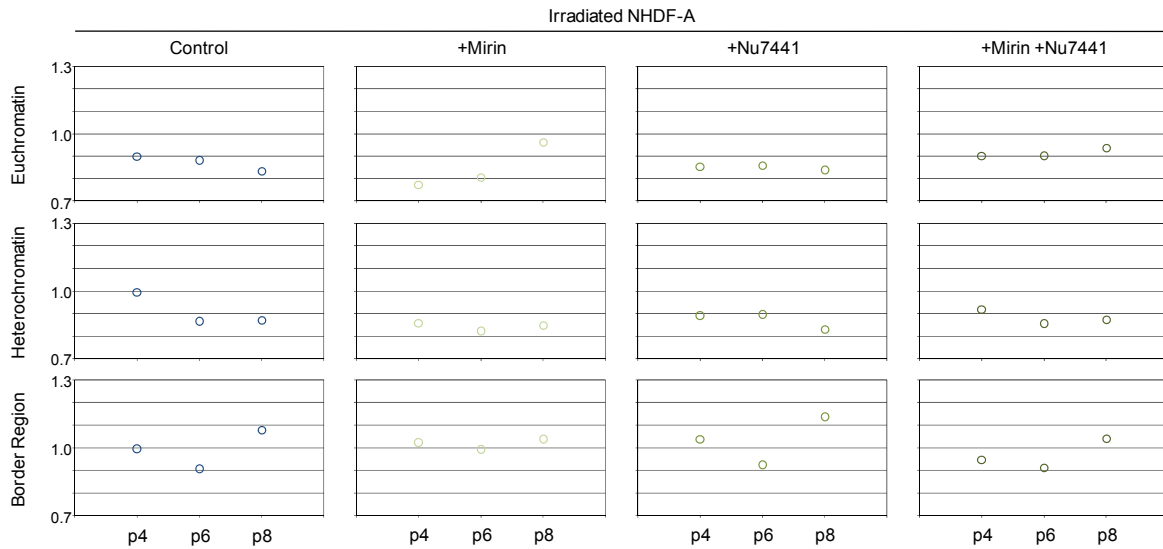
Supplementary Figure 12: Distribution of radiation-induced DSB sites in eu- and heterochromatin of NHEJ-impaired NHDF-A. Similar to NHEJ-proficient cells, the DSB frequency increases in border regions and euchromatin in passage 4, 6, and 8, whereas the DSB frequency in heterochromatin decreases during cultivation. Numbers on y-axis represent enrichment values of DSB frequency compared to random DSB frequency.
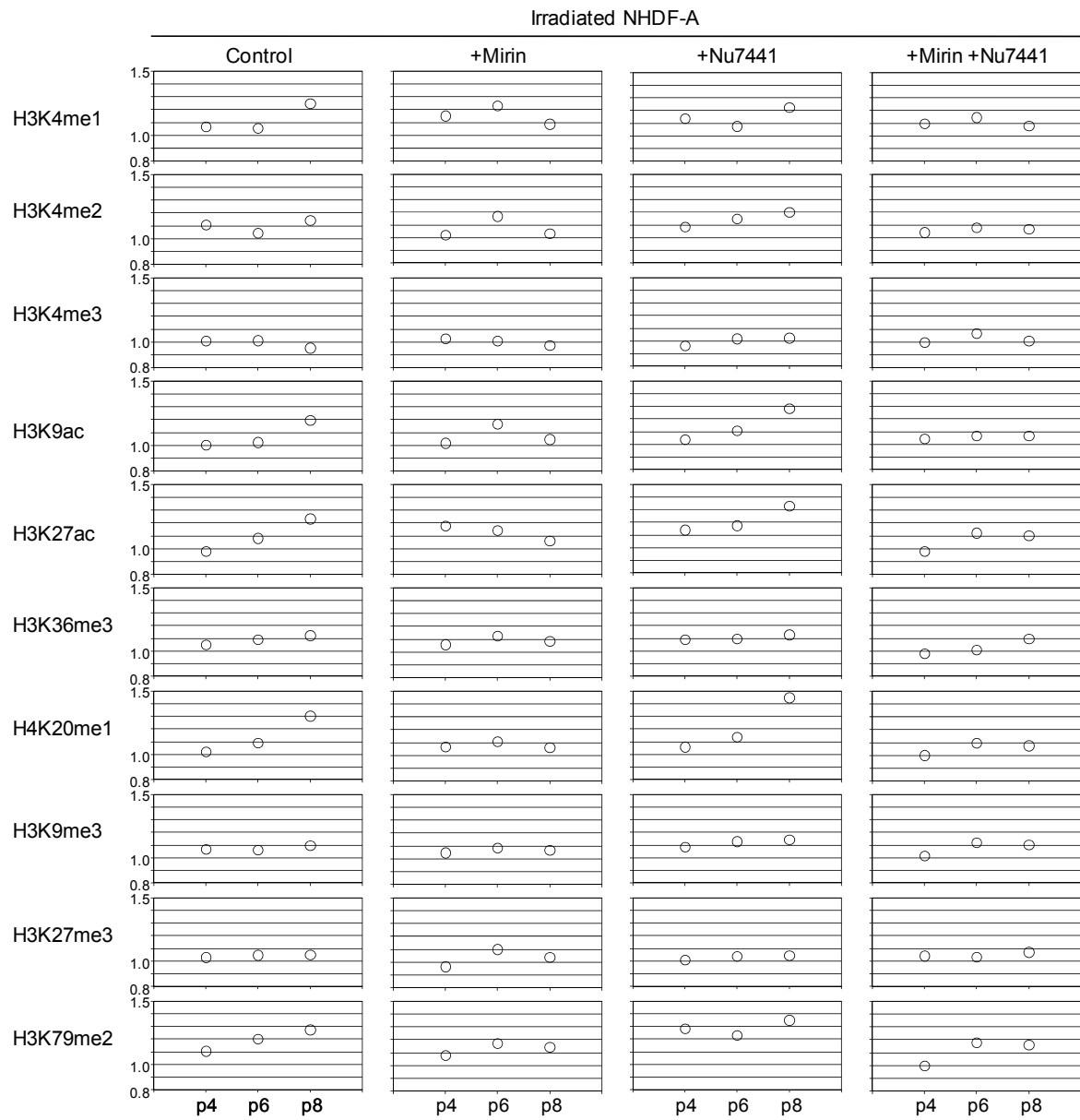
Supplementary Figure 13: Histone modifications associated to radiation-induced DSB sites in NHEJ-impaired NHDF-A in three passages. Numbers on y-axis represent enrichment values of histone frequency compared to histone frequency at randomly-distributed DSB.

+IR Control

p4

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3	H3K79me2
H3K4me1		1,09	1,21	1,15	1,15	1,13	1,11		1,33	1,14
H3K4me2			1,03	1,32	1,29	1,18	1,15	1,21	1,49	1,20
H3K4me3				1,02	1,05	0,97	1,00	1,03	1,25	0,97
H3K9ac					0,99	1,02	1,09	1,04	1,38	1,04
H3K27ac						1,02	1,12	1,25	1,55	1,09
H3K36me3							1,02	1,38	1,33	1,22
H4K20me1								1,13	1,97	1,64
H3K9me3									1,15	2,06
H3K27me3										1,26
H3K79me2										

p6

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3	H3K79me2
H3K4me1		1,04	1,08	1,22	1,22	1,13	1,11		1,25	1,21
H3K4me2			1,04	1,32	1,30	1,25	1,27	1,15	1,44	1,29
H3K4me3				1,01	1,11	0,96	1,03	1,04	1,26	0,99
H3K9ac					1,08	1,05	1,14	1,07	1,46	1,09
H3K27ac						1,08	1,18	1,21	1,63	1,18
H3K36me3							1,11	1,42	1,53	1,26
H4K20me1								1,03	1,96	1,85
H3K9me3									1,18	2,00
H3K27me3										1,23
H3K79me2										

p8

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3	H3K79me2
H3K4me1		1,23	1,19	1,48	1,47	1,38	1,55	1,32	1,47	1,47
H3K4me2			1,04	1,50	1,50	1,28	1,54	1,24	1,46	1,43
H3K4me3				1,13	1,16	0,96	1,25	0,95	1,10	1,09
H3K9ac					1,27	1,19	1,45	1,28	1,66	1,30
H3K27ac						1,22	1,55	1,40	1,93	1,42
H3K36me3							1,43	1,46	1,41
H4K20me1								1,41	2,44	2,15
H3K9me3									1,11	2,53
H3K27me3										1,48
H3K79me2										

+IR +Mirin

p4

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3	H3K79me2
H3K4me1		1,12	1,15	1,29	1,39	1,30	1,24	1,08	1,26	1,21
H3K4me2			0,78	0,96	1,06	0,98	0,93	0,89	1,06	0,93
H3K4me3				0,73	0,86	0,86	0,83	0,77	0,87	0,75
H3K9ac					0,96	1,12	0,99	1,34	1,07
H3K27ac						0,82	1,15	1,10	1,00
H3K36me3							0,86	1,31
H4K20me1								0,75	1,37
H3K9me3									0,84
H3K27me3										
H3K79me2										

p6

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3	H3K79me2
H3K4me1		1,21	1,16	1,33	1,33	1,30	1,29	1,29	1,49	1,35
H3K4me2			0,95	1,31	1,24	1,20	1,18	1,13	1,48	1,20
H3K4me3				0,94	0,98	0,90	0,90	0,86	1,14	0,91
H3K9ac					1,06	1,10	1,09	1,12	1,48	1,12
H3K27ac						1,08	1,12	1,25	1,78	1,16
H3K36me3							1,00	1,33	1,47	1,20
H4K20me1								1,10	1,95	1,62
H3K9me3									1,09	2,11
H3K27me3										1,29
H3K79me2										

p8

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3	H3K79me2
H3K4me1		1,07	1,07	1,24	1,23	1,19	1,16	1,19	1,26	1,21
H3K4me2			0,80	0,98	1,04	0,97	0,97	0,97	1,06	0,96
H3K4me3				0,73	0,81	0,74	0,81	0,77	0,87	0,78
H3K9ac					0,82	0,86	0,89	0,95	1,12	0,88
H3K27ac						0,88	0,96	1,02	1,23	0,96
H3K36me3							0,85	1,14	1,11	1,01
H4K20me1								0,92	1,55	1,36
H3K9me3									0,85	1,74
H3K27me3										0,94
H3K79me2										

+IR +Nu7441

p4

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3	H3K79me2
H3K4me1		1,11	1,10	1,25	1,30	1,23	1,24	1,17	1,34	1,29
H3K4me2			0,88	1,18	1,22	1,07	1,11	1,17	1,17	1,14
H3K4me3				0,85	0,94	0,78	0,87	0,90	1,00	0,89
H3K9ac					0,98	0,91	0,98	0,99	1,28	1,01
H3K27ac						0,97	1,10	1,24	1,41	1,15
H3K36me3							0,95	1,23	1,07	1,09
H4K20me1								1,04	1,72	1,59
H3K9me3									0,86	1,87
H3K27me3										0,95
H3K79me2										

p6

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3	H3K79me2
H3K4me1		1,10	1,07	1,29	1,31	1,26	1,26	1,26	1,31	1,29
H3K4me2			0,80	0,98	1,04	0,93	0,93	0,98	0,98	0,94
H3K4me3				0,72	0,77	0,70	0,71	0,69	0,76	0,73
H3K9ac					0,82	0,86	0,87	0,87	1,13	0,86
H3K27ac						0,92	0,99	1,04	1,16	1,01
H3K36me3							0,81	1,01	0,93	0,95
H4K20me1								0,96	1,41	1,36
H3K9me3									0,81	1,66
H3K27me3										0,79
H3K79me2										

p8

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3	H3K79me2
H3K4me1		1,22	1,17	1,50	1,46	1,38	1,55	1,36	1,50	1,45
H3K4me2			0,79	0,99	0,99	0,97	0,95	0,95	1,06	0,92
H3K4me3				0,68	0,73	0,68	0,69	0,69	0,65
H3K9ac					0,89	0,95	0,97	1,00	1,30	0,90
H3K27ac						0,98	1,06	1,12	1,40	1,01
H3K36me3							0,79	0,99	0,95	0,89
H4K20me1								1,15	1,92	1,56
H3K9me3									0,70	1,30
H3K27me3										0,72
H3K79me2										

+IR +Mirin +Nu7441

p4

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3	H3K79me2
H3K4me1		1,05	1,05	1,22	1,15	1,10	1,13	1,00	1,28	1,10
H3K4me2			0,97	1,31	1,21	1,11	1,18	1,04	1,37	1,15
H3K4me3				1,02	1,06	0,91	1,00	0,91	1,11	0,98
H3K9ac					1,03	0,96	1,02	1,02	1,36	0,99
H3K27ac						1,01	1,12	1,06	1,50	1,11
H3K36me3							0,99	1,21	1,19	1,15
H4K20me1								1,02	1,79	1,55
H3K9me3									0,99	1,78
H3K27me3										1,04
H3K79me2										

p6

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3	H3K79me2
H3K4me1		1,12	1,17	1,30	1,30	1,23	1,27	1,30	1,36	1,23
H3K4me2			1,05	1,32	1,36	1,17	1,24	1,27	1,44	1,26
H3K4me3				1,06	1,16	0,95	1,04	1,15	1,17	1,03
H3K9ac					1,10	1,05	1,17	1,22	1,52	1,12
H3K27ac						1,08	1,25	1,39	1,83	1,25
H3K36me3							1,05	1,46	1,43	1,27
H4K20me1								1,25	2,06	1,80
H3K9me3									1,25	2,34
H3K27me3										1,42
H3K79me2										

p8

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3	H3K79me2
H3K4me1		1,09	1,08	1,25	1,25	1,23	1,24	1,18	1,31	1,19
H3K4me2			1,03	1,34	1,34	1,25	1,23	1,21	1,46	1,23
H3K4me3				1,05	1,11	0,99	1,04	1,01	1,25	0,99
H3K9ac					1,09	1,13	1,21	1,10	1,45	1,11
H3K27ac						1,13	1,21	1,25	1,66	1,23
H3K36me3							1,11	1,43	1,41	1,31
H4K20me1								1,22	2,03	1,76
H3K9me3									1,12	2,22
H3K27me3										1,30
H3K79me2										

Supplementary Figure 14: Radiation-induced and naturally-occurring DSB in NHEJ-impaired NHDF-A cells show association to genomic regions marked by H3K27me3 in combination with H3K4me3, H3K9ac, H3K27ac and H3K36me3. Numbers indicate the enrichment in the histone-per-DSB frequency compared to the random DSB data set. IR: irradiated, p: passage

Natural

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3
H3K4me1		1,06	1,14	1,07	1,14	1,11	1,09	1,03	1,29
H3K4me2			1,01	1,06	1,11	1,02	0,99	1,12	1,24
H3K4me3				0,99	1,02	1,07	1,13	1,47	1,58
H3K9ac					1,06	1,10	1,18	1,36	1,71
H3K27ac						1,13	1,13	1,34	1,66
H3K36me3							1,03	1,14	1,36
H4K20me1								1,03	1,18
H3K9me3									1,03

Photon

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3
H3K4me1		1,05	1,06	1,07	1,07	1,05	1,06	1,10	1,12
H3K4me2			1,01	1,03	1,04	1,07	1,12	1,09	1,11
H3K4me3				1,02	1,00	1,04	1,14	1,16	1,25
H3K9ac					1,04	1,04	1,11	1,15	1,36
H3K27ac						1,06	1,10	1,21	1,38
H3K36me3							1,06	1,15	1,21
H4K20me1								1,07	1,15
H3K9me3									1,01

Proton

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3
H3K4me1		1,05	1,10	1,11	1,10	1,01	1,03	1,09	1,17
H3K4me2			1,08	1,10	1,09	1,06	0,96	1,04	1,04
H3K4me3				1,04	1,03	1,06	1,11	1,12	1,21
H3K9ac					1,05	1,10	1,07	1,18	1,30
H3K27ac						1,06	1,04	1,18	1,39
H3K36me3							1,01	1,16	1,11
H4K20me1								1,06	1,14
H3K9me3									1,00

Carbon

	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K27ac	H3K36me3	H4K20me1	H3K9me3	H3K27me3
H3K4me1		1,02	1,03	1,02	1,06	1,15	1,13	1,13	1,11
H3K4me2			1,03	0,99	1,03	1,10	1,10	1,17	1,16
H3K4me3				0,99	1,07	1,12	1,12	1,28	1,28
H3K9ac					1,07	1,04	1,11	1,22	1,70
H3K27ac						1,10	1,19	1,34	1,86
H3K36me3							1,04	1,16	1,25
H4K20me1								1,03	1,21
H3K9me3									1,03

Supplementary Figure 15: Radiation-induced and naturally-occurring DSB in A549 cells show association to genomic regions marked by H3K27me3 in combination with H3K4me3, H3K9ac, H3K27ac and H3K36me3. Numbers indicate the enrichment in the histone per DSB frequency compared to the random DSB data set. The radiation source is indicated on the left.
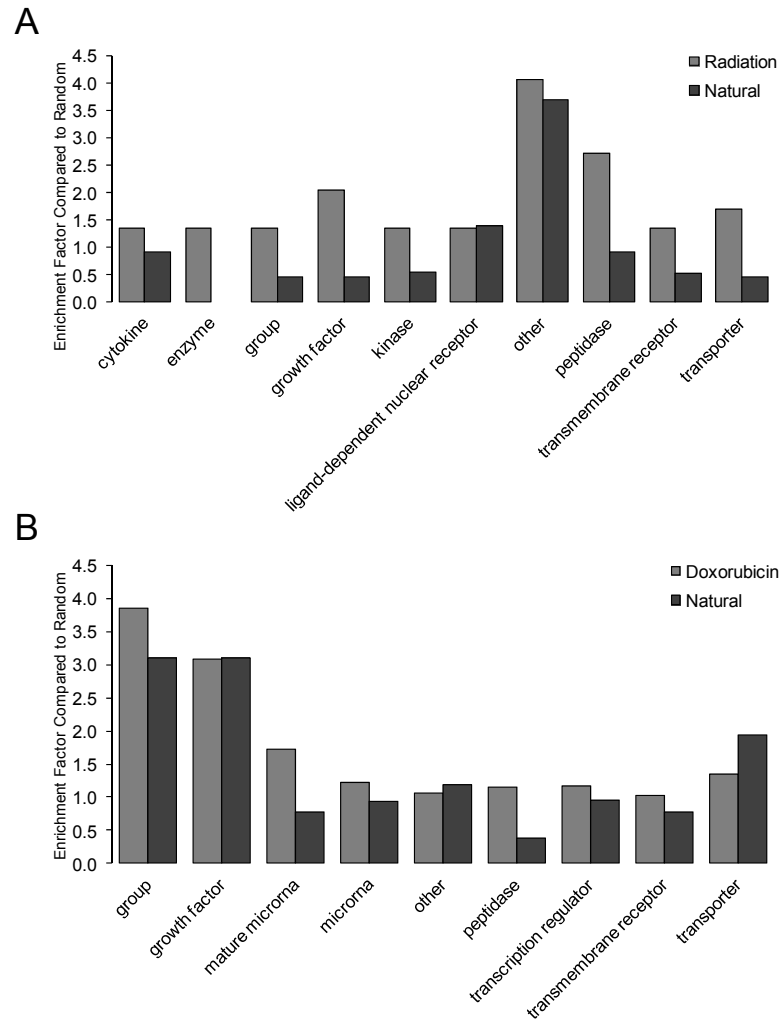
Supplementary Figure 16: Upstream regulators of IDLV-marked genes in (A) irradiated and (B) doxorubicin-treated NHDF-A
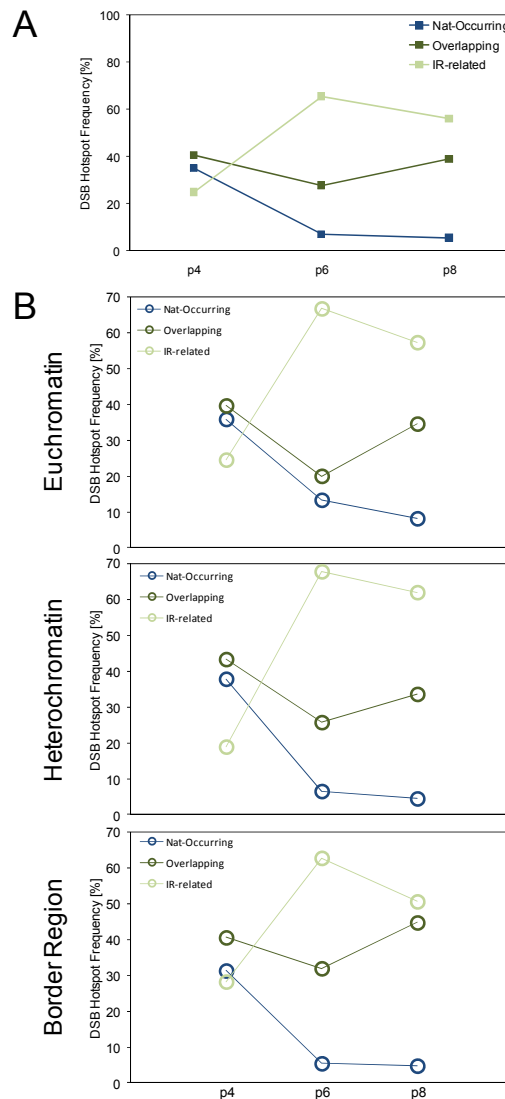
Supplementary Figure 17: Kinetics and distribution of DSB hotspots with respect to chromatin regions. (A) The frequency of naturally-occurring DSB hotspots decreased during expansion, whereas the frequency of radiation-related DSB hotspots increased. (B) Naturally-occurring DSB hotspots decreased in euchromatin, heterochromatin and in border regions. In contreast, the frequency of overlapping DBS hotspots remained constant and radiation-related hotspots increased.

Position in genome downstream of vector-genome junction

Nucleotide	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A	79	61	29	23	32	37	35	36	35	34	34	33	32	34	33	35	35	32	32	32
T	7	19	54	57	33	33	31	30	30	31	25	29	30	34	31	30	29	32	27	28
G	7	11	8	9	15	15	16	14	19	21	23	18	19	18	21	19	20	20	24	19
C	7	9	8	11	20	15	18	19	16	14	17	19	18	14	15	17	16	15	17	20

Supplementary Figure 18: Nucleotide frequency at indicated positions downstream of vector-genome junction. The first two nucleotides (position 1 and 2) have a high frequency of the adenosine, which is most likely to belong to the polyU sequence added to the free DNA ends by Tdt enzyme. Thus, the first nucleotide of the genomic sequence is assumed to be at position 3.
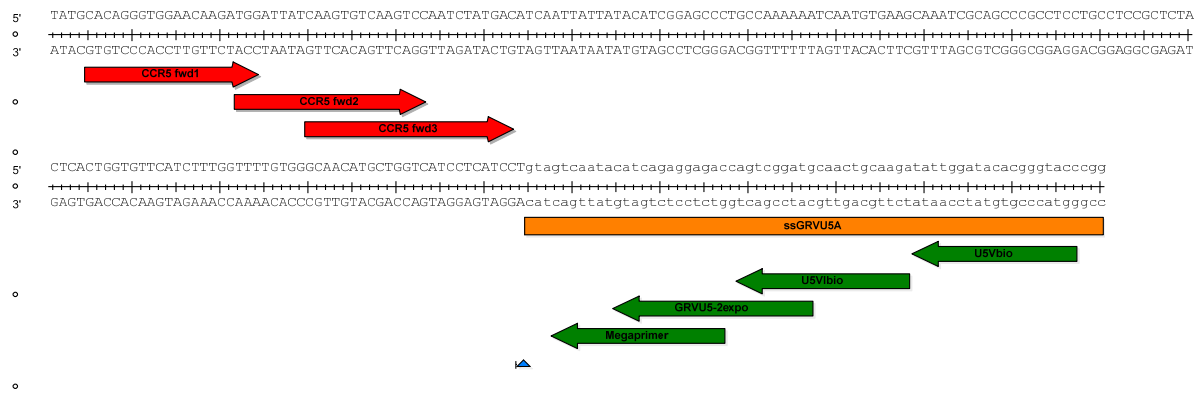
Supplementary Figure 19: CCR5 genomic locus with ligated ssGRVU5A DNA linker. Primers used during locus-specific PCR and LAM-PCR are indicated. The blue triangle marks the CCR5 ZFN target site.

5.2   SupplementaryTables

Supplementary Table 1: Number of unique DSB in cancer cell lines and primary human fibroblasts.

Cell Types	DSB Induction	Source	Unique DSB sites
A549	Irradiation	Photon	2540
	Irradiation	Proton	1102
	Irradiation	Carbon	825
		Natural	188
PC3	Irradiation	Photon	642
	Irradiation	Proton	588
	Irradiation	Carbon	330
		Natural	88
U87	Irradiation	Photon	1261
	Irradiation	Proton	510
	Irradiation	Carbon	330
		Natural	124
NHDF-A	Irradiation	Photon	6959
		Natural	3887
NHDF-A	Chemotherapy	LD5 Doxorubicin	4003
	Chemotherapy	LD10 Doxorubicin	3725
		Natural	263
Random	-		4742

Cell Type	DSB Induction	NHEJ-repair Blocking Agent	Passage	Unique DSB sites
NHDF-A	Irradiation	+Mirin	p4	764
	Irradiation	+Mirin	p6	1459
	Irradiation	+Mirin	p8	3929
	Irradiation	+Nu7441	p4	616
	Irradiation	+Nu7441	p6	1927
	Irradiation	+Nu7441	p8	3131
	Irradiation	+Mirin +Nu7441	p4	787
	Irradiation	+Mirin +Nu7441	p6	2151
	Irradiation	+Mirin +Nu7441	p8	4243
	Irradiation	Control	p4	803
	Irradiation	Control	p6	1194
	Irradiation	Control	p8	1759
		+Mirin	p4	975
		+Mirin	p6	488
		+Mirin	p8	1036
		+Nu7441	p4	897
		+Nu7441	p6	326
		+Nu7441	p8	1710
		+Mirin +Nu7441	p4	735
		+Mirin +Nu7441	p6	393
		+Mirin +Nu7441	p8	667
		Control	p4	610
		Control	p6	453
		Control	p8	329

Supplementary Table 2: Frequency of DSB sites in RefSeq genes, at transcription start site and CpG islands. Integration site data for integrase-competent lentiviruses (ICLV) were obtained from [88]. No values for DSB frequency in NHDF-A at CpG islands were available, since the function was deactivated in HISAP.

| | | | DSB Frequency [%] | | |
| | | | In RefSeq | At TSS | At CpG islands |
Cell Types	DSB Induction	Source	Genes	[+10kb]	[≤1kb]
A549	Irradiation	Photon	43.9	5.2	2.8
	Irradiation	Proton	47.9	6.1	5.9
	Irradiation	Carbon	40.5	4.2	3.0
	Natural		53.2	3.7	5.9
PC3	Irradiation	Photon	45.0	5.6	1.8
	Irradiation	Proton	38.1	5.1	4.1
	Irradiation	Carbon	40.9	3.6	1.7
	Natural		36.4	9.1	3.4
U87	Irradiation	Photon	41.5	4.8	2.1
	Irradiation	Proton	43.8	5.5	3.3
	Irradiation	Carbon	43.0	3.9	1.8
	Natural		48.0	2.4	10.5
NHDF-A	Irradiation	Photon	51.7	5.7	-
	Natural		51.0	6.9	-
NHDF-A	Chemotherapy	LD5 Doxorubicin	65.5	4.6	-
	Chemotherapy	LD10 Doxorubicin	65.0	4.4	-
	Natural		45.7	4.1	-
Random	-	-	40.2	4.7	2.6
ICLV	-	-	74.6	2.2	1.6

Supplementary Table 3: Frequency of lentiviral 'GT' dinucleotide at vector-genome junctions. #: Number of sequences with GT dinucleotide at vector-genome junction.

Cell Types	DSB induction	Source	# 'GT' Dinucleotide	'GT' Frequency	p-value (Fisher's exact test)
NHDF-A	Irradiation	Photon	2090	30.0	1.14E-246
		Natural	2364	60.8	0.00E+00
NHDF-A	Chemotherapy	LD5 Doxorubicin	2560	64.0	0.00E+00
	Chemotherapy	LD10 Doxorubicin	2394	64.3	0.00E+00
		Natural	132	50.2	4.80E-168
Random			230	4.9	

Supplementary Table 4: Frequency of DSB sites located inside DNaseI HS.

Cell Types	DSB induction	Source	DSB Frequency	Enrichment compared to Random	p-value (Fisher's exact test)
A549	Irradiation	Photon	3.43	1.71	3.12E-04
	Irradiation	Proton	3.90	1.95	2.82E-04
	Irradiation	Carbon	4.00	2.00	6.09E-04
		Natural	5.85	2.92	5.99E-04
Random (A549)			2.00	-	-
NHDF-A	Irradiation	Photon	1.87	1.00	9.73E-01
		Natural	1.03	0.55	1.24E-03
NHDF-A	Chemotherapy	LD5	4.42	2.35	5.45E-12
	Chemotherapy	LD10	4.40	2.34	1.36E-11
		Natural	5.24	2.79	1.62E-04
Random (NHDF-A)			1.88	-	-

Cell Type	DSB induction	Passage	NHEJ-repair Blocking Agent	DSB Frequency	Enrichment compared to Random	p-value (Fisher's exact test)
NHDF-A	Irradiation	p4	+Mirin	1.44	0.77	4.01E-01
	Irradiation	p6	+Mirin	1.23	0.66	9.90E-02
	Irradiation	p8	+Mirin	1.86	0.99	9.49E-01
	Irradiation	p4	+Nu7441	2.92	1.56	8.10E-02
	Irradiation	p6	+Nu7441	1.40	0.75	1.78E-01
	Irradiation	p8	+Nu7441	1.31	0.70	5.30E-02
	Irradiation	p4	+Mirin +Nu7441	1.78	0.95	8.51E-01
	Irradiation	p6	+Mirin +Nu7441	1.16	0.62	3.10E-01
	Irradiation	p8	+Mirin +Nu7441	1.46	0.78	1.26E-01
	Irradiation	p4	Control	2.86	1.53	6.60E-02
	Irradiation	p6	Control	1.34	0.71	2.08E-01
	Irradiation	p8	Control	1.42	0.76	2.14E-01
		p4	+Mirin	1.85	0.98	9.49E-01
		p6	+Mirin	1.64	0.87	7.11E-01
		p8	+Mirin	2.41	1.29	2.61E-01
		p4	+Nu7441	2.01	1.07	7.94E-01
		p6	+Nu7441	1.84	0.98	9.63E-01
NHDF-A		p8	+Nu7441	2.22	1.18	3.78E-01
		p4	+Mirin +Nu7441	1.50	0.80	4.74E-01
		p6	+Mirin +Nu7441	2.29	1.22	5.65E-01
		p8	+Mirin +Nu7441	1.65	0.88	6.83E-01
		p4	Control	2.13	1.14	6.65E-01
		p6	Control	4.19	2.23	1.00E-03
		p8	Control	1.13	1.13	7.47E-01
Random (NHDF-A)				1.88	-	-

Supplementary Table 5: Frequency of DSB sites located in border regions, eu- and heterochromatin

Cell Type	DSB Induction	Source	Euchromatin	Heterochromatin	Border Regions
	Irradiation	Photon	28.1	29.8	34.0
	Irradiation	Proton	31.9	26.7	35.9
A549	Irradiation	Carbon	26.3	32.0	32.2
		Natural	29.8	22.3	41.0
	Random (A549)		24.2	34.8	30.5
	Irradiation	Photon	18.0	29.4	37.3
		Natural	18.0	26.4	33.4
NHDF-A	Chemotherapy	LD5 Doxorubicin	17.1	9.7	47.6
	Chemotherapy	LD10 Doxorubicin	16.1	10.6	47.2
		Natural	11.2	29.6	39.0
	Random (NHDF-A)		15.2	34.3	34.0

Cell Type	DSB induction	Passage	NHEJ-repair Blocking Agent	Euchromatin	Heterochromatin	Border Region
	Irradiation	p4	+Mirin	11.8	29.5	34.7
	Irradiation	p6	+Mirin	12.3	28.2	33.7
	Irradiation	p8	+Mirin	14.7	29.1	35.2
	Irradiation	p4	+Nu7441	13.0	30.5	35.2
	Irradiation	p6	+Nu7441	13.1	30.7	31.3
NHDF-A	Irradiation	p8	+Nu7441	12.7	28.5	38.5
	Irradiation	p4	+Mirin +Nu7441	13.7	31.5	32.1
	Irradiation	p6	+Mirin +Nu7441	13.7	29.3	31.0
	Irradiation	p8	+Mirin +Nu7441	14.3	29.9	35.3
	Irradiation	p4	Control	13.7	34.1	33.7
	Irradiation	p6	Control	13.4	29.6	30.7
	Irradiation	p8	Control	12.7	29.8	36.6
		p4	+Mirin	13.8	35.8	32.5
		p6	+Mirin	16.4	28.1	35.5
		p8	+Mirin	16.8	32.1	33.2
		p4	+Nu7441	11.7	37.1	34.3
		p6	+Nu7441	13.8	28.2	40.2
NHDF-A		p8	+Nu7441	18.2	26.4	38.5
		p4	+Mirin +Nu7441	12.5	31.7	35.5
		p6	+Mirin +Nu7441	15.5	28.0	36.9
		p8	+Mirin +Nu7441	13.3	31.9	32.8
		p4	Control	13.4	33.3	34.6
		p6	Control	14.6	36.4	32.9
		p8	Control	11.2	35.0	31.3
	Random (NHDF-A)			15.2	34.3	34.0

Supplementary Table 6: Frequency and enrichment of DSB at CTCF Binding sites compared to random DSB data set.

Cell Types	DSB induction	Source	DSB Frequency in CTCF BS	Enrichment compared to Random	p-value (Fisher's exact test)
A549	Irradiation	Photon	10.8	1.4	2.7E-04
	Irradiation	Proton	10.7	1.3	8.6E-03
	Irradiation	Carbon	13.6	1.7	3.0E-06
	Natural		15.4	1.9	1.3E-03
	Random (A549)		8.0	-	
NHDF-A	Irradiation	Photon	19.5	1.1	3.5E-06
		Natural	19.6	1.1	2.9E-03
NHDF-A	Chemotherapy	LD5 Doxorubicin	32.1	1.8	1.6E-57
	Chemotherapy	LD10 Doxorubicin	32.6	1.9	7.2E-59
		Natural	24.0	1.4	7.0E-03
	Random (NHDF-A)		17.4	-	-

Cell Type	DSB induction	Passage	NHEJ-repair Blocking Agent	DSB Frequency in CTCF BS	Enrichment compared to Random	p-value (Fisher's exact test)
NHDF-A	Irradiation	p4	+Mirin	17.0	1.0	7.9E-01
	Irradiation	p6	+Mirin	17.0	1.0	7.1E-01
	Irradiation	p8	+Mirin	18.0	1.0	4.8E-01
	Irradiation	p4	+Nu7441	18.3	1.1	5.7E-01
	Irradiation	p6	+Nu7441	18.2	1.0	4.7E-01
	Irradiation	p8	+Nu7441	22.9	1.3	2.5E-09
	Irradiation	p4	+Mirin +Nu7441	14.7	0.9	6.4E-02
	Irradiation	p6	+Mirin +Nu7441	15.7	0.9	8.0E-02
	Irradiation	p8	+Mirin +Nu7441	17.7	1.0	7.3E-01
	Irradiation	p4	Control	18.8	1.1	3.4E-01
	Irradiation	p6	Control	15.9	0.9	2.2E-01
	Irradiation	p8	Control	18.9	1.1	1.6E-01
		p4	+Mirin	17.2	1.0	8.9E-01
		p6	+Mirin	20.9	1.2	5.5E-02
		p8	+Mirin	16.7	1.0	5.8E-01
		p4	+Nu7441	15.9	0.9	2.8E-01
		p6	+Nu7441	20.2	1.2	2.0E-01
NHDF-A		p8	+Nu7441	19.8	1.1	2.7E-02
		p4	+Mirin +Nu7441	17.8	1.0	7.9E-01
		p6	+Mirin +Nu7441	17.3	1.0	9.5E-01
		p8	+Mirin +Nu7441	16.6	1.0	6.2E-01
		p4	Control	19.2	1.1	2.8E-01
		p6	Control	19.4	1.1	2.8E-01
		p8	Control	15.8	0.9	4.6E-01
	Random			17.4	-	-

Supplementary Table 7: Average number of histone modifications per DSB in A549 and NHDF-A

Histone Modification	A549				Random
	Photon	Proton	Carbon	Natural	
H3K4me1	0.98	1.34	0.91	1.39	0.80
H3K4me2	0.60	0.92	0.63	0.99	0.50
H3K4me3	0.56	0.71	0.57	0.74	0.45
H3K9ac	0.22	0.29	0.26	0.35	0.18
H3K27ac	0.41	0.58	0.46	0.74	0.32
H3K36me3	0.85	1.13	0.80	1.06	0.66
H4K20me1	0.44	0.70	0.44	0.58	0.36
H3K9me3	0.50	0.38	0.51	0.38	0.48
H3K27me3	0.82	0.66	0.91	0.65	0.67

Histone Modification	NHDF-A					Random
	Photon	Natural	LD5 Doxo	LD10 Doxo	Natural	
H3K4me1	1.14	1.93	1.01	1.05	0.60	0.46
H3K4me2	1.19	1.92	0.67	0.73	0.37	0.31
H3K4me3	1.09	1.71	0.49	0.53	0.32	0.21
H3K9ac	1.01	1.77	0.86	0.86	0.45	0.33
H3K27ac	1.17	1.99	0.91	0.96	0.45	0.33
H3K36me3	1.28	2.26	0.96	0.95	0.55	0.35
H4K20me1	1.41	2.16	0.63	0.69	0.43	0.27
H3K9me3	0.83	1.39	0.35	0.31	0.45	0.43
H3K27me3	1.19	1.75	0.34	0.37	0.47	0.44
H3K79me2	1.38	2.33	0.70	0.74	0.38	0.22

Histone Modification	NHDF-A											
	+IR Ctr			+IR M			+IR N			+IR MN		
	p4	p6	p8	p4	p6	p8	p4	p6	p8	p4	p6	p8
H3K4me1	1.59	1.50	1.85	1.66	1.80	1.57	1.65	1.59	1.82	1.51	1.68	1.56
H3K4me2	1.46	1.43	1.51	1.34	1.55	1.37	1.41	1.50	1.61	1.32	1.43	1.43
H3K4me3	1.31	1.31	1.22	1.32	1.28	1.24	1.27	1.32	1.31	1.28	1.35	1.28
H3K9ac	1.50	1.49	1.81	1.52	1.75	1.54	1.58	1.64	1.91	1.54	1.58	1.56
H3K27ac	1.38	1.53	1.76	1.69	1.60	1.49	1.63	1.69	1.88	1.39	1.62	1.55
H3K36me3	1.20	1.25	1.28	1.29	1.29	1.24	1.16	1.29	1.32	1.13	1.17	1.27
H4K20me1	1.15	1.25	1.49	1.21	1.26	1.20	1.20	1.31	1.68	1.12	1.24	1.21
H3K9me3	1.17	1.23	1.17	1.16	1.22	1.21	1.18	1.35	1.29	1.07	1.30	1.23
H3K27me3	1.31	1.36	1.31	1.18	1.40	1.30	1.21	1.28	1.31	1.27	1.31	1.37
H3K79me2	1.27	1.39	1.48	1.26	1.36	1.33	1.50	1.43	1.57	1.16	1.37	1.35

| Histone Modification | NHDF-A | | | | | | | | | | | |
| | -IR Ctr | | | -IR M | | | -IR N | | | -IR MN | | |
	p4	p6	p8	p4	p6	p8	p4	p6	p8	p4	p6	p8
H3K4me1	1.59	1.66	1.59	1.42	1.84	1.52	1.56	1.89	1.51	1.47	1.64	1.54
H3K4me2	1.30	1.46	1.49	1.35	1.43	1.25	1.35	1.60	1.35	1.32	1.37	1.26
H3K4me3	1.19	1.35	1.23	1.30	1.56	1.21	1.21	1.37	1.26	1.12	1.35	1.21
H3K9ac	1.53	1.61	1.50	1.46	1.74	1.41	1.38	1.73	1.49	1.27	1.45	1.36
H3K27ac	1.50	1.28	1.30	1.38	1.90	1.36	1.48	1.61	1.56	1.36	1.41	1.26
H3K36me3	1.22	1.20	1.17	1.14	1.22	1.23	1.13	1.20	1.17	1.19	1.22	1.21
H4K20me1	1.31	1.12	1.11	1.18	1.74	1.24	1.13	1.35	1.17	1.26	1.25	1.18
H3K9me3	1.13	1.17	1.07	1.26	1.33	1.11	1.17	1.23	1.18	1.11	1.38	1.15
H3K27me3	1.17	1.18	1.36	1.24	1.35	1.28	1.29	1.34	1.27	1.25	1.13	1.26
H3K79me2	1.29	1.23	1.28	1.27	1.68	1.40	1.17	1.44	1.29	1.34	1.26	1.21

Supplementary Table 8: Upstream Regulators associated to genes that show intragenic labeling of radiation-induced DSB

Molecule Type	Upstream Regulator	Target Molecules
Transcription Regulator	AFF1	EPHA7,PTEN
	AJUBA	DOCK1,PTK2
	DNMT3L	IGF2R,SNRPN
	ERG	AR,ARHGAP4,BCR,CAMK1D,CDC42BPB,CYFIP1,DIAPH2,DIAPH3,DIP2A,DOCK1,ELMO1,ETS1,EXT1,FLT1,FYN,GPATCH2L,MAP3K5,MYO10,NLGN1,NRCAM,NRG1,PHACTR2,PLAUR,POU2F1,RGS3,RPS11,RSU1,SVIL,TSC22D3,UTRN,WIPF1
	FOS	ABCC1,ADAM12,AGPS,ALDH1A3,ANK3,ASAP1,B4GALT1,BCL9,CASZ1,COBLL1,DDX3X,DLG1,DOCK1,ELOVL6,FLRT2,GAS2,HIPK3,LAMA3,LARGE,MAP4K4,MLH1,PFN1,PRKCE,PTPRD,QKI,RARA,RIPK1,RXRA,SDK1,SMAD7,SULF2,TAOK1,UTRN,ZHX2
	GABPA	AURKA,FAS,SKP2,SPI1
	HDAC2	ACAN,ATG5,COL1A2,DCLRE1C,FAS,FDXR,LGR6,LIG1,PRIM2,SFRP1,TP63
	HDAC4	CACNA2D1,CHRDL1,CNKSR2,CNTN1,DGKB,DLG2,GABRA1,GPHN,GPM6A,MAPK10,NRCAM,NTM,PRKCA,PRKCB,RAB3C,RGS7
	HOXC6	ABCB1,BMP7,CNTN1,COL23A1,PDE4DIP,SIPA1L1,ZNF536
	HR	ANO1,COL25A1,CSNK2A2,CTNND2,DYM,MAGI2
	HSF1	ABCB1,BMP7,CDK1,FBLN1,FGF1,HERC4,JARID2,JMJD6,MLH1,MSH4,NFATC2,PPP3CB,RAD51B,RAD51D,RARRES3,STAG2,STIP1,TCP1,WNT3
	IKZF1	AQP1,CCND2,DOCK1,DOCK9,IFI16,LHFP,LRP1,NCAM1,NOTCH3,PRKCQ,SPI1,SULF2,VWF
	JARID2	CCND2,PTEN

	JUN	ABCB1,BCL9,CDK1,COL1A2,HNRNPA2B1,IFI16,IGF2R,ITGAV,PLAUR,PRKCE,PTEN,PTPRD,RARA,RXRA,SDK1,SMAD7,SNRPN,STAT1,SULF2,XYLT1
	KLF3	ANK1,KLF12,SELL
	MEOX2	ITGAV,LRP1
	SATB1	ACTG1,CTNNA1,ETS1,ILF3,LPIN2,MAL,MEF2A,NR2C2,PRKCB,RIPK1,SELL,SIPA1L2,SPI1,TAF4B,TAOK1,TMEM117,TNFRSF8,TSC22D3,VTA1
	SPDEF	COL4A1,COL4A2,COL4A5,COL4A6,COL5A1,CYFIP1,ITGA5,PLAUR,PRKCA,SMAD4
	TAF7L	AR,CPA6,FSCN1,HK1,LPIN1,NR6A1,TSKS,ZFX
	TP53	ABCB1,ABCC1,APBB2,AR,ASTN2,ATXN1,AURKA,BAI3,BAK1,BCAS3,BRAF,C12orf5,CAMK2B,CARD11,CCND2,CDC42BPA,CDK1,CEP164,COL1A2,COL4A1,COX7A2,DFFB,DGKA,DLG1,DNAJA2,EDA2R,EGFL6,EIF4G3,ENG,FAS,FDXR,FLRT2,FRMD4A,FYN,GNA13,GTF3C2,HSPA1L,HTT,IFI16,IL4I1,INPP4A,KCNMA1,KIAA0368,LMO3,LPIN1,MAP2K4,MAP4,MAP4K1,MBNL2,MED13L,MIS18A,MLH1,MYH10,MYOF,NCAPG,NCAPH,NEDD8,NPEPPS,NR6A1,NUDT5,PAK3,PCCA,PCDH7,PDCD6IP,PFN1,PIAS2,PIK3C3,PLAUR,PLK1,PPARD,PPP4R2,PRKAG2,PRKAR2A,PRKCB,PRKCE,PRKD1,PRKG1,PSMD1,PSMD12,PTEN,PTK2,RAD50,RECQL4,RPS6KA2,SCP2,SMAD7,SRC,STAT1,STIP1,TACC2,TCF7L2,THBS1,TIMP2,TJP1,TP63,TP73,TRIO,TSC22D3,TULP4,UBE2C,UVRAG,XRCC5,ZNF365
	Yap1	MYH11,MYOCD
Growth Factor	FGF23	AKAP1,ATP1A2,SLC34A1
	TGFB3	ENG,ETS1,TJP1,ZEB2
microRNA	miR-182-5p	ADCY6,BCL2L14,CARD11,CDH4,COL11A2,NCAM1, NFASC,VWF
	miR-185-5p	CDC42,RHOA
	miR-30a-3p	ANKFY1,ARMC8,RSU1,SCAF11
	miR-92a-3p	CCND2,PTEN,TP63
	mir-145	FSCN1,MYH11,MYOCD
	mir-182	ADCY6,BCL2L14,CARD11,CDH4,COL11A2,NCAM1, NFASC,VWF
	mir-183	PTEN,TAOK1
	mir-22	ARRB1,BMP7,MECOM
	mir-8	E2F3,PTEN,ZEB2

Supplementary Table 9: Upstream Regulators associated to genes that show intragenic labeling of doxorubicin-induced DSB

Molecule Type	Upstream Regulator	Target Molecules
Tramscription Regulator	ATF1	CREB1,FTH1,INHBA,RHOA,TEAD1,THBS1
	ATF2	CTNNB1,FN1,MAPK14,MYLK,PDGFRA,PPARG,PPARGC1A,PSEN1,PTEN,ZNF268
	BACH1	CALM1 (includes others),EWSR1,FTH1,SQSTM1,TTC23
	BMI1	DLC1,HK2,LOX,MYT1,NFKB1,NID1,P4HA1,SMAD3,VEZT
	E2F1	ACTR3,ADK,ALG5,ATAD2,BIRC5,BRCA1,CALM1 (includes others),CASP9,CBX5,CDK1,CWC27,DUT,EYA3,HNRNPD,MAPK14,MCL1,MCM3,MFAP1,MSH2,NCL,NCOA3,NDC80,PA2G4,PDK1,PIK3R1,PRIM2,PRKDC,PSAT1,PTPN4,RAD54L,RBBP4,RBFOX2,RBL1,REV3L,RRM1,RUVBL1,SAFB,SMARCA5,TOP2A,TRMT13,TXNRD1,VRK1,WEE1,YWHAE
	E2F4	ATRX,BARD1,BRCA1,CALM1 (includes others),CBX5,CDK1,CENPE,DUT,H2AFZ,LPAR1,MAP3K7,MCM3,MCM6,MSH2,NCL,NDC80,NPM1,PKN2,PRIM2,PRKDC,PSAT1,RAD54L,RBBP4,RBL1,RRM1,SRRM2,TOP2A,TRMT13,TTK
	ELK1	BMPR2,MCL1,MECOM,MYLK,PRKCA,RUNX2,THBS1,TIPARP
	ERG	ABCC4,ADD1,ARFGEF2,ARHGAP22,ARHGAP24,CDC42BPB,CDK5RAP2,CLIP1,CTNNB1,DAAM1,DIAPH2,DIAPH3,DIP2A,DOCK1,DOCK2,DSTN,ETS1,EXT1,MAGI1,MYO1D,MYO5A,NPHP1,NRG1,NUMB,PCNP,PPAP2B,PTPN11,PTPN4,RAB7A,RASA2,RPL22,RYR3,SLIT2,SVIL,TACC1,TMEM263,TSC22D3,UTRN,WDR91
	FOXM1	ADAM17,BIRC5,CDK1,CDKN3,ESR1,LAMA4,LOX,SKP2,VCAN,ZEB1,ZEB2
	GMNN	ANGPT1,ANKRD17,APCDD1,CDX2,CTNNB1,DYRK1A,EFEMP1,GREB1,INHBA,NCKAP1,NUDT21,PDGFRA,TEAD1,TGFB2,TWSG1,TXNRD1,WLS,YAP1
	HAND2	ATP2A2,CCBE1,COL1A2,COL3A1,FAP,GLI3,GREM1,LOC102724428/SIK1,SHH
	HDAC4	CAP2,CNKSR2,DGKB,DLG2,DYRK1A,FGF13,GABRG3,GPHN,HIF1A,MAPK1,MAPK10,NBEA,NTM,PRKCA,SYNGAP1
	HMGA1	BRCA1,CAV2,DAB2,ELK1,INSIG1,KITLG,PPARG,STAT3
	HR	COL25A1,DYM,FGF13,IBTK,KCNIP1,MAGI2,TAF7L,UBR2
	KDM5B	AP4S1,BRCA1,CDK1,DHCR24,DLGAP5,EPB41L1,INSIG1,ITGB1BP1,MAPK8IP3,MCM3,NDC80,PHF20,PSD3,PSIP1,PTPLA,SMC5,SNRPG,SPTSSA,TOP2A,TTK
	MDM2	FOXO3,HIF1A,IGF1R,KAT2B,MDM4,TOP2A
	MTDH	KDM6A,PDCD11
	MYC	ACACA,ADD1,ADK,AEBP2,AKAP12,ANGPT1,ANKRD17,APCDD1,ARF3,ARHGAP22,CAST,CCT3,CD44,CDK1,CDX2,CNTNAP2,COL15A1,COL1A2,COL3A1,COL4A1,COL4A2,COL5A1,COL5A2,COL6A3,COL8A1,CTNNB1,EFEMP1,EFTUD2,EIF2S1,EIF2S2,ENO1,F2R,FAP,FBN1,FECH,FN1,FSTL1,FTH1,GFPT1,GIGYF2,GLS,GLS2,GLUD1,GREB1,GTF2B,GTF2F2,HAPLN1,HERC5,HIF1A,HIVEP2,HK2,HSP90AA1,INHBA,IPO7,IQGAP2,IRF6,ITGB1,JARID2,LOX,MAN2A1,MCM6,MFAP1,NCKAP1,NCL,NFAT5,NFATC3,NPM1,NUDC,NUDT21,PDCD4,PDGFRA,PDK1,PGK1,PHF20,PHF21A,POLR2D,PREP,RAB10,RARS,RBBP4,RRM2B,RUVBL1,SCAMP1,SCEL,SDCBP,SKP2,ST3GAL1,ST3GAL4,TEAD1,TGFB2,THBS1,THBS2,TIMM10B,TNC,TNFRSF10B,TPP2,TWSG1,TXNRD1,WLS,YAP1,YTHDC1
	NEUROG1	ADD3,CEMIP,COL3A1,FAP,FN1,GREM1,INHBA,LGR5,PAPPA,SPOCK1,SULF1,THBS1,

		TMTC2
	RB1	ALG5,BIRC5,BRCA1,CASP9,CBX5,CDC25C,CDK1,CDX2,CWC27,GMPS,IGF1R,KPNA4,MAPK1,MFAP1,MTOR,PPARG,PPARGC1A,PRKAR1A,PTPN4,RBL1,SAFB,SMARCA5,YWHAE,ZEB1,ZEB2
	RBFOX2	PTBP2,PTEN
	SIAH2	ANGPT1,HIF1A,NCOR1
	SMARCE1	BRCA1,BRCA2,CDK1,CENPE,SCN2A
	SOX1	ANGPT1,ANKRD17,APCDD1,CDX2,CTNNB1,EFEMP1,GREB1,INHBA,NCKAP1,NUDT21,PDGFRA,TEAD1,TGFB2,TWSG1,TXNRD1,WLS,YAP1
	SOX11	ADAM10,ADSS,CNOT6,EBF1,HILPDA,IL6ST,ITGB1,MSI2,MYO1B,NREP,RNF122,SETMAR,STAT1
	SOX3	ANGPT1,ANKRD17,APCDD1,CDX2,CTNNB1,EFEMP1,GREB1,INHBA,NCKAP1,NUDT21,PDGFRA,TEAD1,TGFB2,TWSG1,TXNRD1,WLS,YAP1
	SPDEF	CDH11,COL4A1,COL4A2,COL5A1,COL5A2,COL6A3,DKK3,HIF1A,LAMC1,PRKCA,SMAD3,TNC,ZEB1
	TFAP4	CASP9,CD44,DYRK1A,LGR5
	TP53	ABCB4,ACSL3,ADD3,AHCY,AKAP12,ARHGAP5,ASTN2,ASXL1,ATG4A,ATL3,ATXN1,BCAS3,BIRC5,BRCA1,C9,CAND1,CAP1,CDC16,CDC25C,CDC42BPA,CDC42EP3,CDK1,CDKN3,CHUK,COL1A2,COL3A1,COL4A1,COL5A2,COPB1,CSNK1D,CTNNB1,DHCR24,DICER1,DKK3,DLD,DLG1,DNAJA2,DNM1L,DSTN,DUT,EIF4G3,ENPP2,ESR1,ETFA,F2R,FAM120A,FDXR,FGF2,FLRT2,FN1,FOXO3,FSTL1,GLS2,GNA13,GNA14,GSK3B,GSN,GTF3C2,HADHA,HDLBP,HERC5,HIF1A,HK2,HSP90AA1,HSPG2,HTT,IDH1,IGF1R,IGFBP7,IL31RA,IL7,INHBA,INPP4A,IPO7,IPO9,KAT2B,KCNMA1,KIAA0368,KIF23,KITLG,KPNB1,LATS2,LIMA1,LPP,MAN2A1,MAP2K4,MAP4,MAPK1,MCL1,MCM3,MCM6,MDM4,MED13L,MICALL1,MPI,MSH2,MTA1,MTDH,MYBL1,MYH9,MYO6,NDC80,NLRC4,NR6A1,NRAP,NUP153,P4HA1,PAK3,PANK1,PARD6B,PARK7,PCCA,PDCD6IP,PDE4B,PDGFRA,PDK1,PIAS2,PIK3C3,PIK3R4,PLOD2,POLE2,PPM1D,PPP1R13L,PRKAR2A,PRKCZ,PRKG1,PSMD1,PSMD12,PTEN,PTK2,PTPN11,PTPRM,PVRL3,RAD23A,RAD50,RAF1,RBBP4,RBL1,RFC1,RGS12,ROBO1,RPN1,RPN2,RPS6KB1,RRM1,RRM2B,RUNX2,SCMH1,SEC62,SHROOM3,SIRT1,SMAD6,SMC3,SMURF2,SON,SPATA18,SQLE,SRSF3,STAT1,STAU1,STK17A,TANK,TDG,TGFB2,THBS1,TNFRSF10B,TOP2A,TPX2,TRIO,TSC22D3,TTK,UBE2B,UBL3,UIMC1,ULK2,USO1,USP14,UVRAG,VCAN,VCL,VRK1,XRCC5,ZEB1,ZFP36L1
	TP63	ATG4A,BRCA1,CAST,CDC25C,CDC42,CDK1,COL4A1,COL5A1,CYR61,DICER1,F2R,FBN1,FN1,FOXO3,HMGA2,IGFBP7,KIF23,PNPT1,PRKCZ,PTEN,SPATA18,THBS1,TNC,ULK2,UVRAG,WEE1,ZEB1
	TWIST2	CD44,CTNNB1,FN1,ZEB1,ZEB2
	XBP1	ARCN1,ATP2A2,COPB1,COPB2,DERL1,DNAJC10,EDEM1,ERLEC1,ESR1,ETS1,FN1,GOLGA3,GOLGA4,GORASP2,HM13,KDELR2,PDIA3,PPIB,RPN1,RPN2,SDF2L1,SEC22B,SEC23A,SEC31A,SEC61G,SEC63,SMC3,SRP54,SRP68,SSR1,TXN,USO1,XRCC6
Growth Factor	BMP6	ADD3,AKTIP,CD44,ERRFI1,ETV5,FRMD4B,INHBA,SCEL,SLC39A10,SMARCA1,TXNRD1,VCL,ZEB1
	CTGF	CD44,COL4A2,COL8A1,CTNNB1,ENO1,ESR1,FN1,HAPLN1,HIF1A,KITLG,LIMS1,P4HA1,PDE7A,PI4K2B,TMOD3,WNK1

	FGF2	BIRC5,CAPZA2,CDH11,CDX2,CHRM3,COL1A2,CTNNB1,ENO1,FGF2,FN1,FTH1,NF1, PGK1,PPARG,RUNX2,SQSTM1,ST3GAL1,ST3GAL4,TGFB2,TOP2A
	WISP2	CD44,ESR1,FN1,HIF1A,IGFBP7,SMAD3,TGFBR1,ZEB1
microRNA	miR-128-3p	TGFBR1,WEE1
	miR-185-5p	CDC42,RHOA
	miR-192-5p	BIRC5,IGF1R,ZEB1
	miR-199a-5p	BIRC5,DYRK1A,ETS1
	miR-19b-3p	BIRC5,PTEN
	miR-21-5p	MSH2,PDCD4,PIK3R1,PTEN,STAT3
	miR-217-5p	ACACA,PPARGC1A,SIRT1,TRPS1
	miR-29b-3p	COL1A2,COL5A2,DCP2,HDAC4,LAMC1,MAPRE2,NAV3,RERE,ZFP36L1
	miR-30a-3p	ARMC8,KIF1B,LIMA1,MCCC2,RAB8B,SCAF11,WDR44
	mir-1	ADD1,ANXA2,CAP1,FN1,PICALM,RASA1,RHEB,XPO6
	mir-144	KLF7,RAB14,RAC1,RASA1,TLK2,YWHAZ
	mir-154	BIRC5,PTEN
	mir-181	MCL1,PTEN,SPTLC1
	mir-199	DYRK1A,EGLN1,HIF1A,SIRT1
	mir-22	BMPR1B,EDC3,MECOM,YWHAZ
	mir-451	KLF7,RAB14,RAC1,RASA1,TLK2,YWHAZ
	mir-8	ITGB1,PTEN,TGFB2,YAP1,ZEB1,ZEB2,ZFPM2

Supplementary Table 10: Distribution of DSB sites identified in the three passages. Chr: Chromosome, TSS: Transcription Start Site

Treament	Radiation	DSB Frequency in passage			Fold Change	Chr	Position	Upstream of TSS [bp]	Intron/ Exon	Downstream of Gene	RefSeq Gene
		4	6	8							
+IR	Ctr	0.009	0.002	0.001	0.11	6	124852660	0	In3	0	NKAIN2
+IR	Ctr	0.037	0.014	0.005	0.14	8	71528210	7606		0	TRAM1
+IR	Ctr	0.028	0.024	0.013	0.46	10	19146063	0		179123	ARL5B
+IR	Ctr	0.019	0.012	0.013	0.68	15	39600267	0		53219	C15orf54
+IR	Ctr	0.009	0.035	0.007	0.78	6	160280122	0		38387	PNLDC1
+IR	Ctr	0.028	0.002	0.056	2.00	2	237175619	2631		0	ASB18
+IR	M	0.065	0.002	0.003	0.05	7	39788845	0		41130	RALA
+IR	M	0.028	0.889	0.002	0.07	2	236001953	0		37597	SH3BP4
+IR	M	0.009	0.011	0.001	0.11	1	200996263	3435		0	KIF21B
+IR	M	0.009	0.002	0.004	0.44	14	29163728	72559		0	FOXG1
+IR	M	0.009	0.002	0.005	0.56	20	25383405	4918		0	GINS1
+IR	M	0.009	0.006	0.008	0.89	12	69593947	39370		0	CPSF6
+IR	M	0.046	0.023	0.057	1.24	6	87451391	195633		0	HTR1E
+IR	M	0.019	0.043	0.411	21.63	X	40608992	14188		0	MED14
+IR	N	0.009	0.002	0.001	0.11	19	22248131	0	In1	0	ZNF257
+IR	N	0.009	0.002	0.001	0.11	17	60848501	0	In3	0	MARCH10

+IR	N	0.009	0.002	0.001	0.11	6	100799124	0		37627	*SIM1*
+IR	N	0.009	0.002	0.001	0.11	3	194496759	0		86995	*FAM43A*
+IR	N	0.019	0.011	0.011	0.58	2	220242059	0	In13	0	*DNPEP*
+IR	N	0.009	0.049	0.009	1.00	2	9428234	0	In2	0	*ASAP2*
+IR	N	0.019	0.014	0.039	2.05	6	75699157	0		94886	*COL12A1*
+IR	N	0.009	0.040	0.021	2.33	8	63766653	0	In4	0	*NKAIN3*
+IR	N	0.009	0.163	0.032	3.56	X	142717350	0	Ex2	0	*SLITRK4*
+IR	N	0.009	0.002	0.034	3.78	5	132801252	0	In3	0	*FSTL4*
+IR	N	0.019	0.011	0.340	17.89	12	26980249	0	In1	0	*ITPR2*
+IR	MN	0.046	0.002	0.001	0.02	16	24990921	0	In1	0	*ARHGAP17*
+IR	MN	0.093	0.041	0.004	0.04	9	84828995	151991		0	*FAM75B*
+IR	MN	0.083	0.005	0.005	0.06	16	79615445	0		12301	*MAF*
+IR	MN	0.009	0.002	0.001	0.11	7	116876671	0		6598	*ST7*
+IR	MN	0.009	0.002	0.001	0.11	2	138258041	0	In14	0	*THSD7B*
+IR	MN	0.009	0.006	0.001	0.11	5	151230348	0	In7	0	*GLRA1*
+IR	MN	0.037	0.018	0.005	0.14	20	25686101	8632		0	*ZNF337*
-IR	Ctr	0.120	0.032	0.001	0.01	13	59624112	0		615613	*DIAPH3*
-IR	Ctr	0.009	0.002	0.004	0.44	8	103799355	0		39182	*AZIN1*
-IR	M	0.009	0.006	0.001	0.11	3	155453796	0		26605	*C3orf33*
-IR	N	0.019	0.002	0.001	0.05	16	71576571	0		4158	*CHST4*
-IR	N	0.009	0.002	0.001	0.11	1	12156057	0	In2	0	*TNFRSF8*
-IR	N	0.009	0.002	0.001	0.11	21	18509311	376019		0	*CXADR*
-IR	N	0.009	0.002	0.001	0.11	14	68699623	0	In7	0	*RAD51L1*
-IR	N	0.009	0.002	0.004	0.44	2	137507818	240644		0	*THSD7B*
-IR	MN	0.037	0.002	0.001	0.03	12	49271656	12003		0	*RND1*
-IR	MN	0.009	0.002	0.001	0.11	2	40663201	0	In1	0	*SLC8A1*
-IR	MN	0.009	0.002	0.001	0.11	5	96442660	0	In3	0	*LIX1*

Supplementary Table 11: Number and position of the top ten DSB sites over time. TSS: Transcription Start Site; TTS: Transcription Termination Site; +IR: irradiated, Ctr: cells with active NHEJ-repair activity; MN: Mirin + Nu7441 treatment; TSS: Transcription Start Site; TTS: Transcription Termination Site

Radiation	Treatment	Upstream of TSS			In Gene			Downstream of TTS		
		p4	p6	p8	p4	p6	p8	p4	p6	p8
+IR	Ctr	4	3	3	3	4	5	3	3	2
+IR	+Mirin	5	3	5	3	6	2	2	1	3
+IR	+Nu7441	6	1	2	2	5	4	2	4	4
+IR	+MN	2	2	3	5	6	3	3	3	4

Supplementary Table 12: Association of radiation-related DSB hotspots with neoplasia and cancer

Symbol	Entrez Gene Name	Type(s)	Neo-plasia	Cancer
AAMDC	adipogenesis associated, Mth938 domain containing	other		
ARMC1	armadillo repeat containing 1	other		X
ASB1	ankyrin repeat and SOCS box containing 1	transcription regulator		
ATP6V0A1	ATPase, H+ transporting, lysosomal V0 subunit a1	transporter	X	X
ATP6V1H	ATPase, H+ transporting, lysosomal 50/57kDa, V1 subunit H	transporter		
BCCIP	BRCA2 and CDKN1A interacting protein	other	X	X
BCL11B	B-cell CLL/lymphoma 11B (zinc finger protein)	other	X	X
C11orf80	chromosome 11 open reading frame 80	other	X	X
CARD11	caspase recruitment domain family, member 11	kinase	X	X
CD93	CD93 molecule	other	X	X
CDH10	cadherin 10, type 2 (T2-cadherin)	other	X	X
CEBPZ	CCAAT/enhancer binding protein (C/EBP), zeta	other	X	X
CLMN	calmin (calponin-like, transmembrane)	other	X	X
CNOT1	CCR4-NOT transcription complex, subunit 1	other	X	X
CNTNAP2	contactin associated protein-like 2	other	X	X
CTNNA2	catenin (cadherin-associated protein), alpha 2	other	X	X
CWH43	cell wall biogenesis 43 C-terminal homolog (S. cerevisiae)	other	X	X
CYP39A1	cytochrome P450, family 39, subfamily A, polypeptide 1	enzyme		
EDIL3	EGF-like repeats and discoidin I-like domains 3	other	X	X
EDN1	endothelin 1	cytokine	X	X
ELF4	E74-like factor 4 (ets domain transcription factor)	transcription regulator		
EMR2	egf-like module containing, mucin-like, hormone receptor-like 2	other	X	X
ERI1	exoribonuclease 1	enzyme	X	X
ERICH3	glutamate-rich 3	other	X	X
FAM120C	family with sequence similarity 120C	other	X	X
FAM174A	family with sequence similarity 174, member A	other	X	X
FCGR1B	Fc fragment of IgG, high affinity Ib, receptor (CD64)	transmem-brane receptor		X
FPGT-TNNI3K	FPGT-TNNI3K readthrough	other		X
FRMPD4	FERM and PDZ domain containing 4	other	X	X
GALNT5	polypeptide N-acetylgalactosaminyltransferase 5	enzyme	X	X
GINS3	GINS complex subunit 3 (Psf3 homolog)	other	X	X
H2AFY	H2A histone family, member Y	other	X	X
HIST1H1C	histone cluster 1, H1c	other	X	X
HIST1H2AC	histone cluster 1, H2ac	other		

HIST1H3B	histone cluster 1, H3b	other	X	X
HIVEP3	human immunodeficiency virus type I enhancer binding protein 3	transcription regulator	X	X
HS6ST1	heparan sulfate 6-O-sulfotransferase 1	enzyme	X	X
HTRA2	HtrA serine peptidase 2	peptidase	X	X
IL17B	interleukin 17B	cytokine		X
INTS4	integrator complex subunit 4	other	X	X
IRS2	insulin receptor substrate 2	enzyme	X	X
KCTD20	potassium channel tetramerization domain containing 20	ion channel	X	X
KIAA2013	KIAA2013	other		X
LACE1	lactation elevated 1	other		X
MIIP	migration and invasion inhibitory protein	other		
MORN1	MORN repeat containing 1	other	X	X
MYT1L	myelin transcription factor 1-like	transcription regulator	X	X
NDRG4	NDRG family member 4	other	X	X
NKX2-3	NK2 homeobox 3	transcription regulator		
NLGN1	neuroligin 1	enzyme	X	X
OR5AN1	olfactory receptor, family 5, subfamily AN, member 1	G-protein coupled receptor	X	X
OR7A17	olfactory receptor, family 7, subfamily A, member 17	G-protein coupled receptor	X	X
OR7A5	olfactory receptor, family 7, subfamily A, member 5	G-protein coupled receptor	X	X
PDE4D	phosphodiesterase 4D, cAMP-specific	enzyme	X	X
PEX10	peroxisomal biogenesis factor 10	other		X
PHF8	PHD finger protein 8	enzyme	X	X
PLA2G7	phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)	enzyme	X	X
PLAUR	plasminogen activator, urokinase receptor	transmem-brane receptor	X	X
PNLDC1	poly(A)-specific ribonuclease (PARN)-like domain containing 1	other	X	X
PPP1R3B	protein phosphatase 1, regulatory subunit 3B	other		X
PRKX	protein kinase, X-linked	kinase	X	X
RPIA	ribose 5-phosphate isomerase A	enzyme	X	X
RSF1	remodeling and spacing factor 1	transcription regulator	X	X
SCLY	selenocysteine lyase	enzyme	X	X
SETBP1	SET binding protein 1	other	X	X
SIPA1L2	signal-induced proliferation-associated 1 like 2	other	X	X
SPTBN2	spectrin, beta, non-erythrocytic 2	other	X	X

STAT3	signal transducer and activator of transcription 3 (acute-phase response factor)	transcription regulator	X	X
STAT5B	signal transducer and activator of transcription 5B	transcription regulator	X	X
STK38	serine/threonine kinase 38	kinase		
SULT6B1	sulfotransferase family, cytosolic, 6B, member 1	other	X	X
SYNE3	spectrin repeat containing, nuclear envelope family member 3	other	X	X
TAS2R41	taste receptor, type 2, member 41	G-protein coupled receptor		X
TCP1	t-complex 1	other	X	X
TEX36	testis expressed 36	other		X
TFPI	tissue factor pathway inhibitor (lipoprotein-associated coagulation inhibitor)	other	X	X
TIFAB	TRAF-interacting protein with forkhead-associated domain, family member B	other	X	X
TNFRSF8	tumor necrosis factor receptor superfamily, member 8	transmem-brane receptor	X	X
TOP1	topoisomerase (DNA) I	enzyme	X	X
TRIM38	tripartite motif containing 38	other	X	X
UBE2E2	ubiquitin-conjugating enzyme E2E 2	enzyme		
UBE2F	ubiquitin-conjugating enzyme E2F (putative)	enzyme		
WFDC2	WAP four-disulfide core domain 2	other	X	X
XRCC1	X-ray repair complementing defective repair in Chinese hamster cells 1	other	X	X
ZMAT2	zinc finger, matrin-type 2	other	X	X
ZNF280C	zinc finger protein 280C	other	X	X
ZNF33B	zinc finger protein 33B	transcription regulator	X	X
ZYX	zyxin	other	X	X

Supplementary Table 13: Association of doxorubicin-related DSB hotspots with neoplasia and cancer

Symbol	Entrez Gene Name	Type(s)	Neo-plasia	Cancer
AACS	acetoacetyl-CoA synthetase	enzyme	X	X
AAMP	angio-associated, migratory cell protein	other	X	X
AARS	alanyl-tRNA synthetase	enzyme	X	X
ABCA12	ATP-binding cassette, sub-family A (ABC1), member 12	transporter	X	X
ADPRHL1	ADP-ribosylhydrolase like 1	enzyme	X	X
AGBL2	ATP/GTP binding protein-like 2	enzyme	X	X
AKAP12	A kinase (PRKA) anchor protein 12	transporter	X	X
ANKFY1	ankyrin repeat and FYVE domain containing 1	transcription regulator	X	X
ANKRD12	ankyrin repeat domain 12	other	X	X

AP4E1	adaptor-related protein complex 4, epsilon 1 subunit	other	X	X
ARHGEF11	Rho guanine nucleotide exchange factor (GEF) 11	other	X	X
ARID4A	AT rich interactive domain 4A (RBP1-like)	transcription regulator	X	X
ARIH1	ariadne RBR E3 ubiquitin protein ligase 1	enzyme	X	X
ARL4D	ADP-ribosylation factor-like 4D	enzyme	X	X
ARMC9	armadillo repeat containing 9	other	X	X
ASB3/GPR7 5-ASB3	ankyrin repeat and SOCS box containing 3	transcription regulator		X
ASTN2	astrotactin 2	other	X	X
ASZ1	ankyrin repeat, SAM and basic leucine zipper domain containing 1	transcription regulator	X	X
ATP10D	ATPase, class V, type 10D	transporter	X	X
ATP13A3	ATPase type 13A3	transporter	X	X
ATP13A4	ATPase type 13A4	transporter		X
B4GALNT2	beta-1,4-N-acetyl-galactosaminyl transferase 2	enzyme		X
BAZ1B	bromodomain adjacent to zinc finger domain, 1B	transcription regulator	X	X
BIRC6	baculoviral IAP repeat containing 6	enzyme	X	X
BRWD1	bromodomain and WD repeat domain containing 1	transcription regulator	X	X
C14orf166	chromosome 14 open reading frame 166	other	X	X
C1QTNF4	C1q and tumor necrosis factor related protein 4	other		X
C5orf34	chromosome 5 open reading frame 34	other	X	X
C6orf211	chromosome 6 open reading frame 211	other	X	X
CA12	carbonic anhydrase XII	enzyme	X	X
CALD1	caldesmon 1	other	X	X
CANX	calnexin	other	X	X
CASP9	caspase 9, apoptosis-related cysteine peptidase	peptidase	X	X
CCAR1	cell division cycle and apoptosis regulator 1	transcription regulator	X	X
CCBE1	collagen and calcium binding EGF domains 1	other		X
CCDC12	coiled-coil domain containing 12	other	X	X
CCNT1	cyclin T1	transcription regulator	X	X
CDAN1	codanin 1	other		X
CDC25C	cell division cycle 25C	phosphatase		X
CDC27	cell division cycle 27	other	X	X
CDK12	cyclin-dependent kinase 12	kinase	X	X
CDK13	cyclin-dependent kinase 13	kinase	X	X
CELF1	CUGBP, Elav-like family member 1	translation regulator	X	X
CEP97	centrosomal protein 97kDa	other		X
CFAP36	cilia and flagella associated protein 36	other	X	X

CHD8	chromodomain helicase DNA binding protein 8	enzyme	X	X
CHRDL2	chordin-like 2	other		X
CHUK	conserved helix-loop-helix ubiquitous kinase	kinase	X	X
CLNS1A	chloride channel, nucleotide-sensitive, 1A	ion channel		X
CLP1	cleavage and polyadenylation factor I subunit 1	other	X	X
CMSS1	cms1 ribosomal small subunit homolog (yeast)	other		X
CNIH4	cornichon family AMPA receptor auxiliary protein 4	other	X	X
CNTRL	centriolin	transcription regulator	X	X
COL3A1	collagen, type III, alpha 1	other	X	X
COL5A2	collagen, type V, alpha 2	other	X	X
COL8A1	collagen, type VIII, alpha 1	other		X
CORIN	corin, serine peptidase	peptidase	X	X
CPNE3	copine III	kinase	X	X
CPQ	carboxypeptidase Q	peptidase	X	X
CSNK1G1	casein kinase 1, gamma 1	kinase	X	X
CXCR1	chemokine (C-X-C motif) receptor 1	G-protein coupled receptor		X
CXCR2	chemokine (C-X-C motif) receptor 2	G-protein coupled receptor	X	X
DAGLB	diacylglycerol lipase, beta	enzyme		X
DAPP1	dual adaptor of phosphotyrosine and 3-phosphoinositides	other		X
DCUN1D4	DCN1, defective in cullin neddylation 1, domain containing 4	other		X
DDX50	DEAD (Asp-Glu-Ala-Asp) box polypeptide 50	enzyme	X	X
DENND5A	DENN/MADD domain containing 5A	other	X	X
DIP2A	DIP2 disco-interacting protein 2 homolog A (Drosophila)	transcription regulator	X	X
DIS3	DIS3 exosome endoribonuclease and 3'-5' exoribonuclease	enzyme	X	X
DNA2	DNA replication helicase/nuclease 2	enzyme	X	X
DNAAF2	dynein, axonemal, assembly factor 2	other	X	X
DNAJA3	DnaJ (Hsp40) homolog, subfamily A, member 3	other		X
DNAJB6	DnaJ (Hsp40) homolog, subfamily B, member 6	transcription regulator	X	X
DNAJC10	DnaJ (Hsp40) homolog, subfamily C, member 10	enzyme	X	X
DNAJC16	DnaJ (Hsp40) homolog, subfamily C, member 16	other		X
DOCK11	dedicator of cytokinesis 11	other	X	X
DST	dystonin	other	X	X
DTNB	dystrobrevin, beta	other	X	X
DYNC2H1	dynein, cytoplasmic 2, heavy chain 1	other	X	X

DYNLL1	dynein, light chain, LC8-type 1	other	X	X
ECD	ecdysoneless homolog (Drosophila)	transcription regulator	X	X
EFCAB13	EF-hand calcium binding domain 13	other	X	X
EGR1	early growth response 1	transcription regulator	X	X
EHMT1	euchromatic histone-lysine N-methyltransferase 1	transcription regulator	X	X
EIF1	eukaryotic translation initiation factor 1	translation regulator		X
EIF2AK2	eukaryotic translation initiation factor 2-alpha kinase 2	kinase	X	X
EIF2S1	eukaryotic translation initiation factor 2, subunit 1 alpha, 35kDa	translation regulator	X	X
ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2	enzyme	X	X
ERICH5	glutamate-rich 5	other	X	X
ERLEC1	endoplasmic reticulum lectin 1	other	X	X
ETV3L	ets variant 3-like	other	X	X
EXT1	exostosin glycosyltransferase 1	enzyme	X	X
F2R	coagulation factor II (thrombin) receptor	G-protein coupled receptor	X	X
F2RL1	coagulation factor II (thrombin) receptor-like 1	G-protein coupled receptor	X	X
FAM149B1	family with sequence similarity 149, member B1	other	X	X
FAM179B	family with sequence similarity 179, member B	other	X	X
FAM186B	family with sequence similarity 186, member B	other		X
FAM71D	family with sequence similarity 71, member D	other		X
FAM98A	family with sequence similarity 98, member A	other	X	X
FBXO22	F-box protein 22	enzyme	X	X
FILIP1	filamin A interacting protein 1	other	X	X
FMO4	flavin containing monooxygenase 4	enzyme		X
FN1	fibronectin 1	enzyme	X	X
FNBP4	formin binding protein 4	other	X	X
FRMD6	FERM domain containing 6	other	X	X
FYTTD1	forty-two-three domain containing 1	other	X	X
GABBR1	gamma-aminobutyric acid (GABA) B receptor, 1	G-protein coupled receptor	X	X
GAST	gastrin	other	X	X
GCN1L1	GCN1 general control of amino-acid synthesis 1-like 1 (yeast)	translation regulator	X	X
GEMIN5	gem (nuclear organelle) associated protein 5	other	X	X
GIP	gastric inhibitory polypeptide	other		X

GLS	glutaminase	enzyme	X	X
GNPNAT1	glucosamine-phosphate N-acetyltransferase 1	enzyme	X	X
GORASP2	golgi reassembly stacking protein 2, 55kDa	other	X	X
GP5	glycoprotein V (platelet)	other	X	X
GSN	gelsolin	other	X	X
HEATR5B	HEAT repeat containing 5B	other	X	X
HMCN1	hemicentin 1	other	X	X
HMGA2	high mobility group AT-hook 2	enzyme	X	X
HMGXB4	HMG box domain containing 4	other	X	X
HNRNPH3	heterogeneous nuclear ribonucleoprotein H3 (2H9)	other	X	X
IGF2BP1	insulin-like growth factor 2 mRNA binding protein 1	translation regulator	X	X
IGF2BP2	insulin-like growth factor 2 mRNA binding protein 2	translation regulator	X	X
IGFBP7	insulin-like growth factor binding protein 7	transporter	X	X
IMPG1	interphotoreceptor matrix proteoglycan 1	other	X	X
IPO7	importin 7	transporter	X	X
IQGAP2	IQ motif containing GTPase activating protein 2	other	X	X
KANSL1	KAT8 regulatory NSL complex subunit 1	other	X	X
KANSL2	KAT8 regulatory NSL complex subunit 2	other	X	X
KCTD20	potassium channel tetramerization domain containing 20	kinase	X	X
KDELR2	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 2	other	X	X
KDM3B	lysine (K)-specific demethylase 3B	other	X	X
KIAA0196	KIAA0196	other		X
KIAA0226	KIAA0226	other	X	X
KIAA0586	KIAA0586	other		X
KIAA2022	KIAA2022	enzyme	X	X
KIF13A	kinesin family member 13A	transporter	X	X
KLF14	Kruppel-like factor 14	other	X	X
KLHDC2	kelch domain containing 2	other	X	X
KLHL35	kelch-like family member 35	other	X	X
KMT2A	lysine (K)-specific methyltransferase 2A	transcription regulator	X	X
KPNA4	karyopherin alpha 4 (importin alpha 3)	transporter	X	X
LMLN	leishmanolysin-like (metallopeptidase M8 family)	peptidase	X	X
LOX	lysyl oxidase	enzyme	X	X
LRCH3	leucine-rich repeats and calponin homology (CH) domain containing 3	other	X	X
LRP5	low density lipoprotein receptor-related protein 5	transmem-brane receptor	X	X
LRR1	leucine rich repeat protein 1	enzyme	X	X
LRRC57	leucine rich repeat containing 57	other	X	X

LTN1	listerin E3 ubiquitin protein ligase 1	enzyme	X	X
MAML1	mastermind-like 1 (Drosophila)	transcription regulator	X	X
MAP3K2	mitogen-activated protein kinase kinase kinase 2	kinase	X	X
MAP4K5	mitogen-activated protein kinase kinase kinase kinase 5	kinase	X	X
MBTD1	mbt domain containing 1	other	X	X
MCM3AP	minichromosome maintenance complex component 3 associated protein	other	X	X
MDM4	MDM4, p53 regulator	other	X	X
METTL16	methyltransferase like 16	other	X	X
MFAP1	microfibrillar-associated protein 1	other	X	X
MGRN1	mahogunin ring finger 1, E3 ubiquitin protein ligase	enzyme	X	X
MKLN1	muskelin 1, intracellular mediator containing kelch motifs	other	X	X
MLEC	malectin	other	X	X
MPP5	membrane protein, palmitoylated 5 (MAGUK p55 subfamily member 5)	kinase	X	X
MRPL22	mitochondrial ribosomal protein L22	other	X	X
MRPS16	mitochondrial ribosomal protein S16	other		
MSRB3	methionine sulfoxide reductase B3	other	X	X
MTCH2	mitochondrial carrier 2	other		X
MYO6	myosin VI	other	X	X
MYO9A	myosin IXA	enzyme	X	X
MZT1	mitotic spindle organizing protein 1	other	X	X
NBR1	neighbor of BRCA1 gene 1	other	X	X
NCBP2	nuclear cap binding protein subunit 2, 20kDa	other		X
NCOA6	nuclear receptor coactivator 6	transcription regulator	X	X
NCOR1	nuclear receptor corepressor 1	transcription regulator	X	X
NDUFV2	NADH dehydrogenase (ubiquinone) flavoprotein 2, 24kDa	enzyme	X	X
NEMF	nuclear export mediator factor	other		X
NFAT5	nuclear factor of activated T-cells 5, tonicity-responsive	transcription regulator	X	X
NLK	nemo-like kinase	kinase	X	X
NLRC4	NLR family, CARD domain containing 4	other	X	X
NRD1	nardilysin (N-arginine dibasic convertase)	peptidase	X	X
NRG4	neuregulin 4	growth factor		
NSMCE2	non-SMC element 2, MMS21 homolog (S. cerevisiae)	enzyme	X	X
NUP153	nucleoporin 153kDa	transporter	X	X
P4HA1	prolyl 4-hydroxylase, alpha polypeptide I	enzyme	X	X
PAFAH1B1	platelet-activating factor acetylhydrolase 1b, regulatory subunit 1 (45kDa)	enzyme		X
PAIP1	poly(A) binding protein interacting protein 1	translation regulator	X	X

PAK2	p21 protein (Cdc42/Rac)-activated kinase 2	kinase	X	X
PANK3	pantothenate kinase 3	kinase	X	X
PAPPA	pregnancy-associated plasma protein A, pappalysin 1	peptidase	X	X
PCDH10	protocadherin 10	other	X	X
PDIA3	protein disulfide isomerase family A, member 3	peptidase	X	X
PDPR	pyruvate dehydrogenase phosphatase regulatory subunit	enzyme	X	X
PIBF1	progesterone immunomodulatory binding factor 1	other	X	X
PIK3C2B	phosphatidylinositol-4-phosphate 3-kinase, catalytic subunit type 2 beta	kinase	X	X
PIK3C3	phosphatidylinositol 3-kinase, catalytic subunit type 3	kinase	X	X
PKD2L1	polycystic kidney disease 2-like 1	ion channel		X
PLEK2	pleckstrin 2	other	X	X
PLEKHM2	pleckstrin homology domain containing, family M (with RUN domain) member 2	other	X	X
PNKD	paroxysmal nonkinesigenic dyskinesia	other	X	X
PNPT1	polyribonucleotide nucleotidyltransferase 1	enzyme		X
POLD3	polymerase (DNA-directed), delta 3, accessory subunit	transcription regulator	X	X
POLE2	polymerase (DNA directed), epsilon 2, accessory subunit	enzyme	X	X
POLR2B	polymerase (RNA) II (DNA directed) polypeptide B, 140kDa	enzyme	X	X
POP1	processing of precursor 1, ribonuclease P/MRP subunit (S. cerevisiae)	enzyme	X	X
PPP1CB	protein phosphatase 1, catalytic subunit, beta isozyme	phosphatase	X	X
PPP1R15B	protein phosphatase 1, regulatory subunit 15B	phosphatase	X	X
PRG4	proteoglycan 4	other	X	X
PRKAR1A	protein kinase, cAMP-dependent, regulatory, type I, alpha	kinase	X	X
PRRC2C	proline-rich coiled-coil 2C	other		X
PSEN1	presenilin 1	peptidase	X	X
PSMD1	proteasome (prosome, macropain) 26S subunit, non-ATPase, 1	other	X	X
PSMD14	proteasome (prosome, macropain) 26S subunit, non-ATPase, 14	peptidase	X	X
PVRL3	poliovirus receptor-related 3	other	X	X
RAB14	RAB14, member RAS 130ncogenes family	enzyme	X	X
RAC1	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)	enzyme	X	X
RANBP9	RAN binding protein 9	other	X	X
RARS	arginyl-tRNA synthetase	enzyme	X	X
RASGRP3	RAS guanyl releasing protein 3 (calcium and DAG-regulated)	other	X	X
RBM25	RNA binding motif protein 25	other	X	X
RBM5	RNA binding motif protein 5	other	X	X
RBM6	RNA binding motif protein 6	other	X	X

REST	RE1-silencing transcription factor	transcription regulator	X	X
RFC1	replication factor C (activator 1) 1, 145kDa	transcription regulator	X	X
RHBDL3	rhomboid, veinlet-like 3 (Drosophila)	peptidase	X	X
RLIM	ring finger protein, LIM domain interacting	enzyme	X	X
RMDN1	regulator of microtubule dynamics 1	other	X	X
RMND1	required for meiotic nuclear division 1 homolog (S. cerevisiae)	other	X	X
RNF10	ring finger protein 10	enzyme	X	X
RNF139	ring finger protein 139	enzyme	X	X
RNF169	ring finger protein 169	other		X
RNF216	ring finger protein 216	other	X	X
RPGRIP1	retinitis pigmentosa GTPase regulator interacting protein 1	other	X	X
RPLP0	ribosomal protein, large, P0	other	X	X
RPS3	ribosomal protein S3	enzyme		X
RSF1	remodeling and spacing factor 1	transcription regulator	X	X
RSPRY1	ring finger and SPRY domain containing 1	other	X	X
RUFY2	RUN and FYVE domain containing 2	other	X	X
SCD5	stearoyl-CoA desaturase 5	enzyme	X	X
SEC23A	Sec23 homolog A (S. cerevisiae)	transporter	X	X
SEC31A	SEC31 homolog A (S. cerevisiae)	other	X	X
SENP2	SUMO1/sentrin/SMT3 specific peptidase 2	peptidase	X	X
SENP5	SUMO1/sentrin specific peptidase 5	peptidase	X	X
SENP6	SUMO1/sentrin specific peptidase 6	peptidase	X	X
SERF2	small EDRK-rich factor 2	other	X	X
SERPINH1	serpin peptidase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1)	other	X	X
SETD2	SET domain containing 2	enzyme	X	X
SGCB	sarcoglycan, beta (43kDa dystrophin-associated glycoprotein)	other	X	X
SIRT4	sirtuin 4	enzyme	X	X
SIRT5	sirtuin 5	enzyme		X
SLC25A17	solute carrier family 25 (mitochondrial carrier; peroxisomal membrane protein, 34kDa), member 17	transporter	X	X
SMEK2	SMEK homolog 2, suppressor of mek1 (Dictyostelium)	other	X	X
SMG6	SMG6 nonsense mediated mRNA decay factor	enzyme	X	X
SNX1	sorting nexin 1	transporter		X
SPAG9	sperm associated antigen 9	other	X	X
SPATA18	spermatogenesis associated 18	other	X	X
SPATS2	spermatogenesis associated, serine-rich 2	other	X	X
SPDYA	speedy/RINGO cell cycle regulator family member A	other	X	X

SQLE	squalene epoxidase	enzyme	X	X
SQSTM1	sequestosome 1	transcription regulator		X
SRFBP1	serum response factor binding protein 1	other	X	X
SRSF3	serine/arginine-rich splicing factor 3	other		X
ST7	suppression of tumorigenicity 7	other		X
STAT1	signal transducer and activator of transcription 1, 91kDa	transcription regulator	X	X
STAT4	signal transducer and activator of transcription 4	transcription regulator	X	X
STK3	serine/threonine kinase 3	kinase	X	X
STOX1	storkhead box 1	other		X
STRN	striatin, calmodulin binding protein	other		X
SUGCT	succinyl-CoA:glutarate-CoA transferase	enzyme	X	X
SUV420H1	suppressor of variegation 4-20 homolog 1 (Drosophila)	enzyme	X	X
SYNE1	spectrin repeat containing, nuclear envelope 1	other	X	X
TACC1	transforming, acidic coiled-coil containing protein 1	other	X	X
TAF2	TAF2 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 150kDa	transcription regulator	X	X
TATDN1	TatD Dnase domain containing 1	other	X	X
TBC1D23	TBC1 domain family, member 23	other	X	X
TBK1	TANK-binding kinase 1	kinase	X	X
TGM5	transglutaminase 5	enzyme	X	X
TGM7	transglutaminase 7	enzyme		X
THRB	thyroid hormone receptor, beta	ligand-dependent nuclear receptor	X	X
TLK1	tousled-like kinase 1	kinase	X	X
TMCO3	transmembrane and coiled-coil domains 3	transporter	X	X
TMEM44	transmembrane protein 44	other	X	X
TNPO1	transportin 1	transporter	X	X
TOMM70A	translocase of outer mitochondrial membrane 70 homolog A (S. cerevisiae)	transporter	X	X
TP53BP1	tumor protein p53 binding protein 1	transcription regulator	X	X
TPR	translocated promoter region, nuclear basket protein	transporter	X	X
TRIM37	tripartite motif containing 37	enzyme	X	X
TRMT61B	tRNA methyltransferase 61 homolog B (S. cerevisiae)	enzyme		X
TTC19	tetratricopeptide repeat domain 19	other		X
UBE2Q2	ubiquitin-conjugating enzyme E2Q family member 2	enzyme	X	X
UBE3C	ubiquitin protein ligase E3C	enzyme	X	X
UBE4A	ubiquitination factor E4A	enzyme	X	X
UBR2	ubiquitin protein ligase E3 component n-recognin 2	enzyme	X	X

USP15	ubiquitin specific peptidase 15	peptidase	X	X
USP16	ubiquitin specific peptidase 16	peptidase	X	X
USP3	ubiquitin specific peptidase 3	peptidase		X
UTP18	UTP18 small subunit (SSU) processome component homolog (yeast)	other		X
VIT	vitrin	other	X	X
VPS13C	vacuolar protein sorting 13 homolog C (S. cerevisiae)	other	X	X
VPS13D	vacuolar protein sorting 13 homolog D (S. cerevisiae)	other	X	X
WDR19	WD repeat domain 19	other	X	X
WDR76	WD repeat domain 76	other	X	X
WWC1	WW and C2 domain containing 1	transcription regulator	X	X
WWP1	WW domain containing E3 ubiquitin protein ligase 1	enzyme	X	X
WWP2	WW domain containing E3 ubiquitin protein ligase 2	enzyme	X	X
XPNPEP3	X-prolyl aminopeptidase (aminopeptidase P) 3, putative	peptidase	X	X
XPOT	exportin, tRNA	other	X	X
XRCC5	X-ray repair complementing defective repair in Chinese hamster cells 5 (double-strand-break rejoining)	enzyme	X	X
YBEY	ybeY metallopeptidase (putative)	other	X	X
YIPF4	Yip1 domain family, member 4	other	X	X
YTHDF2	YTH domain family, member 2	other		X
ZBTB11	zinc finger and BTB domain containing 11	other	X	X
ZBTB2	zinc finger and BTB domain containing 2	other	X	X
ZDHHC5	zinc finger, DHHC-type containing 5	enzyme	X	X
ZFYVE1	zinc finger, FYVE domain containing 1	other	X	X
ZNF474	zinc finger protein 474	other		X
ZNF572	zinc finger protein 572	other	X	X
ZSCAN29	zinc finger and SCAN domain containing 29	other	X	X
ZZEF1	zinc finger, ZZ-type with EF-hand domain 1	other	X	X

Supplementary Table 14: Association of radiation-related DSB hotspots with neoplasia and cancer in NHEJ-impaired NHDF-A in p6

Symbol	Entrez Gene Name	Type(s)	Neoplasia	Cancer
GRM8	glutamate receptor, metabotropic 8	G-protein coupled receptor	X	X
OR11H12 (includes others)	olfactory receptor, family 11, subfamily H, member 12	other	X	
TP53BP2	tumor protein p53 binding protein 2	other	X	X
BCOR	BCL6 corepressor	transcription regulator		X

ALCAM	activated leukocyte cell adhesion molecule	other	X	X
SCN11A	sodium channel, voltage-gated, type XI, alpha subunit	ion channel	X	
HEG1	heart development protein with EGF-like domains 1	other	X	

Supplementary Table 15: Association of radiation-related DSB hotspots with neoplasia and cancer in NHEJ-impaired NHDF-A in p8

Symbol	Entrez Gene Name	Type(s)	Neoplasia	Cancer
MSH6	mutS homolog 6	enzyme	X	X
BARD1	BRCA1 associated RING domain 1	transcription regulator	X	X
GDNF	glial cell derived neurotrophic factor	growth factor		X
LTA4H	leukotriene A4 hydrolase	enzyme		X
TMEM220	transmembrane protein 220	other		X
C7orf61	chromosome 7 open reading frame 61	other		X
H3F3C	H3 histone, family 3C	other		X
ZNF433	zinc finger protein 433	other		X
EPHA6	EPH receptor A6	kinase		X

Supplementary Table 16: Distribution of therapy-related DSB hotspots

Cell Type	Type of DSB Hotspot	# Hotspots	# Associated Genes	#Hotspots intragenic	#Hotspots associated to Cancer	#Hotspots associated to Neoplasia
NHDF-A	Radiation-related	54	88	34	67	77
	Doxorubicin-related	171	389	168	266	314
NHDF-A +NHEJi	Radiation-related p4	0	0	0	0	0
	Radiation-related p6	25	40	14	6	4
	Radiation-related p8	98	180	56	2	8

Supplementary Table 17: Average number of histones per DSB hotspot. Nat: Naturally-occurring DSB hotspot; Overlap: Overlapping DSB hotspot; IR: Radiation-related DSB hotspot.

Histone Modification	NHDF-A								
	p4			p6			p8		
	Nat	Overlap	IR	Nat	Overlap	IR	Nat	Overlap	IR
H3K4me1	3.2	6.2	3.2	3.0	8.0	3.9	3.8	5.9	4.8
H3K4me2	2.7	3.8	2.3	3.0	6.6	3.7	2.3	4.3	4.3
H3K4me3	1.9	4.0	4.0	3.0	4.9	3.5	3.0	3.4	3.5
H3K9ac	3.5	3.6	3.3	3.0	7.7	3.7	4.0	5.1	4.7
H3K27ac	3.0	3.8	2.8	3.0	7.6	5.1	2.5	5.4	5.7
H3K36me3	2.6	4.4	5.5	3.0	5.5	4.1	2.4	5.0	4.3
H4K20me1	2.8	4.5	3.3	3.0	7.5	3.6	3.0	5.9	4.2
H3K9me3	3.8	3.5	3.0	3.0	5.8	4.0	3.0	4.7	4.1
H3K27me3	2.7	4.1	1.5	3.0	5.7	3.5	6.0	4.4	4.4
H3K79me2	2.5	5.3	3.2	3.0	7.7	3.8	2.5	5.7	4.3

5.3   Figure Index

5.3.1   Figures

5.3.2   Supplementary Figures

5.4   Table Index

5.4.1   Tables

5.4.2   Supplementary Tables

5.5   Zusammenfassung

DNA Doppelstrangbrüche (DSB) können sowohl natürlich als auch nach Bestrahlung auftreten und stellen eine große Gefahr für die Stabilität des Genoms und Zellviabilität dar. Wird ein DSB nicht effizient repariert, können Veränderungen im Genom wie z.B. Deletionen und Vervielfältigungen auftreten, die ein Markenzeichen von Krebszellen sind. Ist der DNA Schaden allerdings zu groß, kann die Zelle Apoptose induzieren, was das Ziel der Strahlentherapie zur Krebsbehandlung ist. Derzeitige Methoden zur Bestimmung von DSB basieren vor allem auf der Markierung von DNA Reparaturproteinen wie ATM (Ataxia telangiectasia mutated), 53BP1 (53 binding protein 1) und γH2AX (phosphoryliertes Histon H2AX) mittels Fluoreszenz-markierten Antikörpern. Diese Proteine bilden innerhalb weniger Minuten nach der DSB Entstehung mikroskopisch-sichtbare Zentren an der DSB Stelle, sogenannte Strahlen-induzierte Zentren (RIF). Obwohl diese Methode Erkenntnisse über die DSB Entstehung und Reparaturkinetik geliefert hat, ist sie limitiert. So kann keine Aussage über die genaue Position des Strangbruchs getroffen werden, da die RIF auf mehrere Megabasen um die DSB Stelle herum vergrößern. Außerdem verschwinden die RIF innerhalb von 24 Stunden nach Bestrahlung, und ermöglichen somit keine weitere Analyse der reparierten DSB Stellen in überlebenden Zellen. Ebenso erlaubt die Immunfärbung keine Aussage über einen möglichen Zusammenhang zwischen der DSB Stelle, der Reparatur und des Schicksals der Zelle. Daher gibt es kaum Informationen darüber, wie Strahlen-induzierte DSB im Genom verteilt sind, wie solche Schäden zur Radioresistenz führen sowie die Entwicklung von einer normalen Zelle hin zu einer Tumorzelle und die Krebsentstehung beeinflussen. Deshalb stellen wir die Hypothese auf, dass DSB im Genom nicht zufällig verteilt sind und dass die Analyse der genom-weiten DSB Verteilung nach Bestrahlung in überlebenden Zellen auf Einzelnukleotidebene unser Verständnis über die Mechanismen, die zur Bestrahlungsresistenz führen, verbessert. Um dieses Ziel zu erreichen, wurden in dieser Arbeit neue Methodenansätze basierend auf der Beobachtung entwickelt, dass Non-Homologous End Joining (NHEJ) Reparatur zum Einbau von DNA Molekülen ins Genom an DSB Stellen führen kann. Für die DSB Markierung wurden daher Integrase-defiziente lentivirale Vektoren (IDLV) in Zellen eingebracht, die nach Bestrahlung an transient-auftretenden DSB Stellen ins Genom stabil eingebaut werden. Diese Integrationsstellen wurden anschließend durch LAM-PCR und Sequenzierung amplifiziert und identifiziert. Insgesamt wurden somit mehr als 20,000 DSB Stellen im Genom bestrahlter Zellen identifiziert und mit natürlich-auftretenden und Doxorubicin-induzierten DSB Stellen verglichen. Die DSB Verteilung wurde mit der Transkriptionsaktivität, verschiedenen Chromatin-Bereichen und Genfunktionen verglichen. Die Analyse ergab, dass in überlebenden Zellpopulationen das DSB-Verteilungsmuster nicht zufällig ist. Therapie-induzierte DSB wurden vermehrt in kleinen genomischen Bereichen, sogenannten Hotspots, die in Grenzregionen von Eu- und Heterochromatin und in oder in der Nähe von Onkogenen, Tumorsuppressorgenen und Enhancer Regionen liegen, nachgewiesen. Die Ergebnisse dieser Arbeit zeigen, dass Strahlen-induzierte Mutationen in überlebenden Zellpopulationen in bevorzugten Bereichen auftreten oder Zellen mit Mutationen in solchen Bereichen selektiert werden. Diese Hotspots können eventuell die Wahrscheinlichkeit für die Entstehung von Therapieresistenzen erhöhen.

5.6   Summary

DNA double strand breaks (DSB) can occur naturally as well as after irradiation and pose a serious threat to genome stability and cell viability. Failure to efficiently repair DSB can result in genomic alterations including deletions and amplifications, which are hallmarks of cancer. If the DNA damage is too severe, apoptosis is induced, which is the main goal in cancer radiotherapy. Current methods for detection of DSB sites following irradiation are mainly based on the immunostaining of DNA repair proteins such ATM (Ataxia telangiectasia mutated), 53BP1 (53 binding protein 1), and γH2AX (phosphorylated histone H2AX) by fluorescent-labeled antibodies. These repair proteins form microscopically-visible foci at the DSB site referred to as radiation-induced foci (RIF) within minutes after DNA damage induction. Even though immunostaining has provided valuable insights into the mechanisms and kinetics of DSB induction and repair, this method has several limitations. Since RIF span up to several megabases around the DSB site, the precise location of the DSB cannot be determined. Moreover, DNA repair proteins disassemble from the DSB site within 24 hours after DSB induction and do not enable DSB site analysis in radiation-surviving cells. Furthermore, immunostaining does not provide any molecular information on DSB localization, repair outcome and cell fate decision. Hence, up to now, there is little information on how radiation-induced and repaired DSB sites are distributed, how radiation-induced DNA damage is being survived and how these damages induce radiation-resistance, cell transformation and carcinogenesis. Thus, we hypothesize that the distribution of radiation-induced DSB sites may not be random and that analyzing the genome-wide distribution of radiation-induced DSB in surviving cells at single-nucleotide resolution will improve our understanding of the mechanisms of radioresistance. In order to achieve this goal, new methodological approaches were developed based on the observation that DNA repair by non-homologous end joining (NHEJ) can lead to the genomic integration of episomal unrelated DNA molecules at the DSB sites. In this thesis, integrase-deficient lentiviral vectors (IDLV) were introduced into target cells and became stably incorporated into the genome at radiation-induced DSB sites. The integration sites were subsequently amplified and identified by LAM-PCR and sequencing. In total, more than 20,000 DSB sites in irradiated and expanded cancer cell lines and primary fibroblasts were obtained and compared to naturally-occurring and doxorubicin-induced DSB. Using this DSB data set, the DSB distribution with respect to transcriptional activity, chromatin structures, as well as gene function and networks was analyzed. In radiation-survivor cells the distribution of DSB was non-random. DSB sites appeared more frequently in genes with specific functions, in genes regulated by specific regulatory networks and in accessible genomic regions. Radiation-induced DSB frequently accumulated in genomic hotspots, which were identified at eu- and heterochromatin boundaries as well as near oncogenes, tumor-suppressors and enhancer elements. The results presented in this thesis indicate that radiation-induced genomic mutations preferably occur or are selected for specific genomic regions in surviving cells. These genomic hotspots of induced DNA damage may increase the likelihood of radioresistance mechanisms and therapy-induced carcinogenesis.

5.7 Abbreviations

Abbreviation	Meaning
3'	3 prime
5'	5 prime
53BP1	p53-binding protein 1
A431	Epidermoid carcinoma cells
A549	Adenocarcinomic human alveolar basal epithelial cells
ac	Acetylated
ALD	X-linked adrenoleukodystrophy
AP	Abasic Sites
AT	Ataxia telangiectasia
ATM	Ataxia-Telangiectasia Mutated
ATP	Adenosine triphosphate
ATR	Ataxia- and Rad3-related
BER	Base Excision Repair
Bio	Biotinylated
BLAST	Basic Local Alignment Search Tool
BLAT	BLAST-like alignment tool
BLS	Bloom's Syndrome
bp	Base pair
BRCA1/2	Breast Cancer 1/2
BSA	Bovine Serum Albumin
Cas9	CRISPR associated nuclease 9
CAT	Chloramphenicol Acetyltransferase
CCR5	C-C chemokine receptor type 5
CD4	Cluster of differentiaton 4
cDNA	complementary DNA
Chr	Chromosome
CMV	Cytomegalovirus
CpG	Cytosine and guanine separated by a phosphate
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CtIP	CtBP-interacting protein
CXCR4	C-X-C chemokine receptor type 4
D64V	Aspartic Acid at amino acid position 64 changed to Valine
DAPI	4',6-diamidino-2-phenylindole
ddC	Dideoxycytidine
DDR	DNA damage response
DHS	DnaseI hypersensitive site
DKFZ	German Cancer Research Center
DMEM	Dulbecco's Modified Eagle Medium
DNA	Dioxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase, catalytic subunit
DKC	Dykeratosis congenita
dNTP	Deoxynucleoside triphosphate

Dox	Doxorubicin
DPBS	Dulbecco's Phosphate Buffered Saline
DPBST	Dulbecco's Phosphate Buffered Saline with 0.2% Tween-20
DSB	DNA double strand break
dsDNA	double-stranded DNA
dUTP	Deoxyuridine triphosphate
EC	Euchromatin
EDTA	Ethylenediaminetetraacetic acid
eGFP	Enhanced green fluorescent protein
env	Envelope
ES	Embryonic stem cells
Expo	Exponential
FA	Fanconi anemia
FACS	Fluorescent Activated Cell Sorting
FCS	Fetal calf serum
fwd	Forward
G1/G0	G1/G0 cell cycle phase
G2/M	G2/mitosis cell cycle checkpoint
g	gravitation
G segment	Gate segment
gag	Group specific antigen
Gp41/120	Glycoprotein 41/120
Gy	Gray
h	Hour
H2AX	Histone 2A variant X
H_2O bidest	Double-distilled water
HC	Heterochromatin
HEK	Human Embryonic Kidney cells
HeLa	Henrietta Lacks
HISAP	High-throughput Insertion Site Analysis Pipeline
HIV	Human Immundeficiency Virus
HIT	Heidelberg Ion Therapy Center
HNSCC	Head and neck squamous cell carcinomas
HR	Homologous Recombination
HTLV	Human T-cell lymphotropic virus
ICL	Interstrand crosslink
ICLV	Integrase-Competent Lentivirus
IDLV	Integrase-Deficient Lentivirus
IgG	Immunglobuline G
IN	Integrase enzyme
IPA	Ingenuity Pathway Analysis
IR	Ionizing radiation
IMDM	Iscove's Modified Dulbecco's Medium
kb	kilobases
kDa	kilo Dalton

keV	kilo electron Volt
l	Liter
LAM-DST	Linker-Amplification-Mediated DSB-Trapping
LAM-PCR	Linear Amplification mediated-PCR
LB	Luria-Bertani broth
LC	Linker cassette
LD	Lethal Dose
LET	Linear Energy Transfer
LTR	Long terminal repeat
LV	Lentivirus
m	Milli (10^{-3})
M	Molar
Mb	Megabase
me1/2/3	Mono-/di-/tri-methylated
min	Minute
miRNA	MicroRNA
ml	Milliliter
Mlu	*Micrococcus luteus*
MMEJ	Microhomology-mediated end joining
MMR	Mismatch repair
Mnase	Micrococcal nuclease
MOI	Multiplicity of infection
MPC	Magnetic particle concentrator
Mre11	Meiotic recombination 11
mRNA	Messenger RNA
Mse	*Micrococcus species*
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide
MYO	Myoglobin
n	Nano (10^{-9})
NBS	Nijmegen breakage syndrome
NCBI	National Center for Biotechnology Information
NCT	National Center for Tumor Diseases
NER	Nucleotide Excision Repair
ng	Nanogram
NGS	Next generation sequencing
NHDF-A	Adult normal human dermal fibroblasts
NHEJ (c/a)	Classical/alternative Non-Homologous End Joining
nm	Nanometer
nr	Non-restrictive
OH	Hydroxyl group
P	Phosphate residue
p	Passage
PARP1	Poly(ADP-ribose)-Polymerase 1
PBS	Primer binding sites
PC3	Human prostate cancer cell line

PCR	Polymerase Chain Reaction
PD	Pyrimidine dimers
PEI	Poly-ethylene Inositol
Pen/Strep	Penicillin/Streptomycin
PFA	Paraformaldehyde
PGK	Phosphoglycerate Kinase 1 promoter
PI	Propidium Iodide
pol	DNA polymerase
PPT	Polypurin tract
pUC	plasmid University of California
q-RT-PCR	Quantitative real time-polymerase chain reaction
R	Redundant
RBE	Relative Biological Effect
RCL	Recombination competent lentivirus
RECQ	Recombination Q
RefSeq	Reference sequence
Rep	Replicate
rev	Reverse
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RNAi	RNA interference
ROS	Reactive Oxygen Species
RRE	Rev responsive element
RSV	Rous sarcoma virus
rpm	rotations per minute
RT	Room Temperature
RTS	Rothmund Thomson Syndrome
S	S cell cycle phase
S.O.C.	Super Optimal broth with Catabolite repression
sec	Seconds
Ser	Serine
SIN	Self-inactivating
SSB	Single-strand break
ssDNA	single-stranded DNA
T segment	Transport segment
TALEN	Transcription activator-like effector nucleases
Taq	*Thermus aquaticus*
TAR	Trans-activation response element
TBE	Tris-borate-EDTA
Tdt	Terminal deoxynucleotidyl transferase
TE	Tris-EDTA
TFBS	Transcription factor binding site
Thr	Threonine
TOP1/2	Topoisomerase 1/2
Tsp	*Thermus species*

TSS	Transcription start site
TUNEL	Tdt-mediated dUTP nick end labeling
U3/5	Unique 3'/5'
U87	Human primary glioblastoma cells
UC	University of California
UCSC	University of California, Santa Cruz
UV	Ultraviolet light
VSV-G	Vesicular Stomatitis Virus – protein G
WPRE	Woodchuck hepatitis virus post-transcriptional regulatory element
w/o	Without
WS	(Alternative) Werner Syndrome
XLF	XRCC4-like factor
XP	Xeroderma pigmentosum
XRCC4	X-ray repair cross-complementing protein 4
ZFN	Zinkfinger nuclease
α	Alpha
β	Beta
γ	Gamma
μ	Micro (10^{-6})
μl	Microliter
μM	Micromolar
ψ	Psi packaging signal

5.8   References

1.   Ciccia, A. and S.J. Elledge, *The DNA damage response: making it safe to play with knives.* Mol Cell, 2010. **40**(2): p. 179-204.
2.   Jackson, S.P. and J. Bartek, *The DNA-damage response in human biology and disease.* Nature, 2009. **461**(7267): p. 1071-8.
3.   Wolters, S. and B. Schumacher, *Genome maintenance and transcription integrity in aging and disease.* Front Genet, 2013. **4**: p. 19.
4.   Kumar, R., et al., *Chromatin modifications and the DNA damage response to ionizing radiation.* Front Oncol, 2012. **2**: p. 214.
5.   Cleaver, J.E., *Defective repair replication of DNA in xeroderma pigmentosum.* Nature, 1968. **218**(5142): p. 652-6.
6.   Cleaver, J.E., *Xeroderma pigmentosum: a human disease in which an initial stage of DNA repair is defective.* Proc Natl Acad Sci U S A, 1969. **63**(2): p. 428-35.
7.   Madhusudan, S. and D.M. Wilson III, *DNA Repair and Cancer: From Bench to Clinic*. 2013, Boca Raton, FL, USA: CRC Press Taylor & Francis Group. 718.
8.   Huertas, P. and S.P. Jackson, *Human CtIP mediates cell cycle control of DNA end resection and double strand break repair.* J Biol Chem, 2009. **284**(14): p. 9558-65.
9.   Chapman, J.R., M.R. Taylor, and S.J. Boulton, *Playing the end game: DNA double-strand break repair pathway choice.* Mol Cell, 2012. **47**(4): p. 497-510.
10.   Botuyan, M.V., et al., *Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair.* Cell, 2006. **127**(7): p. 1361-73.
11.   Sartori, A.A., et al., *Human CtIP promotes DNA end resection.* Nature, 2007. **450**(7169): p. 509-14.
12.   Yu, X., et al., *BRCA1 ubiquitinates its phosphorylation-dependent binding partner CtIP.* Genes Dev, 2006. **20**(13): p. 1721-6.
13.   Lukas, J., C. Lukas, and J. Bartek, *More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance.* Nat Cell Biol, 2011. **13**(10): p. 1161-9.
14.   Cowell, I.G., et al., *gammaH2AX foci form preferentially in euchromatin after ionising-radiation.* PLoS One, 2007. **2**(10): p. e1057.
15.   Goodarzi, A.A., P. Jeggo, and M. Lobrich, *The influence of heterochromatin on DNA double strand break repair: Getting the strong, silent type to relax.* DNA Repair (Amst), 2010. **9**(12): p. 1273-82.
16.   Jakob, B., et al., *DNA double-strand breaks in heterochromatin elicit fast repair protein recruitment, histone H2AX phosphorylation and relocation to euchromatin.* Nucleic Acids Res, 2011. **39**(15): p. 6489-99.
17.   Reynolds, P., et al., *The dynamics of Ku70/80 and DNA-PKcs at DSBs induced by ionizing radiation is dependent on the complexity of damage.* Nucleic Acids Res, 2012. **40**(21): p. 10821-31.
18.   Lomax, M.E., L.K. Folkes, and P. O'Neill, *Biological consequences of radiation-induced DNA damage: relevance to radiotherapy.* Clin Oncol (R Coll Radiol), 2013. **25**(10): p. 578-85.
19.   Hoeijmakers, J.H., *DNA damage, aging, and cancer.* N Engl J Med, 2009. **361**(15): p. 1475-85.
20.   Mimori, T. and J.A. Hardin, *Mechanism of interaction between Ku protein and DNA.* J Biol Chem, 1986. **261**(22): p. 10375-9.
21.   Getts, R.C. and T.D. Stamato, *Absence of a Ku-like DNA end binding activity in the xrs double-strand DNA repair-deficient mutant.* J Biol Chem, 1994. **269**(23): p. 15981-4.
22.   Deriano, L. and D.B. Roth, *Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage.* Annu Rev Genet, 2013. **47**: p. 433-55.
23.   Shiloh, Y. and Y. Ziv, *The ATM protein kinase: regulating the cellular response to genotoxic stress, and more.* Nat Rev Mol Cell Biol, 2013. **14**(4): p. 197-210.
24.   Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors.* Nature, 2005. **437**(7057): p. 376-80.
25.   Bentley, D.R., et al., *Accurate whole human genome sequencing using reversible terminator chemistry.* Nature, 2008. **456**(7218): p. 53-9.
26.   Metzker, M.L., *Sequencing technologies - the next generation.* Nat Rev Genet, 2010. **11**(1): p. 31-46.
27.   Hall, E.J. and A.J. Giaccia, *Radiobiology for the Radiobiologist*. 5th edition ed. 2000, Philadelphia, PA, USA: Lippincott Williams&Wilki.
28.   Denekamp, J., T. Waites, and J.F. Fowler, *Predicting realistic RBE values for clinically relevant radiotherapy schedules.* Int J Radiat Biol, 1997. **71**(6): p. 681-94.
29.   Goodhead, D.T., *Initial events in the cellular effects of ionizing radiations: clustered damage in DNA.* Int J Radiat Biol, 1994. **65**(1): p. 7-17.

30.   Allen, C., et al., *Heavy charged particle radiobiology: using enhanced biological effectiveness and improved beam focusing to advance cancer therapy.* Mutat Res, 2011. **711**(1-2): p. 150-7.

31.   Asaithamby, A., B. Hu, and D.J. Chen, *Unrepaired clustered DNA lesions induce chromosome breakage in human cells.* Proc Natl Acad Sci U S A, 2011. **108**(20): p. 8293-8.

32.   Urushibara, A., et al., *DNA damage induced by the direct effect of He ion particles.* Radiat Prot Dosimetry, 2006. **122**(1-4): p. 163-5.

33.   Ward, J.F., *Some biochemical consequences of the spatial distribution of ionizing radiation-produced free radicals.* Radiat Res, 1981. **86**(2): p. 185-95.

34.   Anderson, J.A., et al., *Participation of DNA-PKcs in DSB repair after exposure to high- and low-LET radiation.* Radiat Res, 2010. **174**(2): p. 195-205.

35.   Harper, J.V., J.A. Anderson, and P. O'Neill, *Radiation induced DNA DSBs: Contribution from stalled replication forks?* DNA Repair (Amst), 2010. **9**(8): p. 907-13.

36.   Groth, P., et al., *Homologous recombination repairs secondary replication induced DNA double-strand breaks after ionizing radiation.* Nucleic Acids Res, 2012. **40**(14): p. 6585-94.

37.   Pommier, Y., et al., *DNA topoisomerases and their poisoning by anticancer and antibacterial drugs.* Chem Biol, 2010. **17**(5): p. 421-33.

38.   Vos, S.M., et al., *All tangled up: how cells direct, manage and exploit topoisomerase function.* Nat Rev Mol Cell Biol, 2011. **12**(12): p. 827-41.

39.   Connelly, J.C. and D.R. Leach, *Repair of DNA covalently linked to protein.* Mol Cell, 2004. **13**(3): p. 307-16.

40.   Quennet, V., et al., *CtIP and MRN promote non-homologous end-joining of etoposide-induced DNA double-strand breaks in G1.* Nucleic Acids Res, 2011. **39**(6): p. 2144-52.

41.   Gomez-Herreros, F., et al., *TDP2-dependent non-homologous end-joining protects against topoisomerase II-induced DNA breaks and genome instability in cells and in vivo.* PLoS Genet, 2013. **9**(3): p. e1003226.

42.   Roy, K., et al., *Hypoxia relieves X-ray-induced delayed effects in normal human embryo cells.* Radiat Res, 2000. **154**(6): p. 659-66.

43.   Kim, G.J., G.M. Fiskum, and W.F. Morgan, *A role for mitochondrial dysfunction in perpetuating radiation-induced genomic instability.* Cancer Res, 2006. **66**(21): p. 10377-83.

44.   Suzuki, K., Yamauchi, M., Suzuki, M., Oka, Y., Yamashita, S., *Involvement of Non-Homologous End-Joining in Radiation-Induced Genomic Instability*, in *Selected Topics in DNA Repair*, C. Chen, Editor. 2011, InTech.

45.   Suzuki, K., *Multistep nature of X-ray-induced neoplastic transformation in mammalian cells: genetic alterations and instability.* J Radiat Res, 1997. **38**(1): p. 55-63.

46.   Goldberg, Z., *Clinical implications of radiation-induced genomic instability.* Oncogene, 2003. **22**(45): p. 7011-7.

47.   Suzuki, K., S. Kodama, and M. Watanabe, *Role of Ku80-dependent end-joining in delayed genomic instability in mammalian cells surviving ionizing radiation.* Mutat Res, 2010. **683**(1-2): p. 29-34.

48.   Suzuki, K., et al., *Radiation-induced DNA damage and delayed induced genomic instability.* Oncogene, 2003. **22**(45): p. 6988-93.

49.   Limoli, C.L., et al., *Differential induction of chromosomal instability by DNA strand-breaking agents.* Cancer Res, 1997. **57**(18): p. 4048-56.

50.   Degrassi, F., M. Fiore, and F. Palitti, *Chromosomal aberrations and genomic instability induced by topoisomerase-targeted antitumour drugs.* Curr Med Chem Anticancer Agents, 2004. **4**(4): p. 317-25.

51.   Rous, P., *A Sarcoma of the Fowl Transmissible by an Agent Separable from the Tumor Cells.* J Exp Med, 1911. **13**(4): p. 397-411.

52.   Temin, H.M. and S. Mizutani, *RNA-dependent DNA polymerase in virions of Rous sarcoma virus.* Nature, 1970. **226**(5252): p. 1211-3.

53.   Baltimore, D., *RNA-dependent DNA polymerase in virions of RNA tumour viruses.* Nature, 1970. **226**(5252): p. 1209-11.

54.   Varmus, H.E., J.M. Bishop, and P.K. Vogt, *Appearance of virus-specific DNA in mammalian cells following transformation by Rous sarcoma virus.* J Mol Biol, 1973. **74**(4): p. 613-26.

55.   Gallo, R.C., *Growth of human normal and leukemic T cells: T-cell growth factor (TCGF) and the isolation of a new class of RNA tumor viruses (HTLV).* Blood Cells, 1981. **7**(2): p. 313-29.

56.   Barre-Sinoussi, F., et al., *Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).* Science, 1983. **220**(4599): p. 868-71.

57.   Coffin, J.M., S.H. Hughes, and H.E. Varmus, *The Interactions of Retroviruses and their Hosts*, in *Retroviruses*, J.M. Coffin, S.H. Hughes, and H.E. Varmus, Editors. 1997: Cold Spring Harbor (NY).

58.   Sakuma, T., M.A. Barry, and Y. Ikeda, *Lentiviral vectors: basic to translational.* Biochem J, 2012. **443**(3): p. 603-18.

59.   Katz, R.A., et al., *The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro.* Cell, 1990. **63**(1): p. 87-95.

60.   Coiras, M., et al., *Understanding HIV-1 latency provides clues for the eradication of long-term reservoirs.* Nat Rev Microbiol, 2009. **7**(11): p. 798-812.

61.   Mitchell, R.S., et al., *Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences.* PLoS Biol, 2004. **2**(8): p. E234.

62.   Wu, X., et al., *Transcription start regions in the human genome are favored targets for MLV integration.* Science, 2003. **300**(5626): p. 1749-51.

63.   Schroder, A.R., et al., *HIV-1 integration in the human genome favors active genes and local hotspots.* Cell, 2002. **110**(4): p. 521-9.

64.   Zufferey, R., et al., *Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery.* J Virol, 1998. **72**(12): p. 9873-80.

65.   Naldini, L., et al., *Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector.* Proc Natl Acad Sci U S A, 1996. **93**(21): p. 11382-8.

66.   Leavitt, A.D., et al., *Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection.* J Virol, 1996. **70**(2): p. 721-8.

67.   Koyama, T., et al., *DNA damage enhances integration of HIV-1 into macrophages by overcoming integrase inhibition.* Retrovirology, 2013. **10**: p. 21.

68.   Mardis, E.R., *The impact of next-generation sequencing technology on genetics.* Trends Genet, 2008. **24**(3): p. 133-41.

69.   Giard, D.J., et al., *In vitro cultivation of human tumors: establishment of cell lines derived from a series of solid tumors.* J Natl Cancer Inst, 1973. **51**(5): p. 1417-23.

70.   DuBridge, R.B., et al., *Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system.* Mol Cell Biol, 1987. **7**(1): p. 379-87.

71.   Kaighn, M.E., et al., *Establishment and characterization of a human prostatic carcinoma cell line (PC-3).* Invest Urol, 1979. **17**(1): p. 16-23.

72.   Beckman, G., et al., *G-6-Pd and Pgm Phenotypes of 16 Continuous Human Tumor Cell Lines - Evidence against Cross-Contamination and Contamination by Hela Cells.* Human Heredity, 1971. **21**(3): p. 238-&.

73.   Mosmann, T., *Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays.* J Immunol Methods, 1983. **65**(1-2): p. 55-63.

74.   Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.

75.   Dupre, A., et al., *A forward chemical genetic screen reveals an inhibitor of the Mre11-Rad50-Nbs1 complex.* Nat Chem Biol, 2008. **4**(2): p. 119-25.

76.   Sharungbam, G.D., et al., *Identification of stable endogenous control genes for transcriptional profiling of photon, proton and carbon-ion irradiated cells.* Radiat Oncol, 2012. **7**: p. 70.

77.   Mullis, K., et al., *Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction.* Cold Spring Harb Symp Quant Biol, 1986. **51 Pt 1**: p. 263-73.

78.   Schmidt, M., et al., *High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR).* Nat Methods, 2007. **4**(12): p. 1051-7.

79.   Schmidt, M., et al., *Detection of retroviral integration sites by linear amplification-mediated PCR and tracking of individual integration clones in different samples.* Methods Mol Biol, 2009. **506**: p. 363-72.

80.   Clark, J.M., *Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases.* Nucleic Acids Res, 1988. **16**(20): p. 9677-86.

81.   Kent, W.J., *BLAT--the BLAST-like alignment tool.* Genome Res, 2002. **12**(4): p. 656-64.

82.   Paruzynski, A., et al., *Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing.* Nat Protoc, 2010. **5**(8): p. 1379-95.

83.   Arens, A., et al., *Bioinformatic clonality analysis of next-generation sequencing-derived viral vector integration sites.* Hum Gene Ther Methods, 2012. **23**(2): p. 111-8.

84.   Abel, U., et al., *Analyzing the number of common integration sites of viral vectors--new methods and computer programs.* PLoS One, 2011. **6**(10): p. e24247.

85.   Gabriel, R., et al., *An unbiased genome-wide analysis of zinc-finger nuclease specificity.* Nat Biotechnol, 2011. **29**(9): p. 816-23.

86.   Lin, Y. and A.S. Waldman, *Capture of DNA sequences at double-strand breaks in mammalian chromosomes.* Genetics, 2001. **158**(4): p. 1665-74.

87.   Nobile, C., J. Nickol, and R.G. Martin, *Nucleosome phasing on a DNA fragment from the replication origin of simian virus 40 and rephasing upon cruciform formation of the DNA.* Mol Cell Biol, 1986. **6**(8): p. 2916-22.

88.   Cartier, N., et al., *Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy.* Science, 2009. **326**(5954): p. 818-23.

89.   Capranico, G., K.W. Kohn, and Y. Pommier, *Local sequence requirements for DNA cleavage by mammalian topoisomerase II in the presence of doxorubicin.* Nucleic Acids Res, 1990. **18**(22): p. 6611-9.

90.   Gross, D.S. and W.T. Garrard, *Nuclease Hypersensitive Sites in Chromatin.* Annual Review of Biochemistry, 1988. **57**: p. 159-197.

91.   Voigt, P., W.W. Tee, and D. Reinberg, *A double take on bivalent promoters.* Genes Dev, 2013. **27**(12): p. 1318-38.

92.   Tessitore, A., et al., *MicroRNAs in the DNA Damage/Repair Network and Cancer.* Int J Genomics, 2014. **2014**: p. 820248.

93.   Morris, L.G., et al., *Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation.* Nat Genet, 2013. **45**(3): p. 253-61.

94.   Wong, C.C., et al., *Inactivating CUX1 mutations promote tumorigenesis.* Nat Genet, 2014. **46**(1): p. 33-8.

95.   Takita, J., et al., *Aberrations of NEGR1 on 1p31 and MYEOV on 11q13 in neuroblastoma.* Cancer Sci, 2011. **102**(9): p. 1645-50.

96.   Balint, I., et al., *Cloning and characterisation of the RBCC728/TRIM36 zinc-binding protein from the tumor suppressor gene region at chromosome 5q22.3.* Gene, 2004. **332**: p. 45-50.

97.   Saitou, M., et al., *Identification of the TCL6 genes within the breakpoint cluster region on chromosome 14q32 in T-cell leukemia.* Oncogene, 2000. **19**(23): p. 2796-802.

98.   Thomas, G., et al., *A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1).* Nat Genet, 2009. **41**(5): p. 579-84.

99.   Vainio, P., et al., *High-throughput transcriptomic and RNAi analysis identifies AIM1, ERGIC1, TMED3 and TPX2 as potential drug targets in prostate cancer.* PLoS One, 2012. **7**(6): p. e39801.

100.   Reed, J.A., et al., *SKI pathways inducing progression of human melanoma.* Cancer Metastasis Rev, 2005. **24**(2): p. 265-72.

101.   Geyik, E., et al., *Investigation of the association between ATP2B4 and ATP5B genes with colorectal cancer.* Gene, 2014. **540**(2): p. 178-82.

102.   Szczepanski, M.J., et al., *Triggering of Toll-like receptor 4 expressed on human head and neck squamous cell carcinoma promotes tumor development and protects the tumor from immune attack.* Cancer Res, 2009. **69**(7): p. 3105-13.

103.   Dupont, W.D., et al., *Protein phosphatase 2A subunit gene haplotypes and proliferative breast disease modify breast cancer risk.* Cancer, 2010. **116**(1): p. 8-19.

104.   Iorns, E., et al., *Identification of CDK10 as an important determinant of resistance to endocrine therapy for breast cancer.* Cancer Cell, 2008. **13**(2): p. 91-104.

105.   Linja, M.J., et al., *Expression of androgen receptor coregulators in prostate cancer.* Clin Cancer Res, 2004. **10**(3): p. 1032-40.

106.   Bi, X., et al., *Drosophila caliban, a nuclear export mediator, can function as a tumor suppressor in human lung cancer cells.* Oncogene, 2005. **24**(56): p. 8229-39.

107.   Li, Q., et al., *A death receptor-associated anti-apoptotic protein, BRE, inhibits mitochondrial apoptotic pathway.* J Biol Chem, 2004. **279**(50): p. 52106-16.

108.   Hurov, K.E., C. Cotta-Ramusino, and S.J. Elledge, *A genetic screen identifies the Triple T complex required for DNA damage signaling and ATM and ATR stability.* Genes Dev, 2010. **24**(17): p. 1939-50.

109.   Hoefig, K.P., et al., *Eri1 degrades the stem-loop of oligouridylated histone mRNAs to induce replication-dependent decay.* Nat Struct Mol Biol, 2013. **20**(1): p. 73-81.

110.   Meeks-Wagner, D. and L.H. Hartwell, *Normal stoichiometry of histone dimer sets is necessary for high fidelity of mitotic chromosome transmission.* Cell, 1986. **44**(1): p. 43-52.

111.   Adamson, B., et al., *A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response.* Nat Cell Biol, 2012. **14**(3): p. 318-28.

112.   Tang, W., et al., *BIRC6 promotes hepatocellular carcinogenesis: Interaction of BIRC6 with p53 facilitating p53 degradation.* Int J Cancer, 2014.

113.   Neumann, B., et al., *Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes.* Nature, 2010. **464**(7289): p. 721-7.

114.   Singleton, B.K., et al., *The C terminus of Ku80 activates the DNA-dependent protein kinase catalytic subunit.* Mol Cell Biol, 1999. **19**(5): p. 3267-77.

115.   Huynh, K.D., et al., *BCoR, a novel corepressor involved in BCL-6 repression.* Genes Dev, 2000. **14**(14): p. 1810-23.

116.   Banath, J.P., S.H. Macphail, and P.L. Olive, *Radiation sensitivity, H2AX phosphorylation, and kinetics of repair of DNA strand breaks in irradiated cervical cancer cell lines.* Cancer Res, 2004. **64**(19): p. 7144-9.

117.   Neumaier, T., et al., *Evidence for formation of DNA repair centers and dose-response nonlinearity in human cells.* Proc Natl Acad Sci U S A, 2012. **109**(2): p. 443-8.

118.   Bouquet, F., C. Muller, and B. Salles, *The loss of gammaH2AX signal is a marker of DNA double strand breaks repair only at low levels of DNA damage.* Cell Cycle, 2006. **5**(10): p. 1116-22.

119.   Muslimovic, A., et al., *Numerical analysis of etoposide induced DNA breaks.* PLoS One, 2009. **4**(6): p. e5859.

120.   Iacovoni, J.S., et al., *High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome.* EMBO J, 2010. **29**(8): p. 1446-57.

121.   Markova, E., N. Schultz, and I.Y. Belyaev, *Kinetics and dose-response of residual 53BP1/gamma-H2AX foci: co-localization, relationship with DSB repair and clonogenic survival.* Int J Radiat Biol, 2007. **83**(5): p. 319-29.

122.   Wanisch, K. and R.J. Yanez-Munoz, *Integration-deficient lentiviral vectors: a slow coming of age.* Mol Ther, 2009. **17**(8): p. 1316-32.

123.   Wang, M., et al., *PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways.* Nucleic Acids Res, 2006. **34**(21): p. 6170-82.

124.   Neal, J.A., et al., *Inhibition of homologous recombination by DNA-dependent protein kinase requires kinase activity, is titratable, and is modulated by autophosphorylation.* Mol Cell Biol, 2011. **31**(8): p. 1719-33.

125.   Lieber, M.R., *The mechanism of human nonhomologous DNA end joining.* J Biol Chem, 2008. **283**(1): p. 1-5.

126.   Bozas, A., et al., *Genetic analysis of zinc-finger nuclease-induced gene targeting in Drosophila.* Genetics, 2009. **182**(3): p. 641-51.

127.   Brussel, A. and P. Sonigo, *Analysis of early human immunodeficiency virus type 1 DNA synthesis by use of a new sensitive assay for quantifying integrated provirus.* J Virol, 2003. **77**(18): p. 10119-24.

128.   Nightingale, S.J., et al., *Transient gene expression by nonintegrating lentiviral vectors.* Mol Ther, 2006. **13**(6): p. 1121-32.

129.   Gabriel, R., et al., *Comprehensive genomic access to vector integration in clinical gene therapy.* Nat Med, 2009. **15**(12): p. 1431-6.

130.   Cornu, T.I. and T. Cathomen, *Targeted genome modifications using integrase-deficient lentiviral vectors.* Mol Ther, 2007. **15**(12): p. 2107-13.

131.   Seo, J., et al., *Genome-wide profiles of H2AX and gamma-H2AX differentiate endogenous and exogenous DNA damage hotspots in human cells.* Nucleic Acids Res, 2012. **40**(13): p. 5965-74.

132.   Seo, J., et al., *Genome-wide reorganization of histone H2AX toward particular fragile sites on cell activation.* Nucleic Acids Res, 2014. **42**(2): p. 1016-25.

133.   Barlow, J.H., et al., *Identification of early replicating fragile sites that contribute to genome instability.* Cell, 2013. **152**(3): p. 620-32.

134.   Zhou, V.W., A. Goren, and B.E. Bernstein, *Charting histone modifications and the functional organization of mammalian genomes.* Nat Rev Genet, 2011. **12**(1): p. 7-18.

135.   Peng, J.C. and G.H. Karpen, *Heterochromatic genome stability requires regulators of histone H3 K9 methylation.* PLoS Genet, 2009. **5**(3): p. e1000435.

136.   Guelen, L., et al., *Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions.* Nature, 2008. **453**(7197): p. 948-51.

137.   Wen, B., et al., *Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells.* Nat Genet, 2009. **41**(2): p. 246-50.

138.   Li, Z., et al., *Histone H4 Lys 20 monomethylation by histone methylase SET8 mediates Wnt target gene activation.* Proc Natl Acad Sci U S A, 2011. **108**(8): p. 3116-23.

139.   Chapman-Rothe, N., et al., *Chromatin H3K27me3/H3K4me3 histone marks define gene sets in high-grade serous ovarian cancer that distinguish malignant, tumour-sustaining and chemo-resistant ovarian tumour cells.* Oncogene, 2013. **32**(38): p. 4586-92.

140.   Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.

141.   Kraemer, A., et al., *MicroRNA-mediated processes are essential for the cellular radiation response.* Radiat Res, 2011. **176**(5): p. 575-86.

142.   Kraemer, A., et al., *Cell survival following radiation exposure requires miR-525-3p mediated suppression of ARRB1 and TXN1.* PLoS One, 2013. **8**(10): p. e77484.

143.   Wang, J., et al., *Repression of ATR pathway by miR-185 enhances radiation-induced apoptosis and proliferation inhibition.* Cell Death Dis, 2013. **4**: p. e699.

144.   Zhang, J., et al., *Loss of microRNA-143/145 disturbs cellular growth and apoptosis of human epithelial cancers by impairing the MDM2-p53 feedback loop.* Oncogene, 2013. **32**(1): p. 61-9.

145.   Skvortsov, S., et al., *Rac1 as a potential therapeutic target for chemo-radioresistant head and neck squamous cell carcinomas (HNSCC).* Br J Cancer, 2014. **110**(11): p. 2677-87.

146.   Meistrich, M.L., et al., *Low levels of chromosomal mutations in germ cells derived from doxorubicin-treated stem spermatogonia in the mouse.* Cancer Res, 1990. **50**(2): p. 370-4.

147.   Goodarzi, A.A., et al., *ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin.* Mol Cell, 2008. **31**(2): p. 167-77.

148.   Olszak, A.M., et al., *Heterochromatin boundaries are hotspots for de novo kinetochore formation.* Nat Cell Biol, 2011. **13**(7): p. 799-808.

149.   Yuen, K.W., B. Montpetit, and P. Hieter, *The kinetochore and cancer: what's the connection?* Curr Opin Cell Biol, 2005. **17**(6): p. 576-82.

150.   Vader, G., et al., *Protection of repetitive DNA borders from self-induced meiotic instability.* Nature, 2011. **477**(7362): p. 115-9.

151.   Motea, E.A. and A.J. Berdis, *Terminal deoxynucleotidyl transferase: the story of a misguided DNA polymerase.* Biochim Biophys Acta, 2010. **1804**(5): p. 1151-66.

152.   Srinivasan, M., D. Sedmak, and S. Jewell, *Effect of fixatives and tissue processing on the content and integrity of nucleic acids.* Am J Pathol, 2002. **161**(6): p. 1961-71.

153.   Wang, K., et al., *Molecular engineering of DNA: molecular beacons.* Angew Chem Int Ed Engl, 2009. **48**(5): p. 856-70.

154.   Crosetto, N., et al., *Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing.* Nat Methods, 2013. **10**(4): p. 361-5.

155.   Tsai, S.Q., et al., *GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.* Nat Biotechnol, 2015. **33**(2): p. 187-97.

156.   Wang, X., et al., *Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors.* Nat Biotechnol, 2015. **33**(2): p. 175-8.

157.   Frock, R.L., et al., *Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases.* Nat Biotechnol, 2015. **33**(2): p. 179-86.

5.9   Publications and congress attendances

Publications:

Mientus, M., Brady, S., Angelov, A., Zimmermann, P., Wemheuer, B., Schuldes, J., Daniel, R., Liebl, W. (2013) Thermostable Xylanase and β-Glucanase Derived from the Metagenome of the Avachinsky Crater in Kamchatka (Russia). **Current Biotechnology** 2(4): 284-293(10)

Ni, M., Zimmermann, P.K., Kandasamy, K., Lai, W., Li, Y., Leong, M.F., Wan, A.C., Zink, D. (2012) The use of a library of industrial materials to determine the nature of substrate-dependent performance of primary adherent human cells. **Biomaterials** 33(2): 353-364

Zimmermann, P.K., Chiblak, S., Weber, C., Käppel, C., Debus, J., von Kalle, C., Abdollahi, A., Schmidt, M., Nowrouzi A. Genome-Wide Detection of Radiation-Induced DNA Double Strand Breaks and Long-Term Genomic Instability at Single Nucleotide Resolution. (Manuscript submitted)

Nowrouzi, A., Aoi, T., Herbst, F., Deichmann, A., Arens, A., Kara, N., Zimmermann, P.K., Abel, U., Utikal, J., Glimm, H., Schmidt, M., Yamanaka, S., von Kalle, C. Retrovirus Insertion in iPSC Identifies Genes Facilitating Somatic Reprogramming. (Manuscript in preparation)

Conference attendances:

March 1-6, 2015	Keystone Symposia on Molecular and Cellular Biology: Genomic Instability and DNA Repair
	Whistler, British Columbia, Canada
	Contribution: Poster
December 19th, 2014	DKFZ PhD Poster Presentation
	Heidelberg, Germany
	Contribution: Poster
July 9th-11th, 2014	19th Annual PhD Retreat
	Weil der Stadt, Germany
	Contribution: Oral Presentation
May 21-24, 2014	17th Annual Meeting of the American Society of Gene and Cellular Therapy
	Washington, D.C., USA
	Contribution: Poster
October 25-28, 2013	European Society of Gene and Cellular Therapy and Stem Cell Clonality and Genome Stability Retreat Collaborative Congress
	Madrid, Spain
	Contribution: Poster
March 3-8, 2013	Keystone Symposia on Molecular and Cellular Biology: Genomic Instability and DNA Repair
	Banff, Alberta, Canada
	Contribution: Poster

6. DANKSAGUNG

Auf diesen letzten Seiten möchte ich die Möglichkeit nutzen, allen meinen Kollegen, Freunden und meiner Familie für die spannenden Jahre meiner Promotion zu danken.

Als erstes möchte ich mich bei Herrn Prof. Dr. Christof von Kalle dafür danken, dass ich diese spannende Arbeit in seiner Abteilung am NCT durchführen durfte. Außerdem möchte ich mich für die konstruktiven Diskussionen sowie für die zahlreichen Möglichkeiten, die Ergebnisse der Arbeit auf zahlreichen internationalen Konferenzen zu präsentieren, bedanken.

Außerdem gilt mein Dank Herrn Prof. Dr. Stefan Wiemann für seine Bereitschaft als Zweitgutachter dieser Doktorarbeit sowie als Mitglied meines Thesis Advisory Commitees zu agieren. Daneben möchte ich auch Herrn Prof. Dr. Jürgen Debus für seine Teilnahme am Thesis Advisory Commitees danken und Herrn Dr. Dr. Amir Abdollahi für die tolle Kooperation und seine Bereitschaft als Teilnehmer meines Prüfungskommitees aufzutreten. Ein großer Dank geht auch an Herrn Dr. Manfred Schmidt für die Betreuung, seine Ratschläge, die Korrekturen und die Möglichkeit diese Arbeit in seiner Arbeitsgruppe durchzuführen.

Eine besondere Stellung bei dieser Arbeit nimmt Dr. Ali Nowrouzi ein, dem ich für die direkte Betreuung im Labor und die Unterstützung in den vergangenen dreieinhalb Jahren in der Abteilung zutiefst dankbar bin und der mir ein sehr guter Freund geworden ist. Ohne die wertvollen Ratschläge, Diskussionen, die gemeinsamen Unternehmungen in der Freizeit und das leckere Essen wäre diese Arbeit nicht möglich gewesen. Unsere gemeinsame Zeit werde ich in sehr guter Erinnerung behalten und vermissen. Vielen Dank für die schöne Zeit.

Vielen Dank möchte ich auch an meine ehemaligen und derzeitigen Kollegen Malaika Rabenstein, Ina Kutschera, Christine Lulay, Steffie Laier, Kai Lukat, Hannah Pfeff, Lena Fuchs, Martina Burk, Steffie Laufs, Eliana Ruggiero, Anna Paruzynski, Simone Scholz, Tim Rath, Cynthia Bartholomä, Richard Gabriel, Irene Gil-Farina, Stefan Wilkening, Christine Engeland, Friederike Herbst und Annika Mengering für die tolle Atmosphäre, die Unterstützung und die lustigen Stunden im Labor richten. Weiterhin bedanke ich mich bei meinen ehemaligen Masterstudenten Paranchai Boonsawat und Mahdi Akbarpour, die einen großen Beitrag geleistet haben, sowie den Abteilungen von Amir Abdollahi, Stefan Fröhling/ Claudia Scholl und Hanno Glimm. Ein spezielles Dankeschön geht auch an Raffaele Fronza und Saira Afzal, die mich bei der Auswertung der bioinformatischen Daten unterstützen.

Zuletzt geht ein besonderer Dank an meine Freunde und meine Familie, die mich immer in meinem Studium und während meiner Promotion unterstützt haben. Eine herausragende Stellung nehmen meine Eltern und mein Bruder ein. Trotz der Entfernung und der wenigen Treffen im Jahr habt ihr mich während des Studiums und der Doktorarbeit immer unterstützt und mir immer einen wunderbaren Aufenthalt zu Hause in Berlin ermöglicht. Ohne eure Hilfe wäre ich nicht soweit gekommen. Dafür gebührt euch ein besonderer Dank! Außerdem danke ich Elke, Melanie, Peter und Fabi für die tollen und erholsamen Stunden in der Pfalz, das leckere Essen und Trinken sowie die Möglichkeit im Weinberg mitzuhelfen. Eine Reise zu euch ist immer ein schöner Urlaub! Zudem geht ein großer Dank an meine Frau Sarah, mit der ich die gemeinsame Zeit im Studium, in der Freizeit und während Promotion genießen durfte. Du warst und bist mir immer eine große Unterstützung, und ohne dich wäre mir das Studium, die schöne Zeit in Singapur und die Promotion nicht möglich gewesen. Ich freue mich sehr auf unsere weiteren Herausforderungen und eine tolle Zukunft.

Ein großer Dank geht auch an meine Freunde. Vor allem zu nennen ist dabei Christian Weber, für die Unterstützung in der Zellkultur, die zahlreichen lustigen Pausen, die MTB Trips, Superbowl Abende und die viele

Zeit, die wir sonst noch zusammen hatten! Danke möchte ich auch an Friederike Knipping, Christine Käppel, Christian Großardt, Aaron und Jessica Tideman richten, mit denen ich viele schöne gemeinsame Stunden in vielen Kletterhallen, Kinos, Bars und Restaurants in der Rhein-Neckar Region verbringen durfte. Ein großer Dank geht auch an meine Berliner Freunde Sylvia und Simone sowie Gerrit und Daria. Außerdem möchte ich meiner Hockeymannschaft für die vielen Trainingseinheiten und Reisen sowie Erol für die unzähligen Trainingsstunden auf dem Spinningrad, an der Langhantel und dem TRX danken, die mir halfen, die Gedanken wieder neu zu sortieren, neue Energie zu tanken und endlich richtig fit zu werden. ☺

Natürlich darf man bei einer Arbeit über die Verteilung von DNA Schäden nicht die zahlreichen Putzkräfte vergessen, die jeden Morgen meine Konzentration am Computer mit dem Staubsauger störten. Euch widme ich dieses Werk. Ohne Kaffee und Döner wäre diese Arbeit auch nicht durchführbar gewesen, daher danke ich allen Barista in Heidelberg und der Dönerbude am Nahkauf. Außerdem bin ich den Polizistinnen und Polizisten in Heidelberg dankbar, dass sie jeden Morgen meine rote Autoplakette in der Umweltzone übersehen haben und mir so viel Strafgeld erspart blieb. Danke auch an die Förster im Odenwald für die guten Mountainbike Bedingungen auf den zahllosen Trails rund um den Weißen Stein und den Königsstuhl. Ohne eure herausragende Arbeit wäre so manch ein Ride nicht ganz so flowig geworden. Als großer Sportjunky muss ich natürlich auch einen Dank an die Sportartikelhersteller Nike, Asics und Adidas sowie den Hockeymarken TK und Voodoo richten, die mich super unterstützten viel Zeit im Labor zu sparen. Zu tiefstem Dank verpflichtet bin ich auch der „Voice of G100" Jury für die durchweg positive Kritik an meinem Gesangstalent. Sie haben mich dazu inspiriert meine Gesangskarriere auch weiterhin mit größtem Einsatz in anderen Laboratorien dieser Welt zu verfolgen. Ich werde euch auch weiterhin zahlreiche Kostproben meiner Gesangseinlagen kostenlos zukommen lassen. Zuletzt bedanke ich mich bei den Programmdirektoren von RTL, RTL2 und Vox, die mich mit ihrem allabendlichen Programm daran erinnern, dass es definitiv die richtige Entscheidung war zu studieren und promovieren.