Article

Genomic Adaption and Mutational Patterns in a HaCaT Subline Resistant to Alkylating Agents and Ionizing Radiation

Reinhard Ullmann [1],*, Benjamin Valentin Becker [2], Simone Rothmiller [3], Annette Schmidt [3], Horst Thiermann [3], Hanns Leonhard Kaatsch [1], Gerrit Schrock [1], Jessica Müller [1], Julia Jakobi [1], Richard Obermair [1], Matthias Port [1] and Harry Scherthan [1]

[1] Bundeswehr Institute of Radiobiology Affiliated to the University of Ulm, Neuherbergstr. 11, D-80937 Munich, Germany; HannsLeonhardKaatsch@bundeswehr.org (H.L.K.); gerrit2686@hotmail.com (G.S.); Jessica4Mueller@bundeswehr.org (J.M.); JuliaJakobi@bundeswehr.org (J.J.); richard.obermair@web.de (R.O.); MatthiasPort@bundeswehr.org (M.P.); harryscherthan@bundeswehr.org (H.S.)

[2] Bundeswehr Central Hospital, Department of Radiology and Neuroradiology, Rübenacherstrasse 170, D-56072 Koblenz, Germany; benjamin3becker@bundeswehr.org

[3] Bundeswehr Institute of Pharmacology and Toxicology, Neuherbergstr. 11, D-80937 Munich, Germany; Simone1Rothmiller@bundeswehr.org (S.R.); annette.schmidt@unibw.de (A.S.); HorstThiermann@bundeswehr.org (H.T.)

* Correspondence: reinhard1ullmann@bundeswehr.org

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: Sulfur mustard (SM) is a chemical warfare agent that can damage DNA via alkylation and oxidative stress. Because of its genotoxicity, SM is cancerogenic and the progenitor of many chemotherapeutics. Previously, we developed an SM-resistant cell line via chronic exposure of the popular keratinocyte cell line HaCaT to increasing doses of SM over a period of 40 months. In this study, we compared the genomic landscape of the SM-resistant cell line HaCaT/SM to its sensitive parental line HaCaT in order to gain insights into genetic changes associated with continuous alkylation and oxidative stress. We established chromosome numbers by cytogenetics, analyzed DNA copy number changes by means of array Comparative Genomic Hybridization (array CGH), employed the genome-wide chromosome conformation capture technique Hi-C to detect chromosomal translocations, and derived mutational signatures by whole-genome sequencing. We observed that chronic SM exposure eliminated the initially prevailing hypotetraploid cell population in favor of a hyperdiploid one, which contrasts with previous observations that link polyploidization to increased tolerance and adaptability toward genotoxic stress. Furthermore, we observed an accumulation of chromosomal translocations, frequently flanked by DNA copy number changes, which indicates a high rate of DNA double-strand breaks and their misrepair. HaCaT/SM-specific single-nucleotide variants showed enrichment of C > A and T > A transversions and a lower rate of deaminated cytosines in the CpG dinucleotide context. Given the frequent use of HaCaT in toxicology, this study provides a valuable data source with respect to the original genotype of HaCaT and the mutational signatures associated with chronic alkylation and oxidative stress.

Keywords: chronic exposure; DNA alkylation; genome rearrangements; HaCaT; irradiation; mutational signatures; oxidative stress; resistance; sulfur mustard

1. Introduction

Sulfur mustard (SM) is infamous for its use as a chemical warfare agent on the battlefield of Ypres during the First World War in 1917 and in several other more recent military conflicts. Even currently, SM poses a continuing threat, as it is probably the most widely distributed chemical weapon to date [1–3].

A variety of mechanisms have been proposed to explain the tissue-dependent cytotoxic effects of SM [4]. With respect to its genotoxicity, the high bifunctional alkylating activity

of SM can lead to DNA adducts and crosslinks, particularly involving the N7-position of deoxyguanosine and, to a lesser extent, the N3-position of deoxyadenosine [5–11]. In addition, DNA can be damaged by oxidative stress due to SM-induced depletion of endogenous antioxidants [12–15]. One outcome of SM-induced DNA damage can be an increase in PARP activity resulting in NAD+ depletion, decreased glycolysis, and subsequent protease release, leading to reduced cellular fitness [16].

As a consequence of its genotoxicity, SM is a well-established risk factor for bronchial carcinoma and several other tumor types [17–22]. At the same time, SM is the progenitor of many chemically and/or functionally related chemotherapeutics. Already in 1943, nitrogen mustard, a SM derivative, was used to treat a patient with non-Hodgkin lymphoma [23,24]. Since then, nitrogen mustards have been further improved in terms of efficacy and side effects and are regularly used in clinics for the treatment of various types of cancer [25]. The flipside of therapy with this kind of alkylating drug is that tumor cells can acquire resistance [26,27] and patients can develop therapy-associated secondary tumors later on [28]. These complications underscore the need to obtain more detailed knowledge about how alkylating drugs can alter the genomic landscape and in which way these changes might contribute to the development of resistance and secondary tumors.

In previous work, we generated a cell line resistant to SM (HaCaT/SM) [29,30]. The resistant cell population originates from the highly popular keratinocyte cell culture model HaCaT, which is an immortal hypotetraploid cell line that is nontumorigenic and has retained its capacity to differentiate [31,32]. HaCaT displays a rather stable karyotype [33,34] and adaptability to genotoxic stressors [35,36]. The resistant cell line HaCaT/SM emerged after chronic exposure of parental HaCaT cells to increasing doses of SM over a period of 40 months [30]. Further studies demonstrated that resistance was not limited to SM but extended to further nine cytostatics, with several of them currently used in tumor therapy [37]. This makes HaCaT/SM a suitable model to study the spectrum of genetic alterations associated with prolonged exposure and subsequent resistance to alkylating agents and oxidative stress. Using a broad spectrum of methods, ranging from classical cytogenetics to whole-genome sequencing, we performed a comprehensive analysis of chromosomal aberrations and point mutations in HaCaT and its SM-resistant derivative. We present evidence that chronic SM exposure promoted the expansion of a hyperdiploid cell population with an elevated rate of structural chromosomal aberrations and a characteristic mutational signature.

2. Results

2.1. Sulfur Mustard Exposure Promoted Clonal Expansion of a Hyperdiploid Cell Population with Higher Tolerance to Ionizing Radiation

Previous studies showed that SM resistance in HaCaT/SM went along with higher proliferative capability, increased cell survival, and smaller cell nuclei [29]. To scrutinize whether this reduction in nuclear size is due to either higher chromatin compaction or lower DNA content per cell, we prepared chromosome spreads of the parental and the resistant cell line and determined their average chromosome number. It was found that the parental cell line HaCaT was composed of at least two cell populations, one with 76 (in 76% of cells) and one with 56 chromosomes (in 24% of cells) on average, while in HaCaT/SM the distribution of chromosome numbers cells suggests a more homogenous cell population with a median of 55 chromosomes (Figure 1). This indicates that chronic exposure to SM induced positive selection of the hyperdiploid cell population at the cost of the hypotetraploid one that dominated the parental cell line. This change in ploidy was in line with DNA content measurements of the two cell lines (Supplementary Table S1).
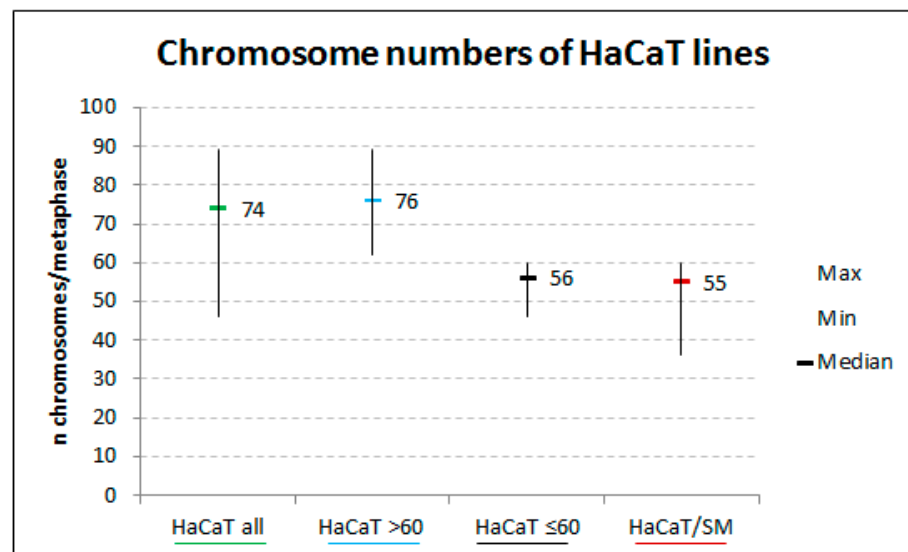
Figure 1. Chromosome numbers in HaCaT/sulfur mustard (SM) and HaCaT. HaCaT parents (all) and subclones (≤60 and >60 chromosomes) are shown. The bars indicate the median. The whiskers indicate the range (min, max). Numbers are based on the analysis of 75 metaphases/cell line.

Previous studies revealed that the mechanisms mediating resistance in HaCaT/SM also protect from other chemical stressors [37]. Through exposure to increasing doses of X-rays, we were also able to demonstrate significant differences in cellular survival in response to ionizing radiation (Figure 2). Exposure to both ionizing radiation (IR) and alkylating agents can result in increased levels of reactive oxygen species (ROS), suggesting a central role of oxidative stress response in the cross-resistance of HaCaT/SM.
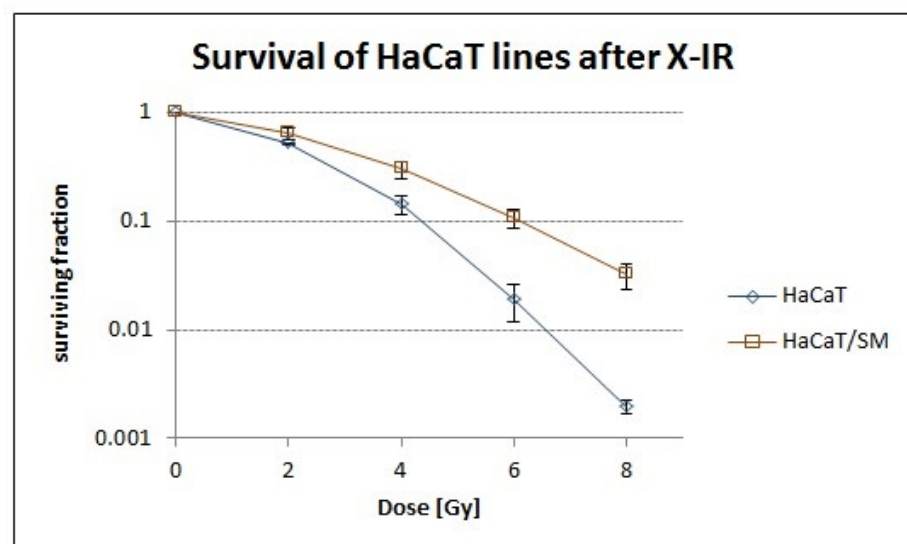

Figure 2. Survival of HaCaT after irradiation. Cellular survival after increasing doses of X-ray ionizing radiation (IR) shows a radioresistant phenotype of the HaCaT/SM cell line, especially at higher doses, as revealed by clonogenic assay (data are based on three replicate experiments ± SD).

2.2. Shared and Unique Structural Chromosomal Aberrations in HaCaT and HaCaT/SM

Next, we performed a genome-wide screen for chromosomal translocations. Translocations drastically change the spatial proximity of chromosomes within the nucleus. This fact can be used to infer chromosomal translocation partners and chromosomal breakpoints on the basis of analysis of chromosome interaction probabilities by means of Hi-C [38–42] (Figure 3; Supplementary Table S2).
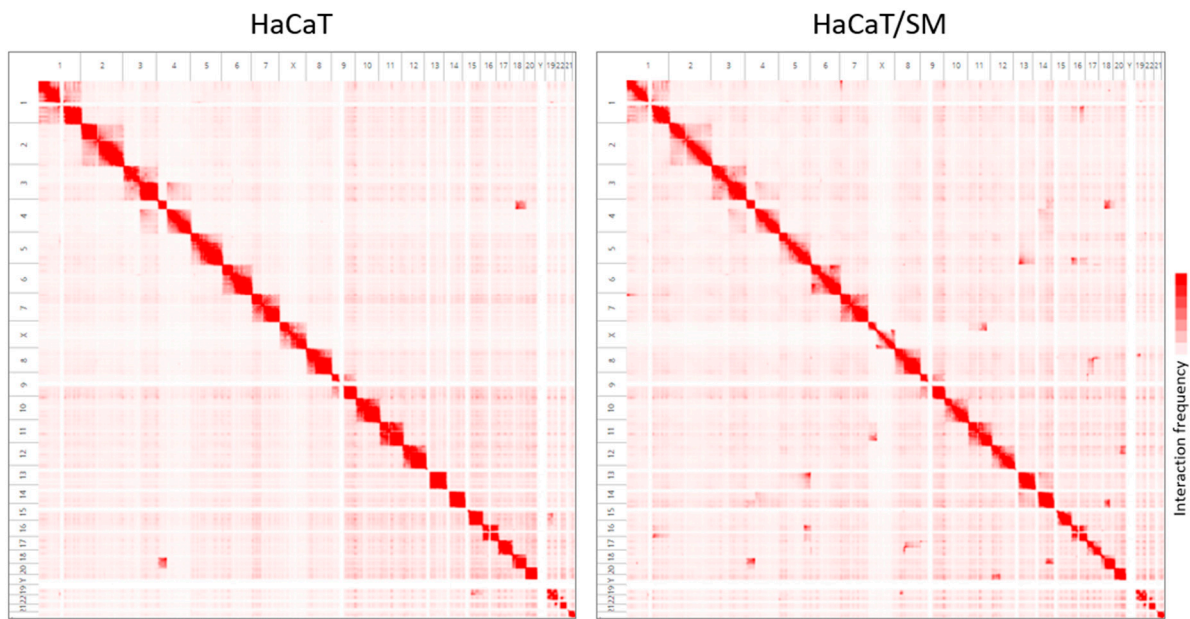
Figure 3. Genome-wide detection of chromosomal translocations by Hi-C interaction matrices for HaCaT (**left**) and HaCaT/SM (**right**). Red color saturation corresponds to the number of Hi-C interactions between chromosomal regions and, hence, their probability of spatial proximity within the nucleus. Chromosome numbers are given at the top and to the left of each matrix. See text for further explanations. Juicebox was employed for visualization of Hi-C data [43]. Enlarged views of t(4;18) and t(19;22) are exemplarily provided in Supplementary Figure S1. Genomic coordinates of translocations are given in Supplementary Table S2.

In addition to the three translocations present in both cell populations, Hi-C interaction matrices revealed seven exclusive translocations in HaCaT/SM and one translocation t(15;19) specific to HaCaT. The latter was most likely originally present in both cell lines before it was eliminated by chromosomal loss in HaCaT/SM (Figure 4). Two of the three translocations shared, t(3;4) and t(4;18), were previously reported by Boukamp and colleagues as structural abnormalities already detectable at passage five of HaCaT [31,33]. In contrast, the seven HaCaT/SM-specific translocations do not match any translocation reported by these authors, whether for early or later passages of HaCaT.

We next analyzed chromosomal gains and losses in HaCaT and HaCaT/SM by array Comparative Genomic Hybridization (array CGH) and quantitative analysis of whole-genome sequencing reads. The copy number alterations identified using both methods largely overlapped, although we also detected copy number alterations exclusively present in one of the two datasets. These differences are most likely due to the fact that the DNA used for array CGH analysis was isolated from a later passage than the DNA used for whole-genome sequencing. All DNA copy segments defined by circular binary segmentation [44] of array CGH-derived $\log_2$ ratios are provided in Supplementary Table S3.

Both HaCaT and HaCaT/SM karyotypes were characterized by complex patterns of chromosomal changes. Comparative analysis revealed extensive karyotypic differences between the two HaCaT cell populations ranging from gross chromosomal rearrangements encompassing several megabases to submicroscopic changes, partly overlapping fragile sites such as Fra1B, Fra2K, Fra3B, and Fra16D. Noteworthy, the gene *WWOX* mapping to Fra16D harbored two deletions. The first one, chr16: 78,371,638–78,384,899 (hg19), was also present in a fraction of parental HaCaT cells, while the second one, chr16: 78,465, 122–78,622,373 (hg19), was exclusively found in HaCaT/SM. We also observed elimination of the surplus chromosome 15 in HaCaT/SM, which was present in the parental cells. A numerical loss of chromosome 15 was previously reported in the context of malignant transformation of HaCaT by HRAS [34,45].
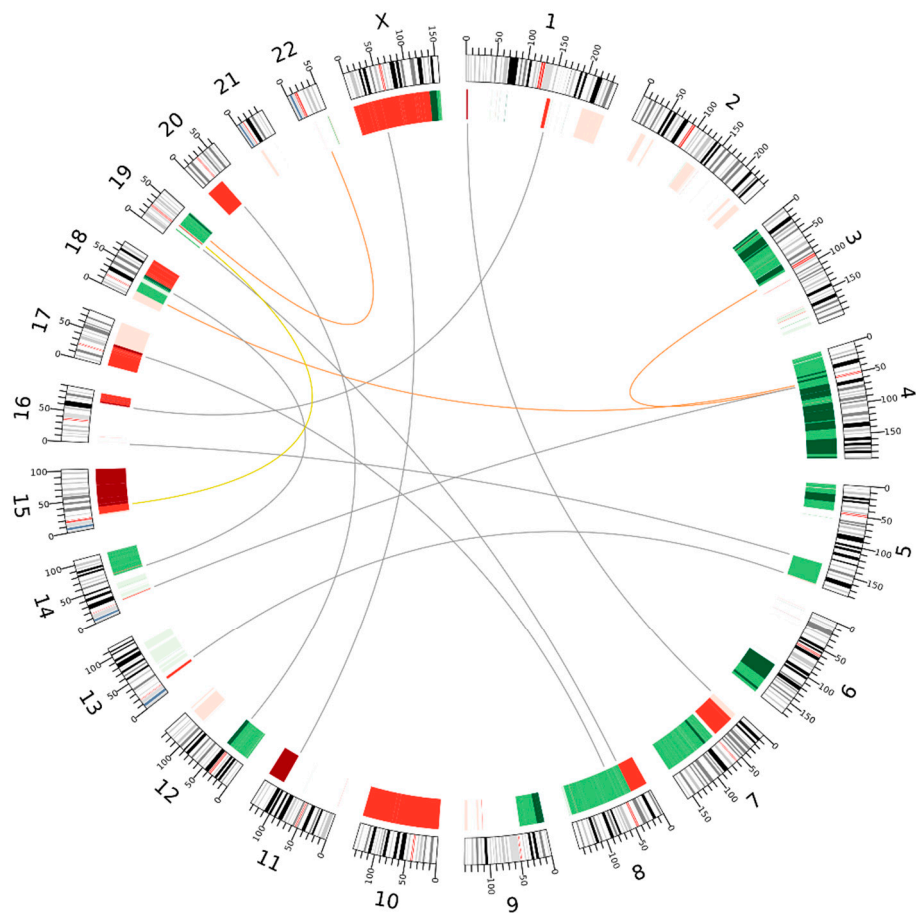
Figure 4. Circos plot of chromosomal translocations and DNA copy number differences in HaCaT and HaCaT/SM. Copy number differences between HaCaT/SM and HaCaT are plotted next to the radially oriented chromosome ideograms. Green indicates higher and red indicates lower DNA copy number in HaCaT/SM relative to HaCaT. The degree of copy number change as defined by their array CGH $\log_2$ ratio is indicated by three different grades of color saturation to enable the distinction of moderate from high copy number changes (dark color: $\log_2$ threshold ± 0.5, medium color: $\log_2$ threshold 0.2, and light color: $\log_2$ threshold 0.12; see Section 4 for details). Orange links within the ideogram show translocations common to both cell lines, whereas yellow and gray links show those specific to HaCaT and HaCaT/SM, respectively. Coordinates of translocation breakpoints are provided in Supplementary Table S2. Visualization was done by means of Circos [46].

A great proportion of HaCaT/SM specific copy number alterations were adjacent to translocation breakpoints (Figure 4). On the example of chromosome X, integrative analysis of Hi-C and array CGH data revealed that seemingly simple and independent structural aberrations were connected and resulted in considerable conformational changes in the affected chromosome (Figure 5). This rearrangement of chromosome X was exclusively present in array CGH and Hi-C data, but absent in whole-genome sequencing (WGS) data, indicating that this change occurred at a later stage of SM exposure.
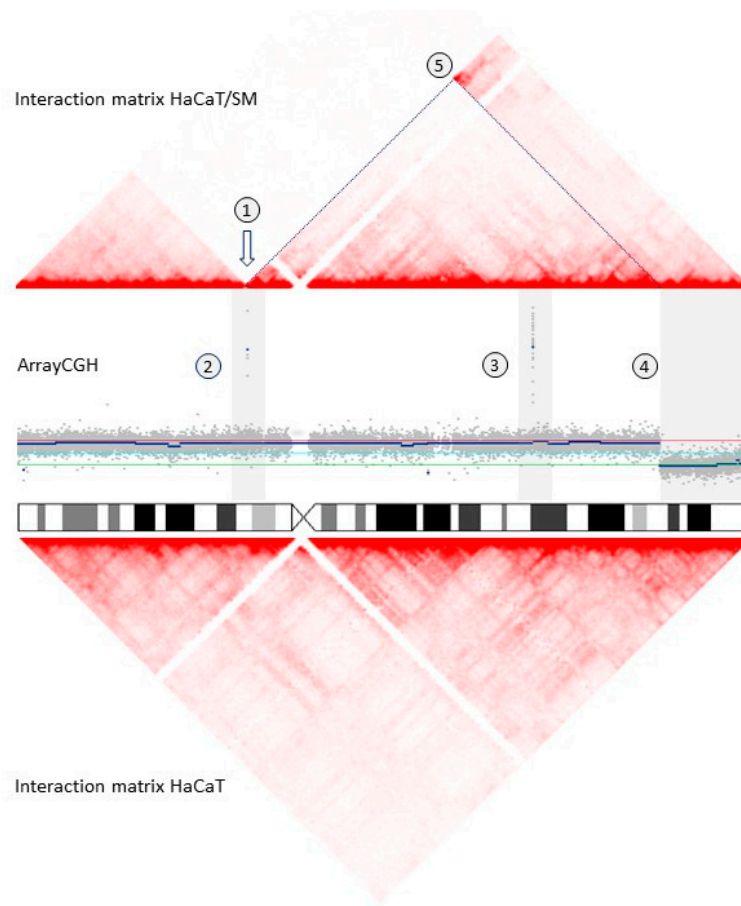
Figure 5. Changes in DNA copy number and chromatin conformation suggest complex rearrangement of chromosome X. Triangular Hi-C interaction matrices for HaCaT/SM and HaCaT are plotted above and below the array CGH log$_2$ ratio plots. Increased red color saturation at the tip of a triangle connecting two genomic positions at the base line indicates higher interaction probability (exemplarily highlighted in the Hi-C results of HaCaT/SM by dashed lines): (1) translocation breakpoint t(11;X); (2) small deletion (chrX: 48,928,825–48,942,873) adjacent to the translocation breakpoint; (3) deletion chrX:109,358,110–109,515,685; (4) duplication chrX:148,572,165–155,257,126; (5) increased chromosomal interaction probability indicating spatial proximity of the chromosomal duplication and translocation breakpoints (highlighted by gray dashed lines). Chromosome coordinates refer to hg19.

2.3. Genomic Characteristics at Sites of DNA Double-Strand Breaks

According to the quantitative analysis of whole-genome sequencing data, we created a manually curated dataset of 75 HaCaT/SM-specific chromosomal breakpoints resolved at a single-base level (Supplementary Table S4). As for non-B-conformations, seven of those breakpoints mapped to inverted repeats and two were located within direct repeats. No G-quadruplex motifs were identified at or near the breakpoints (100 bp up- and downstream interval). We submitted 100 bp sequence intervals surrounding the breakpoints to MEME Suite [47] for sequence motif discovery. Three sequence motifs were significantly overrepresented (Supplementary Figure S2). Yet, none of the motifs had a specific location with respect to the chromosomal breakpoint, and two of the three significantly overrepresented motifs discovered by MEME Suite were identified in sequence intervals that mapped to either SINEs or LINEs. In order to check whether the overlap with repetitive sequences exceeds what can be expected by chance, we employed the R package GenometriCorr [48]. Although 33 breakpoints intersected repetitive elements, neither overlap probability nor relative distance to repetitive elements differed significantly from expec-

tation (Kolmogorov–Smirnov test of normal distribution of relative distances $p = 0.763$; projection test $p = 0.464$).

2.4. Genome-Wide Distribution of Single-Nucleotide Variants

A comparison of single-nucleotide variants (SNV) called from the WGS datasets of HaCaT and HaCaT/SM revealed that 3,974,694 SNVs were common to both cell lines, while 610,408 and 199,342 SNVs were exclusively found in the progenitor and the resistant subline, respectively (read statistics of WGS data comprising the numbers of uniquely mapped reads, mean read length, coverage, and sequencing depth are summarized in Supplementary Table S5). As expected, and exemplarily depicted for chromosome 8 (Supplementary Figure S3), many of these presumptive unique SNVs showed strong correlation with DNA copy number state. Therefore, search for HaCaT/SM-specific mutational patterns was limited to a subset of SNVs as detailed in Section 4 to minimize the misleading influence of cell line-specific chromosomal gains and losses on the definition of HaCaT/SM-specific mutational signatures.

Altogether, this subset of SNVs, mapping to eight chromosomal regions lacking prominent differences in DNA copy number and local bias in mutation frequency between the two cell populations, comprised 488,217 common, 41,944 HaCaT-specific, and 14,787 HaCaT/SM-specific SNVs. As can be inferred from the mutation type frequency plot provided in Figure 6, with reference to the pyrimidine of the DNA double strand, the set of HaCaT/SM-specific SNVs was characterized by a higher rate of C > A and T > A transversions at the cost of T > C and C > T transitions, the latter particularly in the CpG dinucleotide context. An increase in C > A transversions can also be caused by 8-oxoguanine modifications arising in the course of sequencing library preparation [49]. As this artefact should affect both DNA strands with similar frequency, we tested the strand-specificity of all genic SNVs. This analysis revealed a significant bias of C > A transversions toward the transcribed strand (Supplementary Figure S4), making a preparation-derived artefact unlikely.
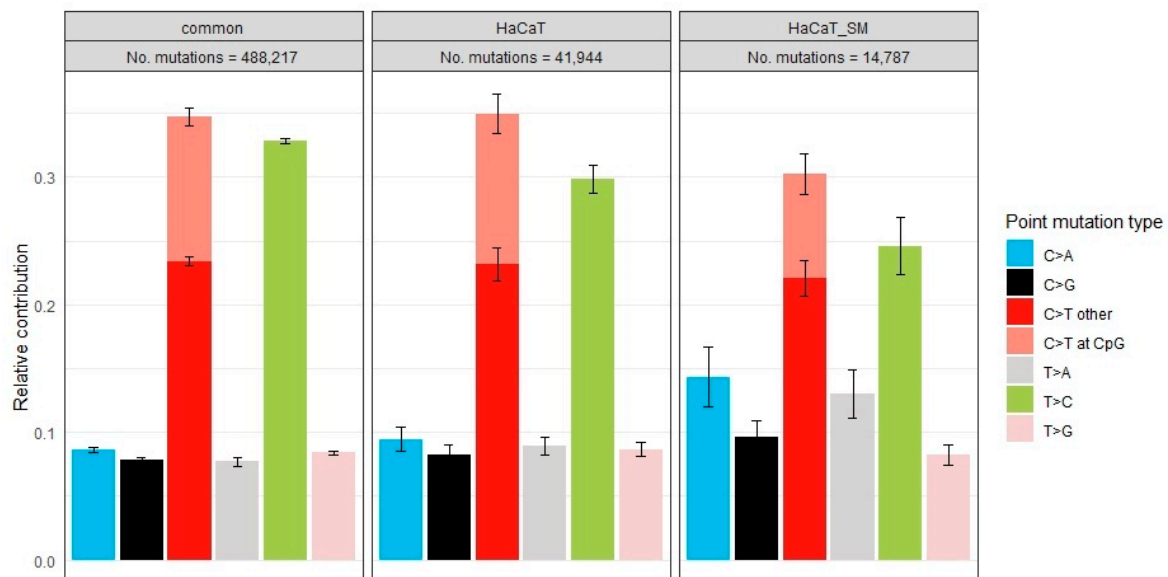

Figure 6. Mutation type frequency plot of common and unique single-nucleotide variants (SNVs). Numbers of mutations are given at the top. The bars represent the relative contribution of each mutation type with reference to the pyrimidine of each basepair. A color legend is provided to the right. The error bars indicate the 95% confidence interval.

2.5. Mutational Signatures

The mutational landscape of a given cell is shaped by the additive contribution of various mutational processes. Each of these processes can exhibit bias with regard to mutation type and sequence context. The term "mutational signatures" has been coined for these distinct mutational footprints. In most tumors, the overall mutational landscape is a composite of several mutational signatures [50,51]. At first, we tested whether HaCaT/SM-specific SNV subsets derived from these eight selected chromosomal regions described above can be distinguished from common and HaCaT-specific SNV subsets. Unsupervised hierarchical clustering of the 24 SNV subsets (common, HaCaT, and HaCaT/SM for eight chromosomal regions) based on the relative contribution of previously published 30 mutational signatures (Mutation Signatures v2–Cosmic) [52] grouped six out of eight HaCaT/SM SNV subsets in one cluster (Figure 7a). In line with the lower prevalence of C > T transitions at CpG dinucleotides in HaCaT/SM, age-associated signature #1 contributed less to HaCaT/SM-specific SNVs when compared to its contribution to common and HaCaT specific SNVs. Although HaCaT is a keratinocyte cell line, we failed to identify a major contribution of ultraviolet (UV) exposure-associated signature # 7 (Figure 7a).

In a next step, we tried to define a mutational signature most representative for the set of HaCaT/SM-specific SNVs. For this purpose, the non-negative matrix factorization (NMF) algorithm [53] implemented in the software package *Mutational Patterns* [54] was employed to extract three mutational signatures from the 24 SNV subsets. The extraction of three mutational signatures should account for the fact that HaCaT/SM-specific SNVs are unlikely induced by SM alone, but include at least a fraction of mutations caused by cell culture conditions and the continuous action of mutational processes already present in HaCaT. However, it is clear that three mutational signatures cannot cover the full complexity of mutational events and, therefore, should be seen as a compromise helping to avoid overstretching the analysis by too many signatures.

Unsupervised hierarchical clustering based on the relative contribution of these three signatures to the overall mutational spectrum of both HaCaT populations clustered the same six out of eight selected chromosomal regions as observed in the clustering experiment described above (Figure 7b). The mutational signature characteristic for HaCaT/SM was dominated by an increased rate of transversions, thereby reflecting the mutation type bias already identified in the HaCaT/SM-specific SNV set (Figure 7c).
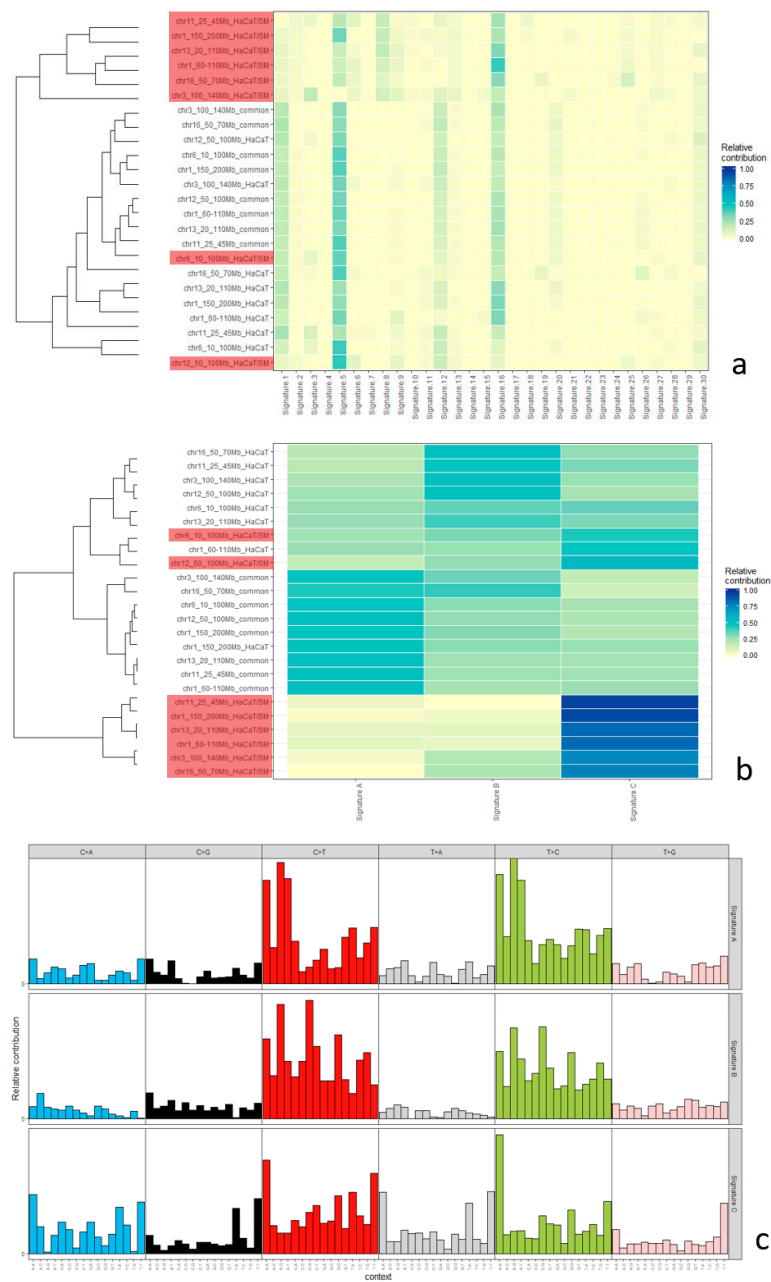
Figure 7. Mutational signatures identified in HaCaT/SM-specific SNV (next page). (**a**) Unsupervised hierarchical clustering based on the relative contribution of 30 previously published mutational signatures. The numbering of signatures refers to Mutation Signatures v2—COSMIC. Names given to the left of each heatmap indicate the chromosomal region of the respective SNV set and whether the SNV set is common or unique to HaCaT and HaCaT/SM, respectively. HaCaT/SM specific SNV sets are additionally highlighted in red. (**b**) Unsupervised hierarchical clustering based on the relative contribution of mutational signatures as extracted by non-negative matrix factorization from the 24 SNV sets representing common and unique SNVs in HaCaT and HaCaT/SM, respectively. HaCaT/SM-specific SNV sets are highlighted in red. (**c**) Relative contribution of each mutation type with reference to the pyrimidine of each basepair for the three signatures shown in (**b**). Each mutation type is shown in the trinucleotide context, i.e., in the middle of a trinucleotide with all possible combinations of neighbors to the left and right. All images were generated by means of the *Mutational Patterns* software package [54].

3. Discussion

In this study we undertook a genome-wide characterization of chromosomal rearrangements and mutational patterns in the SM-resistant HaCaT cell line. We identified genetic changes associated with chronic exposure to SM, which may impact resistance toward various alkylating agents and ionizing irradiation.

3.1. Change in Ploidy and Structural Chromosomal Aberrations

We observed that chronic exposure of the parental HaCaT cell line to SM induced the expansion of a hyperdiploid subclone at the cost of the originally prevailing hypotetraploid cell population. This reversed the initial polyploidization at passage 5 of the HaCaT cell line reported by Boukamp and colleagues, which was deemed essential for autonomous growth of HaCaT [31,55], but apparently no longer for the hyperdiploid HaCaT/SM line. The expansion of a cell clone with lower DNA content was rather unexpected in light of previous reports that linked polyploidization to increased tolerance and adaptability toward genotoxic stressors [56].

In addition to differences in chromosome number, both cell lines also considerably differed with respect to structural chromosomal aberrations. The rather high number of chromosomal translocations exclusive to HaCaT/SM tempted us to speculate that these aberrations might not have evolved gradually and independently, but in a saltatory and interdependent manner, as described for a specific aberrational pattern termed chromoplexy [57]. Yet, we failed to identify any complex chained translocations, the main characteristic of chromoplexy. Notably, the majority of translocation breakpoints were associated with changes in DNA copy number. Unless these unbalanced translocations are due to a nonreciprocal mechanism of interchromosomal DNA double-strand break repair, this observation suggests that a considerable fraction of translocations preceded the numerical changes in chromosomes, which then under selection pressure wiped out one of the derivative chromosomes.

It remains unsure to what extent these genetic changes can be directly attributed to SM exposure or alternatively represent pre-existing mutations that became detectable by SM-induced clonal selection. Either way, the high number of structural chromosomal aberrations points toward a compromised or altered DNA double-strand break repair in HaCaT/SM. Spot-checking at the level of focal chromosomal aberrations unearthed focal copy number alterations affecting several genes implicated in DNA damage response and repair, including *FHIT*, *RAD18*, *RAD51b*, and *WWOX*. Of note, the two genes lost due to the expression of two common fragile sites (Figure 8), *FHIT* (FRA3B) and *WWOX* (FRA16D), have already been discussed in the context of DNA repair pathway choice, altered efficiency of homology-dependent DNA repair, and replication-stress-induced genomic instability, respectively, as well as resistance to irradiation and various chemical stressors [58–60].

3.2. Distribution of Single-Nucleotide Variants and Mutational Signatures

The number of SNVs exclusively identified in HaCaT/SM was ~3 times lower compared to those specific to HaCaT. This rather unexpected low rate of exclusive SNVs in the HaCaT/SM genome might be a consequence of reduced cellular heterogeneity as a result of SM-induced clonal selection and/or the lower cellular DNA content in the hyperdiploid cell population. Mutation type frequencies and comparison to a collection of 30 mutational signatures (Mutation Signatures v2—Cosmic) indicate that the mutational mechanism involving deamination of 5-methylcytosines, a signature typically associated with age, contributed less to the set of HaCaT/SM-specific SNVs when compared to SNVs common to both cell lines or exclusively detected in HaCaT. Surprisingly for keratinocyte-derived cell lines and despite the presence of UV-typical mutations in *TP53* [61], we failed to prove any prominent impact of UV-associated Cosmic Signature #7 both in HaCaT and in HaCaT/SM, which might be due to the fact that HaCaT cells originate from a site not extensively exposed to sunlight [31].
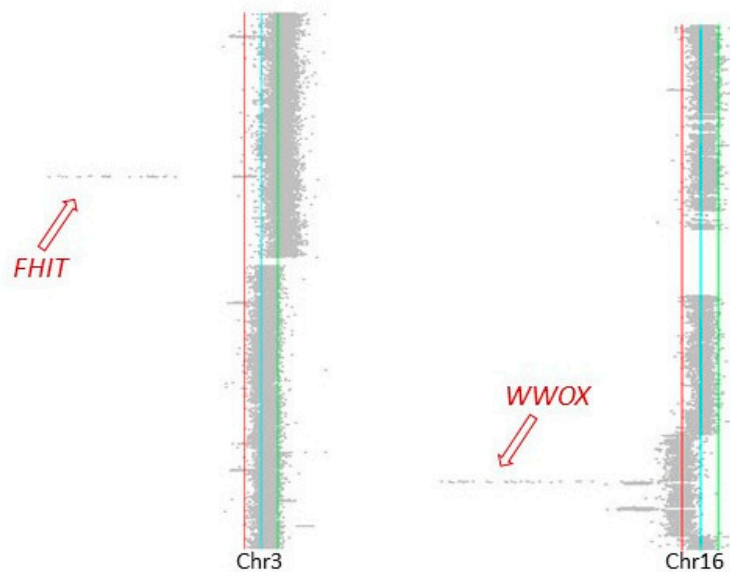
Figure 8. Focal deletions of *FHIT* and *WWOX* due to expression of FRA3B and FRA16D. Array CGH results for HaCaT versus HaCaT/SM for chromosome 3 and 16. Green and red lines indicate $\log_2$ ratios of 0.4 and -0.4, respectively. Arrows point to the deletions encompassing *FHIT* and *WWOX*.

HaCaT/SM-specific SNVs were further characterized by an increased rate of transversions. In particular, C > A transversions have recently been linked to the generation of 8-oxo-guanine in the course of sequencing library preparation [49]. However, a major influence of this technical artefact on the differences in C > A mutation frequency observed here is rather unlikely, given the significant strand bias of C > A changes and the fact that both cell lines were analyzed using the very same protocols and bioinformatics pipeline. In vivo, C > A transversions are frequently caused by oxidative stress, either by mispairing or by compromised base excision repair of 8-oxo-guanine or formamidopyrimidine-guanine (Fapy-G), an alternative oxidation product of guanine [62,63]. In accordance, Kucab and colleagues observed a dominance of C > A transversions in various trinucleotide contexts after triggering oxidative stress in stem cells by means of potassium bromate. Interestingly, the same authors hardly observed C > A transversions when exposing the same stem cells to various alkylating agents [64]. This may suggest that not alkylation but oxidative stress is the main driver of point mutations associated with SM exposure. Noteworthy in this context is our observation that HaCaT/SM is less sensitive to irradiation, which supports the idea that resistance of HaCaT/SM is mainly based on a better way to handle oxidative damage that is also a feature of ionizing radiation exposure. A considerable contribution of oxidative stress to the mutational burden would also be of relevance in the context of chemotherapy with alkylating agents, as these treatments have already been reported to induce oxidative stress [65].

Yet, there are several uncertainties attached to speculations on the main source of point mutations in HaCaT/SM. Bulky DNA adducts can also lead to depurinization and consequently transversions (reviewed in [66]), and in vitro experiments have demonstrated that N7 alkylation of guanine can also result in Fapy-G lesions [62,67–69]. Moreover, the mutational pattern generated by chronic exposure of HaCaT/SM, with extensive chance of adaption and selection, might differ from that induced by acute intoxication of nonadapted cells. It is also likely that a fraction of the observed mutations emerged independently of SM exposure with several of them already being present in some cells before SM administration. In any case, the high frequency of both C > A and T > A transversion is suggestive of an increased occurrence of apurinic sites, as a consequence of either oxidative damage repair or alkylation-induced DNA repair, which are then preferentially complemented with adenine following the so-called A-rule [70–73].

In summary, we described the complex genomic changes associated with chronic exposure of the keratinocyte cell line HaCaT to sulfur mustard. In addition to a better understanding of the genetic consequences of prolonged treatment of cells with an alkylating and oxidative stress provoking agents, our findings remind of the fact that toxicological long-term experiments can drastically alter the genetic characteristics of this highly popular cell culture test system and most likely many others.

4. Materials and Methods

4.1. Cultivation and Authentication of HaCaT

HaCaT cells were kindly provided by Prof. Dr. N. Fusenig (German Cancer Research Center, Heidelberg, Germany) and cultivated in Dulbecco's modified Eagle medium supplemented with glutamine and 10% fetal bovine serum (Gibco, Carlsbad, CA, USA). Cells were maintained at 37 °C and 5% CO_2 in a humidified atmosphere. For cell passaging, HaCaT cells were detached with 0.05% trypsin in 1 mM EDTA (Gibco, Carlsbad, CA, USA) for 5 min and reseeded in fresh culture medium twice a week. The generation of HaCaT/SM is described in Schmidt et al. [30].

Authentication of HaCaT and its SM-resistant subline was based on the reidentification of cytogenetic aberrations as reported in the original description of HaCaT [31] and HaCaT-specific *TP53* mutations identified by Lehmann and colleagues [61]. The common origin of HaCaT and HaCaT/SM was further confirmed by evaluating the extent of shared single-nucleotide variants (SNVs) and DNA copy number changes. PCR testing revealed that both cell lines were infected with mycoplasma during an advanced stage of the experiments, which is, however, not considered to have a major influence on the results since the contamination affected both cell lines used for the comparisons.

4.2. Chromosome Preparation and Quantification of DNA Content

The HaCaT parental cell line and the SM-resistant cell line HaCaT/SM [30] in exponential growth were arrested at metaphase with 0.05 mg/mL Colcemid (Gibco, Carlsbad, CA, USA) for 2 h and then harvested by treatment with trypsin. Chromosome spreads were obtained according to standard acetic acid/methanol (1/3) fixation protocols [74]. Slides were stored at 22 °C until use. DAPI-stained metaphases were analyzed for chromosome number using a Metafer4 imaging system (MetaSystems).

For quantification of DNA content, 1.5×10^6 cells of HaCaT and HaCaT/SM were used, respectively. DNA isolation was performed with a Gentra Puregene Cell Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Briefly, cell pellets were lysed in 300 μL of cell lysis solution for 10 min at room temperature (RT). After addition of 100 μL protein precipitation solution, incubation on ice for 5 min, and centrifugation (3 min, $16,000 \times g$), the supernatant was transferred into a new tube, and 300 μL of pure isopropanol was added. After another centrifugation, the pelleted DNA was washed with 300 μL of 70% ethanol by centrifugation. The DNA pellet was air-dried and rehydrated in 100 μL of DNA hydration solution for 1 h at 65 °C. DNA concentration was determined by NanoDrop 8000 (Thermo Fisher Scientific, Waltham, MA, USA) in biological duplicates from three independent experiments. HaCaT and HaCaT/SM were compared by the unpaired two-sample Wilcoxon test and p-values < 0.05 were considered statistically significant.

4.3. Whole-Genome Sequencing and Variant Calling

Whole-genome sequencing and primary data analysis comprising quality trimming, mapping, and variant calling were done by a commercial core facility (Eurofins, Ebersberg, Germany). In brief, 2×125 bp reads were generated for HaCaT and HaCaT/SM on a HiSeq2500 (Illumina, San Diego, CA, USA). Adapter sequences and low-quality bases (Phred < 20) were clipped off by means of Trimmomatic [75]. Reads < 40 bp after clipping were excluded from further analysis. Remaining reads were mapped to Hg38 employing BWA-MEM using the default settings [76]. Duplicate reads were removed and variants were called with VARScan 2.3 [77]. Common variants were distinguished from those exclusively

present in HaCaT or HaCaT SM by means of the *vcf-isec* command implemented in VCF Tools [78].

4.4. Analysis of DNA Copy Number Changes

DNA copy number changes in HaCaT and HaCaT/SM were analyzed by means of array CGH and quantitative analysis of whole-genome sequencing data. For comparability of array CGH data, which referred to hg19, WGS data of HaCaT and HaCaT/SM were remapped to hg19 by means of BWA-MEM [76] and quality-filtered (>q30) using Samtools [79]. ACE [80] and its included functions of QDNASeq [81] were used for estimation of DNA copy number according to genomic binning of GC bias-corrected read counts with a bin size of 1000 bp. DNA used for array CGH was isolated from a later cell passage with longer exposure to SM than the DNA used for whole-genome sequencing. In the first two array CGH experiments, DNA isolated from the two HaCaT populations was separately hybridized against reference DNA derived from GM12878, which is a lymphoblastoid cell line that lacks gross structural chromosomal aberrations. GM12878 was obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research (Repository number GM12878). In a third experiment, HaCaT and HaCaT/SM were hybridized against each other in order to directly verify DNA copy number differences between these two HaCaT cell populations. Labeling of DNA and hybridization onto a 400 k SurePrint G3 Human CGH Array (Agilent; 4448 A) were performed according to the manufacturer's recommendations. Feature Extraction Software 12.0.3.1 was employed for primary data analysis. Circular Binary Segmentation [44] was performed with DNACopy [82] implemented in GenomeCAT [83]. In order to distinguish moderate from high DNA copy number alterations in the Circos plot presented in Figure 4, we categorized chromosomal changes according to their $\log_2$ ratio shift (thresholds for each fragment defined by Circular Binary Segmentation: ±0.12, ±0.2, and ±0.5). IGV [84] and GenomeCAT [83] were used to visualize the results. All array CGH data refer to genome release hg19. Coordinates of fragile sites were taken from the HumCFS database [85].

4.5. Genomic Characteristics at Sites of DNA Double-Strand Breaks

For identification of chromosomal breakpoints at single-base resolution, binning results of WGS data (see above) were plotted in the IGV browser and screened by visual inspection for reads overlapping possible breakpoints. Only breakpoints confirmed by three independent breakpoint-spanning reads exclusively present in HaCaT/SM were considered for further analysis, which was mainly based on R packages downloaded from the Bioconductor depository (https://bioconductor.org/). BED files were transformed into GenomicRange objects using the import function of Rtracklayer [86]. DNAShapeR [87] was employed to automatically retrieve the sequences surrounding the breakpoints as FASTA files. These files were then uploaded to the webtool MEME suite for identification of significantly overrepresented sequence motifs (http://meme-suite.org) [47]. Statistical significance of breakpoint overlap with repetitive elements was tested with GenometriCorr [48]. The genomic coordinates of non-B DNA conformations were obtained from non-B DB [88].

4.6. Analysis of Chromosomal Translocations by In Situ Hi-C

Genome-wide probability of spatial proximity of chromosomal segments was determined by means of in situ Hi-C following the protocol developed by Rao and colleagues [89] using one million cells of HaCaT and HaCaT/SM. Hi-C libraries were paired-end sequenced (2 × 75 bp) on a NextSeq500 (Illumina, San Diego, CA, USA). The software package Juicer was employed for alignment and data analysis using the default setting [90]. Translocation breakpoints were defined by visual inspection of the Hi-C interaction matrices as described previously [42] employing the webtool Juicebox [43] and Circos [46].

4.7. Genomic Distribution of Variants and Mutational Signatures

For visualizing the distribution of SNVs across the genome, the Bedtools *map* command was used to sum the number of SNVs per 100 kb genomic bin proceeding from three bedGraph files listing the chromosomal positions of variants both shared and unique to HaCaT and HaCaT/SM (Bedtools v2.17.0; [91]). Results were plotted as heatmaps in the Integrated Genomics Viewer (IGV 2.3.88; [84]). Many of those SNVs identified as unique in either HaCaT or HaCaT/SM are not a consequence of SM treatment, but simply "unique" due to alterations in DNA copy number state, e.g., loss of one chromosome in only one of the two HaCaT cell populations. Therefore, we limited our search for possible SM-related mutational signatures to SNVs mapping to eight chromosomal regions that neither showed any chromosomal differences between HaCaT and HaCaT/SM nor any apparent copy number-associated regional enrichment of SNVs as determined by visual inspection. SNVs mapping to these eight regions, which encompassed 405 Mb of the genome, were extracted from the genome-wide vcf files using the filtering options provided by VCFtools [78]. The resulting 24 vcf files separately representing unique and common SNVs in each of the eight regions were imported into *Mutational Patterns* [54] for further analysis. We followed the workflow described in the vignette of this R package to determine the relative contribution of the six possible base substitution types, to generate 96 trinucleotide count matrices, to extract mutational signatures utilizing non-negative matrix factorization (factorization rank $n = 3$), and to compare our SNV sets to previously published COSMIC mutational signatures (https://cancer.sanger.ac.uk/cosmic/signatures_v2) [50,92,93].

References

1. Rall, D.P.; Pechura, C.M. Effects on health of mustard gas. *Nature* **1993**, *366*, 398–399. [CrossRef] [PubMed]
2. Szinicz, L. History of chemical and biological warfare agents. *Toxicology* **2005**, *214*, 167–181. [CrossRef] [PubMed]
3. World Health Organization. Chemical agents and related occupations. *IARC Monogr. Eval. Carcinog. Risks Hum.* **2012**, *100*, 9–562.
4. Somani, S.M.; Babu, S.R. Toxicodynamics of sulfur mustard. *Int. J. Clin. Pharmacol. Ther. Toxicol.* **1989**, *27*, 419–435.
5. Dębiak, M.; Panas, A.; Steinritz, D.; Kehe, K.; Bürkle, A. High-throughput analysis of DNA interstrand crosslinks in human peripheral blood mononuclear cells by automated reverse FADU assay. *Toxicology* **2011**, *280*, 53–60. [CrossRef]

6. Roberts, J.J.; Brent, T.P.; Crathorn, A.R. Evidence for the inactivation and repair of the mammalian DNA template after alkylation by mustard gas and half mustard gas. *Eur. J. Cancer* **1971**, *7*, 515–524. [CrossRef]
7. Shahin, S.; Cullinane, C.; Gray, P.J. Mitochondrial and nuclear DNA damage induced by sulphur mustard in keratinocytes. *Chem. Biol. Interact.* **2001**, *138*, 231–245. [CrossRef]
8. Walker, I.G. Intrastrand bifunctional alkylation of DNA in mammalian cells treated with mustard gas. *Can. J. Biochem.* **1971**, *49*, 332–336. [CrossRef]
9. Ludlum, D.B.; Ritchie, P.A.; Hagopian, M.; Niu, T.Q.; Yu, D. Detection of sulfur mustard-induced DNA modifications. *Chem. Biol. Interact.* **1994**, *91*, 39–49. [CrossRef]
10. Fidder, A.; Moes, G.W.; Scheffer, A.G.; van der Schans, G.P.; Baan, R.A.; de Jong, L.P.; Benschop, H.P. Synthesis, characterization, and quantitation of the major adducts formed between sulfur mustard and DNA of calf thymus and human blood. *Chem. Res. Toxicol.* **1994**, *7*, 199–204. [CrossRef]
11. Mood, M.B.; Hefazi, M. The pharmacology, toxicology, and medical treatment of sulphur mustard poisoning. *Fundam. Clin. Pharmacol.* **2005**, *19*, 297–315. [CrossRef] [PubMed]
12. Pohanka, M.; Stetina, R.; Svobodova, H.; Nedecky, B.R.; Jilkova, M.; Sochor, J.; Sobotka, J.; Adam, V.; Kizek, R. Sulfur mustard causes oxidative stress and depletion of antioxidants in muscles, livers, and kidneys of Wistar rats. *Drug Chem. Toxicol.* **2013**, *36*, 270–276. [CrossRef] [PubMed]
13. Layali, I.; Shahriary, A.; Talatappe, N.R.; Tahmasbpour, E.; Rostami, H.; Harchegani, A.B. Sulfur mustard triggers oxidative stress through glutathione depletion and altered expression of glutathione-related enzymes in human airways. *Immunopharmacol. Immunotoxicol.* **2018**, *40*, 290–296. [CrossRef] [PubMed]
14. Harchegani, A.B.; Tahmasbpour, E.; Borna, H.; Imamy, A.; Ghanei, M.; Shahriary, A. Free Radical Production and Oxidative Stress in Lung Tissue of Patients Exposed to Sulfur Mustard: An Overview of Cellular and Molecular Mechanisms. *Chem. Res. Toxicol.* **2018**, *31*, 211–222. [CrossRef] [PubMed]
15. Husain, K.; Dube, S.N.; Sugendran, K.; Singh, R.; das Gupta, S.; Somani, S.M. Effect of topically applied sulphur mustard on antioxidant enzymes in blood cells and body tissues of rats. *J. Appl. Toxicol.* **1996**, *16*, 245–248. [CrossRef]
16. Papirmeister, B.; Gross, C.L.; Meier, H.L.; Petrali, J.P.; Johnson, J.B. Molecular basis for mustard-induced vesication. *Fundam. Appl. Toxicol.* **1985**, *5*, S134–S149. [CrossRef]
17. Weiss, A.; Weiss, B. Carcinogenesis due to mustard gas exposure in man, important sign for therapy with alkylating agents. *Dtsch. Med. Wochenschr.* **1975**, *100*, 919–923. [PubMed]
18. Yamada, A.; Hirose, F.; Miyanishi, M. An autopsy case of bronchial carcinoma found in a patient succumbed to occupational mustard gas poisoning. *Gan* **1953**, *44*, 216–218.
19. Tokuoka, S. Early cancer and related changes in the bronchial epithelium of former mustard gas workers. *Gan Kagaku Ryoho* **1985**, *12*, 708–713.
20. Tokuoka, S.; Hayashi, Y.; Inai, K.; Egawa, H.; Aoki, Y.; Akamizu, H.; Eto, R.; Nishida, T.; Ohe, K.; Kobuke, T. Early cancer and related lesions in the bronchial epithelium in former workers of mustard gas factory. *Acta Pathol. Jpn.* **1986**, *36*, 533–542. [CrossRef]
21. Dacre, J.C.; Goldman, M. Toxicology and pharmacology of the chemical warfare agent sulfur mustard. *Pharmacol. Rev.* **1996**, *48*, 289–326. [PubMed]
22. World Health Organization. Overall evaluations of carcinogenicity: An updating of IARC Monographs volumes 1 to 42. *IARC Monogr. Eval. Carcinog. Risks Hum. Suppl.* **1987**, *7*, 1–440.
23. de Vita, V.T., Jr.; Chu, E. A history of cancer chemotherapy. *Cancer Res.* **2008**, *68*, 8643–8653. [CrossRef] [PubMed]
24. Goodman, L.S.; Wintrobe, M.M.; Dameshek, W.; Goodman, M.J.; Gilman, A.; McLennan, M.T. Nitrogen mustard therapy; use of methyl-bis (beta-chloroethyl) amine hydrochloride and tris (beta-chloroethyl) amine hydrochloride for Hodgkin's disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders. *J. Am. Med. Assoc.* **1946**, *132*, 126–132. [CrossRef] [PubMed]
25. Varela, B.D.; Ai, Y.; Liang, D.; Xue, F. Nitrogen Mustards as Anticancer Chemotherapies: Historic Perspective, Current Developments and Future Trends. *Curr. Top. Med. Chem.* **2019**, *19*, 691–712. [CrossRef]
26. Panasci, L.; Xu, Z.-Y.; Bello, V.; Aloyz, R. The role of DNA repair in nitrogen mustard drug resistance. *Anticancer Drugs* **2002**, *13*, 211–220. [CrossRef]
27. Buller, A.L.; Clapper, M.L.; Tew, K.D. Glutathione S-transferases in nitrogen mustard-resistant and -sensitive cell lines. *Mol. Pharmacol.* **1987**, *31*, 575–578.
28. Boffetta, P.; Kaldor, J.M. Secondary malignancies following cancer chemotherapy. *Acta Oncol.* **1994**, *33*, 591–598. [CrossRef]
29. Wolf, M.; Siegert, M.; Rothmiller, S.; Scheithauer, N.; Strobelt, R.; Steinritz, D.; Worek, F.; Thiermann, H.; Schmidt, A. Characterization of sulfur mustard resistant keratinocyte cell line HaCaT/SM. *Toxicol. Lett.* **2016**, *244*, 49–55. [CrossRef]
30. Schmidt, A.; Steinritz, D.; Thiermann, H. Development of the sulfur mustard resistant keratinocyte cell line HaCaT/SM. *Toxicol. Lett.* **2016**, *244*, 44–48. [CrossRef]
31. Boukamp, P.; Petrussevska, R.T.; Breitkreutz, D.; Hornung, J.; Markham, A.; Fusenig, N.E. Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *J. Cell Biol.* **1988**, *106*, 761–771. [CrossRef] [PubMed]
32. Schoop, V.M.; Mirancea, N.; Fusenig, N.E. Epidermal organization and differentiation of HaCaT keratinocytes in organotypic coculture with human dermal fibroblasts. *J. Investig. Dermatol.* **1999**, *112*, 343–353. [CrossRef] [PubMed]

33. Boukamp, P.; Popp, S.; Altmeyer, S.; Hülsen, A.; Fasching, C.; Cremer, T.; Fusenig, N.E. Sustained nontumorigenic pheno-type correlates with a largely stable chromosome content during long-term culture of the human keratinocyte line HaCaT. *Genes Chromosomes Cancer* **1997**, *19*, 201–214. [CrossRef]

34. Fusenig, N.E.; Boukamp, P. Multiple stages and genetic alterations in immortalization, malignant transformation, and tumor progression of human skin keratinocytes. *Mol. Carcinog.* **1998**, *23*, 144–158. [CrossRef]

35. He, Y.Y.; Pi, J.; Huang, J.L.; Diwan, B.A.; Waalkes, M.P.; Chignell, C.F. Chronic UVA irradiation of human HaCaT keratinocytes induces malignant transformation associated with acquired apoptotic resistance. *Oncogene* **2006**, *25*, 3680–3688. [CrossRef]

36. Pi, J.; He, Y.; Bortner, C.; Huang, J.; Liu, J.; Zhou, T.; Qu, W.; North, S.L.; Kasprzak, K.S.; Diwan, B.A.; et al. Low level, long-term inorganic arsenite exposure causes generalized resistance to apoptosis in cultured human keratinocytes: Potential role in skin co-carcinogenesis. *Int. J. Cancer* **2005**, *116*, 20–26. [CrossRef]

37. Schmidt, A.; Wolf, M.; Rothmiller, S.; Worek, F.; Steinritz, D.; Thiermann, H. Cytostatic resistance profile of the sulfur mustard resistant keratinocyte cell line HaCaT/SM. *Toxicol. Lett.* **2018**, *293*, 16–20. [CrossRef]

38. Harewood, L.; Kishore, K.; Eldridge, M.D.; Wingett, S.; Pearson, D.; Schoenfelder, S.; Collins, V.P.; Fraser, P. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. *Genome Biol.* **2017**, *18*, 125. [CrossRef]

39. Burton, J.N.; Adey, A.; Patwardhan, R.P.; Qiu, R.; Kitzman, J.O.; Shendure, J. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **2013**, *31*, 1119–1125. [CrossRef]

40. Engreitz, J.M.; Agarwala, V.; Mirny, L.A. Three-dimensional genome architecture influences partner selection for chromosomal translocations in human disease. *PLoS ONE* **2012**, *7*, e44196. [CrossRef]

41. Barutcu, A.R.; Lajoie, B.R.; McCord, R.P.; Tye, C.E.; Hong, D.; Messier, T.L.; Browne, G.; van Wijnen, A.J.; Lian, J.B.; Stein, J.L.; et al. Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome Biol.* **2015**, *16*, 214. [CrossRef] [PubMed]

42. Steininger, A.; Ebert, G.; Becker, B.V.; Assaf, C.; Möbs, M.; Schmidt, C.A.; Grabarczyk, P.; Jensen, L.R.; Przybylski, G.K.; Port, M.; et al. Genome-Wide Analysis of Interchromosomal Interaction Probabilities Reveals Chained Translocations and Overrepresentation of Translocation Breakpoints in Genes in a Cutaneous T-Cell Lymphoma Cell Line. *Front. Oncol.* **2018**, *8*, 183. [CrossRef] [PubMed]

43. Robinson, J.T.; Turner, D.; Durand, N.C.; Thorvaldsdottir, H.; Mesirov, J.P.; Aiden, E.L. Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *Cell Syst.* **2018**, *6*, 256–258. [CrossRef]

44. Venkatraman, E.S.; Olshen, A.B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **2007**, *23*, 657–663. [CrossRef] [PubMed]

45. Boukamp, P.; Peter, W.; Pascheberg, U.; Altmeier, S.; Fasching, C.; Stanbridge, E.J.; Fusenig, N.E. Step-wise progression in human skin carcinogenesis in vitro involves mutational inactivation of p53, rasH oncogene activation and additional chromosome loss. *Oncogene* **1995**, *11*, 961–969.

46. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [CrossRef]

47. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [CrossRef]

48. Favorov, A.; Mularoni, L.; Cope, L.M.; Medvedeva, Y.; Mironov, A.A.; Makeev, V.J.; Wheelan, S.J. Exploring massive, genome scale datasets with the GenometriCorr package. *PLoS Comput. Biol.* **2012**, *8*, e1002529. [CrossRef]

49. Costello, M.; Pugh, T.J.; Fennell, T.J.; Stewart, C.; Lichtenstein, L.; Meldrim, J.C.; Fostel, J.L.; Friedrich, D.C.; Perrin, D.; Dionne, D.; et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* **2013**, *41*, e67. [CrossRef]

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

51. Alexandrov, L.B.; Zainal, S.N.; Wedge, D.C.; Campbell, P.J.; Stratton, M.R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **2013**, *3*, 246–259. [CrossRef] [PubMed]

52. Tate, J.G.; Bamford, S.; Jubb, H.C.; Sondka, Z.; Beare, D.M.; Bindal, N.; Boutselakis, H.; Cole, C.G.; Creatore, C.; Dawson, E.; et al. COSMIC: The Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **2019**, *47*, D947. [CrossRef] [PubMed]

53. Lee, D.D.; Seung, H.S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, *401*, 788–791. [CrossRef] [PubMed]

54. Blokzijl, F.; Janssen, R.; van Boxtel, R.; Cuppen, E. MutationalPatterns: Comprehensive genome-wide analysis of mutational processes. *Genome Med.* **2018**, *10*, 33. [CrossRef]

55. Lim, S.; Ganem, N.J. Tetraploidy and tumor development. *Oncotarget* **2014**, *5*, 10959–10960. [CrossRef]

56. Schoenfelder, K.P.; Fox, D.T. The expanding implications of polyploidy. *J. Cell Biol.* **2015**, *209*, 485–491. [CrossRef]

57. Baca, S.C.; Prandi, D.; Lawrence, M.S.; Mosquera, J.M.; Romanel, A.; Drier, Y.; Park, K.; Kitabayashi, N.; MacDonald, T.Y.; Ghandi, M.; et al. Punctuated evolution of prostate cancer genomes. *Cell* **2013**, *153*, 666–677. [CrossRef]

58. Hosseini, S.A.; Horton, S.; Saldivar, J.C.; Miuma, S.; Stampfer, M.R.; Heerema, N.A.; Huebner, K. Common chromosome fragile sites in human and murine epithelial cells and FHIT/FRA3B loss-induced global genome instability. *Genes Chromosomes Cancer* **2013**, *52*, 1017–1029. [CrossRef]

59. Schrock, M.S.; Batar, B.; Lee, J.; Druck, T.; Ferguson, B.; Cho, J.H.; Akakpo, K.; Hagrass, H.; Heerema, N.A.; Xia, F.; et al. Wwox-Brca1 interaction: Role in DNA repair pathway choice. *Oncogene* **2017**, *36*, 2215–2227. [CrossRef]
60. Ishii, H.; Wang, Y.; Huebner, K. A Fhit-ing role in the DNA damage checkpoint response. *Cell Cycle* **2007**, *6*, 1044–1048. [CrossRef]
61. Lehman, T.A.; Modali, R.; Boukamp, P.; Stanek, J.; Bennett, W.P.; Welsh, J.A.; Metcalf, R.A.; Stampfer, M.R.; Fusenig, N.; Rogan, E.M.; et al. p53 mutations in human immortalized epithelial cell lines. *Carcinogenesis* **1993**, *14*, 833–839. [CrossRef] [PubMed]
62. Tudek, B. Imidazole ring-opened DNA purines and their biological significance. *J. Biochem. Mol. Biol.* **2003**, *36*, 12–19. [PubMed]
63. Gehrke, T.H.; Lischke, U.; Gasteiger, K.L.; Schneider, S.; Arnold, S.; Müller, H.C.; Stephenson, D.S.; Zipse, H.; Carell, T. Unexpected non-Hoogsteen-based mutagenicity mechanism of FaPy-DNA lesions. *Nat. Chem. Biol.* **2013**, *9*, 455–461. [CrossRef] [PubMed]
64. Kucab, J.E.; Zou, X.; Morganella, S.; Joel, M.; Nanda, A.S.; Nagy, E.; Gomez, C.; Degasperi, A.; Harris, R.; Jackson, S.P.; et al. A Compendium of Mutational Signatures of Environmental Agents. *Cell* **2019**, *177*, 821–836. [CrossRef]
65. Conklin, K.A. Chemotherapy-associated oxidative stress: Impact on chemotherapeutic effectiveness. *Integr. Cancer Ther.* **2004**, *3*, 294–300. [CrossRef] [PubMed]
66. Loeb, L.A. Apurinic sites as mutagenic intermediates. *Cell* **1985**, *40*, 483–484. [CrossRef]
67. Gruppi, F.; Hejazi, L.; Christov, P.P.; Krishnamachari, S.; Turesky, R.J.; Rizzo, C.J. Characterization of nitrogen mustard formamidopyrimidine adduct formation of bis(2-chloroethyl)ethylamine with calf thymus DNA and a human mammary cancer cell line. *Chem. Res. Toxicol.* **2015**, *28*, 1850–1860. [CrossRef]
68. Minko, I.G.; Christov, P.P.; Li, L.; Stone, M.P.; McCullough, A.K.; Lloyd, R.S. Processing of N(5)-substituted formamidopyrimidine DNA adducts by DNA glycosylases NEIL1 and NEIL3. *DNA Repair (Amst.)* **2019**, *73*, 49–54. [CrossRef]
69. Boysen, G.; Pachkowski, B.F.; Nakamura, J.; Swenberg, J.A. The formation and biological significance of N7-guanine adducts. *Mutat. Res.* **2009**, *678*, 76–94. [CrossRef]
70. Sagher, D.; Strauss, B. Insertion of nucleotides opposite apurinic/apyrimidinic sites in deoxyribonucleic acid during in vitro synthesis: Uniqueness of adenine nucleotides. *Biochemistry* **1983**, *22*, 4518–4526. [CrossRef]
71. Schaaper, R.M.; Kunkel, T.A.; Loeb, L.A. Infidelity of DNA synthesis associated with bypass of apurinic sites. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 487–491. [CrossRef] [PubMed]
72. Shibutani, S.; Takeshita, M.; Grollman, A.P. Translesional synthesis on DNA templates containing a single abasic site. A mechanistic study of the "A rule". *J. Biol. Chem.* **1997**, *272*, 13916–13922. [CrossRef] [PubMed]
73. Obeid, S.; Blatter, N.; Kranaster, R.; Schnur, A.; Diederichs, K.; Welte, W.; Marx, A. Replication through an abasic DNA lesion: Structural basis for adenine selectivity. *EMBO J.* **2010**, *29*, 1738–1747. [CrossRef]
74. Scherthan, H.; Abend, M.; Muller, K.; Beinke, C.; Braselmann, H.; Zitzelsberger, H.; Kohn, F.M.; Pillekamp, H.; Schiener, R.; Das, O.; et al. Radiation-induced late effects in two affected individuals of the Lilo radiation accident. *Radiat. Res.* **2007**, *167*, 615–623. [CrossRef]
75. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
76. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.
77. Koboldt, D.C.; Zhang, Q.; Larson, D.E.; Shen, D.; McLellan, M.D.; Lin, L.; Miller, C.A.; Mardis, E.R.; Ding, L.; Wilson, R.K. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **2012**, *22*, 568–576. [CrossRef]
78. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef]
79. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
80. Poell, J.B.; Mendeville, M.; Sie, D.; Brink, A.; Brakenhoff, R.H.; Ylstra, B. ACE: Absolute copy number estimation from low-coverage whole-genome sequencing data. *Bioinformatics* **2019**, *35*, 2847–2849. [CrossRef]
81. Scheinin, I.; Sie, D.; Bengtsson, H.; van de Wiel, M.A.; Olshen, A.B.; van Thuijl, H.F.; van Essen, H.F.; Eijk, P.P.; Rustenburg, F.; Meijer, G.A.; et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res.* **2014**, *24*, 2022–2032. [CrossRef] [PubMed]
82. Olshen, A.B.; Venkatraman, E.S.; Lucito, R.; Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **2004**, *5*, 557–572. [CrossRef] [PubMed]
83. Tebel, K.; Boldt, V.; Steininger, A.; Port, M.; Ebert, G.; Ullmann, R. GenomeCAT: A versatile tool for the analysis and integrative visualization of DNA copy number variants. *BMC Bioinform.* **2017**, *18*, 19. [CrossRef] [PubMed]
84. Robinson, J.T.; Thorvaldsdottir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [CrossRef] [PubMed]
85. Kumar, R.; Nagpal, G.; Kumar, V.; Usmani, S.S.; Agrawal, P.; Raghava, G.P.S. HumCFS: A database of fragile sites in human chromosomes. *BMC Genom.* **2019**, *19*, 985. [CrossRef]
86. Lawrence, M.; Gentleman, R.; Carey, V. rtracklayer: An R package for interfacing with genome browsers. *Bioinformatics* **2009**, *25*, 1841–1842. [CrossRef]

87. Chiu, T.P.; Comoglio, F.; Zhou, T.; Yang, L.; Paro, R.; Rohs, R. DNAshapeR: An R/Bioconductor package for DNA shape prediction and feature encoding. *Bioinformatics* **2016**, *32*, 1211–1213. [CrossRef]

88. Cer, R.Z.; Donohue, D.E.; Mudunuri, U.S.; Temiz, N.A.; Loss, M.A.; Starner, N.J.; Halusa, G.N.; Volfovsky, N.; Yi, M.; Luke, B.T.; et al. Non-B DB v2.0: A database of predicted non-B DNA-forming motifs and its associated tools. *Nucleic Acids Res.* **2013**, *41*, D94–D100. [CrossRef]

89. Rao, S.S.; Huntley, M.H.; Durand, N.C.; Stamenova, E.K.; Bochkov, I.D.; Robinson, J.T.; Sanborn, A.L.; Machol, I.; Omer, A.D.; Lander, E.S.; et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **2014**, *159*, 1665–1680. [CrossRef]

90. Durand, N.C.; Shamim, M.S.; Machol, I.; Rao, S.S.; Huntley, M.H.; Lander, E.S.; Aiden, E.L. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* **2016**, *3*, 95–98. [CrossRef]

91. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [CrossRef] [PubMed]

92. Alexandrov, L.B.; Kim, J.; Haradhvala, N.J.; Huang, M.N.; Tian Ng, A.W.; Wu, Y.; Boot, A.; Covington, K.R.; Gordenin, D.A.; Bergstrom, E.N.; et al. The repertoire of mutational signatures in human cancer. *Nature* **2020**, *578*, 94–101. [CrossRef] [PubMed]

93. Zainal, S.N.; Alexandrov, L.B.; Wedge, D.C.; van Loo, P.; Greenman, C.D.; Raine, K.; Jones, D.; Hinton, J.; Marshall, J.; Stnebbigs, L.A.; et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* **2012**, *149*, 979–993. [CrossRef] [PubMed]

94. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [CrossRef]

95. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets-update. *Nucleic Acids Res.* **2013**, *41*, D991–D995. [CrossRef]