




Article

Unraveling the Heterogeneous Mutational Signature of Spontaneously Developing Tumors in $MLH1^{-/-}$ Mice

Yvonne Saara Gladbach ^{1,2,3} , Leonie Wiegele ⁴, Mohamed Hamed ¹ ,
Anna-Marie Merckenschläger ¹, Georg Fuellen ¹, Christian Junghans ⁴ and Claudia Maletzki ^{4,*} 

¹ Institute for Biostatistics and Informatics in Medicine and Ageing Research (IBIMA), Rostock University Medical Center, University of Rostock, 18057 Rostock, Germany; yvonne.gladbach@uni-rostock.de (Y.S.G.); mohamed.hamed@uni-rostock.de (M.H.); anna-marie.merkenschlager@uni-rostock.de (A.-M.M.); fuellen@med.uni-rostock.de (G.F.)

² Faculty of Biosciences, Heidelberg University, 69120 Heidelberg, Germany

³ Division of Applied Bioinformatics, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT) Heidelberg, 69120 Heidelberg, Germany

⁴ Department of Internal Medicine, Medical Clinic III - Hematology, Oncology, Palliative Medicine, Rostock University Medical Center, University of Rostock, 18057 Rostock, Germany; leonie.wiegele@uni-rostock.de (L.W.); christian.junghans@med.uni-rostock.de (C.J.)

* Correspondence: claudia.maletzki@med.uni-rostock.de; Tel.: +49-381-494-5764; Fax: +49-381-494-5898

Received: 6 August 2019; Accepted: 26 September 2019; Published: 2 October 2019



Abstract: Mismatch repair deficient (MMR-D) tumors exemplify the prototypic hypermutator phenotype. Owing to the high mutation rates, plenty of neo-antigens are present on the tumor cells' surface, ideally shared among different cancer types. The $MLH1$ knock out mouse represents a preclinical model that resembles features of the human MMR-D counterpart. While these mice develop neoplasias in a sequential twin-peaked manner (lymphomas > gastrointestinal tumors (GIT)) we aimed at identification of underlying molecular mechanisms. Using whole-genome sequencing, we focused on (I) shared and (II) mutually exclusive mutations and describe the process of ongoing mutational events in tumor-derived cell cultures. The landscape of $MLH1^{-/-}$ tumors is heterogeneous with only a few shared mutations being detectable among different tumor entities (*ARID1A* and *IDH2*). With respect to coding microsatellite analysis of MMR-D-related target genes, partial overlap was detectable, yet recognizing shared antigens. The present study is the first reporting results of a comparison between spontaneously developing tumors in MMR-D driven tumorigenesis. Additionally to identifying *ARID1A* as potential causative mutation hotspot, this comprehensive characterization of the mutational landscape may be a good starting point to refine therapeutic concepts.

Keywords: tumor mutational burden; exclusive/shared mutations; predicted tumor antigens

1. Introduction

Comprehensive genomic sequencing is an increasingly common practice in oncological precision medicine. The technological improvements in analyzing diseases at the genomic, transcriptomic, and epigenetic levels allow for identification of characteristic genetic changes. To detect disease-causing variants (so-called driver mutations) and discover therapeutic target genes, whole-exome sequencing (WES) is a highly meaningful sequencing method [1]. By applying this technique, the coding region of the genome is captured and sequenced at high depth. The interrogation of the exome in clinical

diagnosis raises challenges of uncovering mutational profiles of heterogeneous tumors and may allow the development of new immunotherapeutic approaches [2,3].

Closely linked to hypermutation and high immunogenicity is the *MLH1* gene, belonging to the DNA mismatch-repair (MMR) family. Germline, as well as somatic *MLH1* mutations drive tumorigenesis. The hallmarks of the resulting tumors are mismatch-repair deficiency (MMR-D) and an exceedingly high tumor mutational burden (TMB) [4]. The latter is characterized by frameshift mutations and, in consequence, by functionally impaired proteins as well as elevated frequencies of non-synonymous mutations. A high neoantigen load reflects high TMB. The tumor cells' MHC molecules, capable of eliciting anticancer T cell responses, present these mutated self-peptides. However, immune evasion, such as loss of the MHC class I subunit, beta-2-microglobulin as well as upregulation of immune-checkpoint molecules (PD-1/PD-L1) is quite common and most tumor-infiltrating T cells show signs of exhaustion [5].

Clinically, MMR-D is associated with sporadic as well as hereditary cancer, with the latter being due to mono- or biallelic MMR germline mutations. Individuals with monoallelic germline mutations suffer from Lynch syndrome and have an 80% lifetime risk of developing cancer [6]. The bi-allelic counterpart is constitutional mismatch-repair deficiency (CMMR-D), associated with a complex and slightly different tumor spectrum than that seen in Lynch-associated cancers [7,8]. As there is a high likelihood of tumorigenesis for both syndromes, attempts to delay or even prevent tumor formation are ongoing (e.g., *clinicaltrials.gov.identifiers*: NCT01885702; NCT02813824; NCT02497820; NCT03070574). Moreover, preclinical tumor models may provide new insights into tumor biology, helping to refine prevention and therapy.

MLH1^{-/-} knockout mice constitute an ideal preclinical model for hereditary tumor syndromes. With a virtually 100 % penetrance, these mice develop a variety of malignancies, including highly aggressive lymphomas in different organ sites, gastrointestinal tumors, and malignant skin lesions [9,10]. The sequential appearance of these neoplasias, often in a mutually exclusive fashion ("either-or" principle in terms of tumor type) raises the question of how the biological and/or molecular mechanisms of these neoplasias relate to each other. To gain insights into the underlying molecular alterations that affect tumor formation and thus act as drivers, we herein analyzed the number and type of somatic mutations arising at the bottom of a germline MMR-D. Exclusive as well as shared alterations were identified, explaining the heterogeneous clinical presentation.

2. Results

2.1. *In Vivo* and *Ex Vivo* Data

MLH1^{-/-} mice are prone to different tumors. Data from 90 mice reveal a heterogeneous distribution pattern and the median age of onset varies, with a high prevalence of early Non-Hodgkin T cell lymphomas, primarily arising in the thymus and spleen (43.5 and 18.5%, respectively), after which mice develop epithelial tumors of the gastrointestinal tract (GIT) tumours, primarily located in the duodenum or lymphoid skin lesions (26.0 and 12.0%, respectively). Besides, gender-specific differences are apparent with female mice developing lymphomas more frequently than GIT and vice versa in males (Figure 1A).

For our comprehensive analysis, we included a set of four tumor samples, mainly covering the spectrum observed in *MLH1*^{-/-} mice (Table 1). Since all *MLH1*^{-/-}-associated tumors harbor MHC class I on their surface, which is preserved in the resulting *in vitro* cell culture as well as in the allograft [9], phenotyping focused on immune evasion markers, such as PD-L1, CTLA-4, LAG-3, and TIM-3. As anticipated, lymphomas expressed all markers in high abundance (Figure 1B). By contrast, PD-L1 and CTLA-4 surface expression on GIT was heterogeneous and found on only 20-40% of cells, respectively (Figure 1B; referring to P11 and P12, respectively). LAG-3 and TIM-3 expression was confined to tumor-infiltrating lymphocytes and thus below 5% (Figure 1B; referring to P14 and P15, respectively and [3]). Generally, lymphocytic infiltrates in *MLH1*^{-/-}-associated GIT showed an equal distribution

between $CD3^+CD4^+$ and $CD3^+CD8^+$ T cells (Figure 1C, right bars with dashed lines and representative immunofluorescence images). However, marked differences were found between the two GIT cases #7450 and #328, nicely reflecting the heterogeneity even among GIT with similar genetic background. As can be taken from Figure 1C, the value of both cell types in #7450 (dotted plot) exceeds the mean value, while in #328 (checked plot) infiltrating cell numbers are far lower than on average.

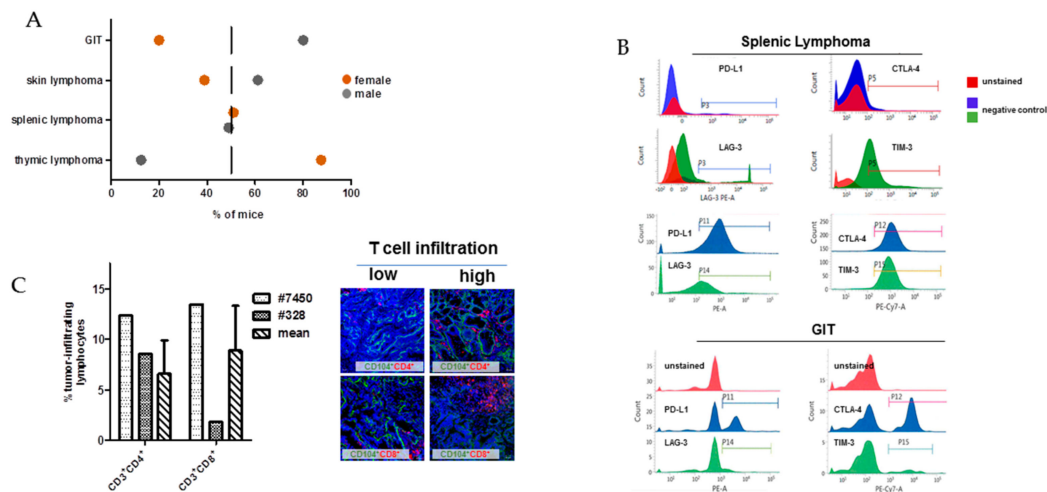


Figure 1. General distribution of tumors in $MLH1^{-/-}$ mice and tumor phenotyping of immunoregulatory markers. (A) Type and frequency of tumors in $MLH1^{-/-}$ mice; data were taken from a total of 90 mice in which spontaneous tumorigenesis was observed. Females were more likely to develop thymic lymphomas, while males were prone to gastrointestinal tract (GIT). Generalized splenic lymphomas were found in both genders. (B) Flow cytometric phenotyping of $MLH1^{-/-}$ tumors reveals a high abundance of immune-checkpoint molecules on lymphomas. The phenotype of GIT is different with only a few cells expressing these exhaustion markers. Shown data refer to % numbers of cells measured in a live gate on total white blood cells. As controls, normal lymphocytes from healthy mice were stained with evasion markers (upper histogram plots) (C) Numbers of tumor-infiltrating lymphocytes in GIT as determined by flow cytometry. Left bar—#7450; middle bar - #328; right bar—mean value of tumor-infiltrating lymphocytes in GIT of different $MLH1^{-/-}$ mice. Mean \pm SD value result from $n = 10$ individual mice. Values are given as percentage of lymphocytes measured from 50,000 events in a live gate. Representative immunofluorescence images of GIT showing either low or high T-cell infiltration. Original magnification 20 \times .

Table 1. Tumor samples included in this study.

$MLH1^{-/-}$ Number	Sex	Sample Type/Origin	Time of Onset [Weeks]
328	♂	GIT/duodenum	29.5
7450	♀	GIT/duodenum	41.0
1351	♀	Lymphoma/spleen	17.4
1444	♀	Lymphoma/skin	36.0

GIT—gastrointestinal tumor.

2.2. Mutational Landscape of $MLH1^{-/-}$ Tumors

The different tumor spectra detectable in $MLH1^{-/-}$ mice prompted us to perform a whole-exome sequencing analysis. This experiment confirmed our observations of high TMB. For the confirmatory analysis, we initially considered all point mutations. TMB ranged from 39 mutations/Megabase (MB) (in GIT #7450) to 943 mutations/MB (in GIT #328). Breaking this down to non-synonymous substitutions, i.e., missense and nonsense mutations, high mutational burden was, by the definition of

ref [11] (>10 mut/MB), preserved, with 3/4 samples even being ultra-hypermuted (>100 mut/MB) (Figure 2A).

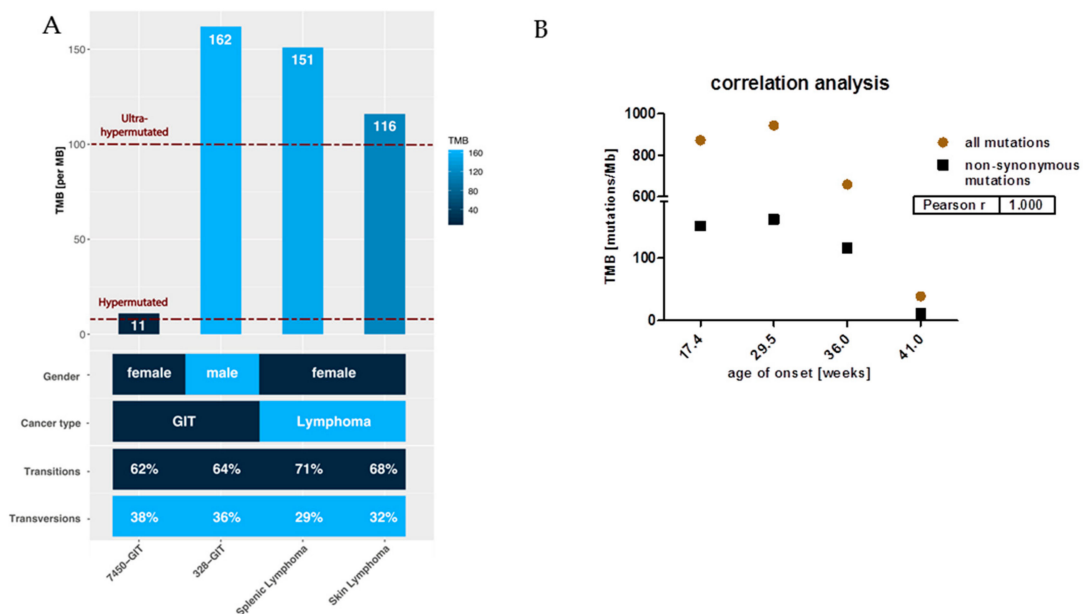


Figure 2. Mutational load and the ratio of non-synonymous mutations compared to all mutations in GIT versus lymphomas. (A) Each bar shows the tumor mutational burden (TMB) per Megabase (MB) for GIT and lymphomas, whereby cases of TMB > 100 are defined as ultra-hypermuted; these are GIT #328 and the lymphomas. Furthermore, in the heatmap, gender, cancer type, and the amount of transitions and transversions are shown in percent. (B) The number of mutations in relation to TMB and to the age of onset.

Most mutations in cancer genomes, however, are ‘passengers’ and do not bear strong imprints of selection. Apart from this, the status of hypermutation complicates the identification of driver mutations simply due to the sheer abundance of somatic variants [12]. As $MLH1^{-/-}$ -associated lymphomas showed a more aggressive in vivo growth behavior and develop considerably earlier than GIT, we expected a different molecular make-up of the lymphomas and possibly also a higher mutational burden. While the former was actually true, the mutational load was not generally higher. Indeed, the lymphomas’ mutation rates were highest, but the two GIT cases #7450 and #328 itself differed significantly from each other (Figure 2). The #328s’ mutational load was mostly comparable with that of the splenic and skin lymphomas (Figure 2). The very early formation of this GIT before the mean age of onset (~33 weeks for GIT), may explain our finding best (Figure 2B). Also, there was an inverse correlation between TMB and age of onset (Figure 2B).

Missense mutations were more frequent than nonsense mutations, and base changes were mainly due to transitions ($C > T$; $A > G$) (Figure 2A). High levels of transversion are associated with PD-L1 abundance that functions as a biomarker for immune-checkpoint inhibition. In a previous study, PD-L1 abundance was analyzed on $MLH1^{-/-}$ -tumors [13]. In line with our recent findings on high percentage numbers of transitions instead of transversions, we also detected only low-moderate PD-L1 expression (<30%), and if detectable, this was mainly confined to infiltrating immune cells, see ref. [13].

2.3. Mutation Types and Shared Mutations

Comparative whole-exome sequencing of the $MLH1^{-/-}$ GIT #7450 and #328 with the lymphoma revealed the mutanome of these samples (Figures 3 and 4). Considering the number of mutations per type, marked differences were found in the #328 compared to the #7450 GIT and both lymphomas (Figure 3A). As anticipated, the number of silent mutations was high whereas the nonsense mutation

rate was low in all four tumor samples. Focusing on the non-synonymous mutations demonstrated similar amounts of these for the #328 GIT and the splenic lymphoma, but slightly increased numbers compared to the skin lymphoma. In fact, #328 from a male mouse had the highest mutation rate in all mutation types, compared to the female counterpart GIT and to the other cancers. Furthermore, lymphomas and GIT shared 640 mutations (Figure 3B), although a high amount of single nucleotide variations (SNV) for each case was exclusive. Interestingly, the lymphomas shared a higher amount of SNV with 684 vs. only 10 in the GIT.

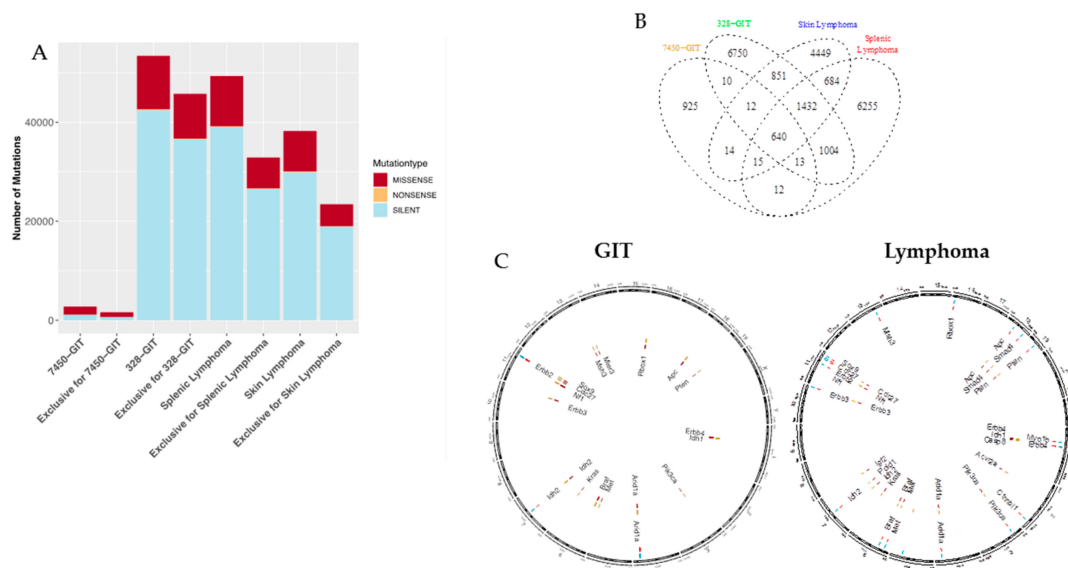


Figure 3. Mutational landscape and shared mutations in GIT versus lymphoma. (A) The statistics for each single nucleotide variations (SNV) type (missense, nonsense, and silent) are shown for the $MLH1^{-/-}$ GIT primary tumors and the lymphomas including the distribution of the exclusivity of the SNV for each condition. (B) The GIT and lymphomas are showing 640 shared SNV in addition to their exclusive mutations. (C) Pairwise comparative mutational analysis of $MLH1^{-/-}$ GIT and lymphomas. Ideogram plots show the genomic distributions of the missense mutations occurring in the annotated/known genes for the $MLH1^{-/-}$ GIT (#7450 and #328, left side). The outermost track (track 0) is the cytoband of the mm10 assembly, followed by four tracks that visualize the SNV belonging to a specific condition ($MLH1^{-/-}$ GIT primary tumors (#7450 and #328, spleen and skin lymphoma). All missense SNV for $MLH1^{-/-}$ GIT primary tumors are blue (#7450, track 1) and yellow (#328, track 3) respectively with their corresponding coordinate on the mouse reference genome mm10 cytoband. The exclusive missense SNV for the $MLH1^{-/-}$ GIT primary tumors are red (tracks 2 and 4) and annotated with a corresponding ID containing the SNV position and the affected gene name. Analogously to for the GIT, the splenic lymphoma (blue) and the skin lymphoma (yellow) are compared for all missense SNV in a pairwise fashion (right side). Red are the exclusive SNV respectively.

In dissecting the mutational landscape (Figure 3C) we initially focused on shared mutations among all cancers most likely functioning as drivers for $MLH1$ -associated tumorigenesis. With this analysis, a surprisingly low number of shared mutations were detectable. These included mutations in the following genes: *AT-rich interaction domain 1A* (*ARID1A*) and *isocitrate dehydrogenase 2* (*IDH2*). *ARID1A*, located on chromosome 4, encodes a protein of the SWI/SNF family, contributing to the large ATP-dependent chromatin-remodeling complex. It is an epigenetic modifier that functions as a tumor suppressor; gene mutations have been reported in several malignancies [14]. *IDH2* is different, usually being amplified and/or overexpressed in tumors. Mutations are associated with poor clinical outcome, well reflecting the clinical presentation of biallelic MMR-D-driven malignancies.

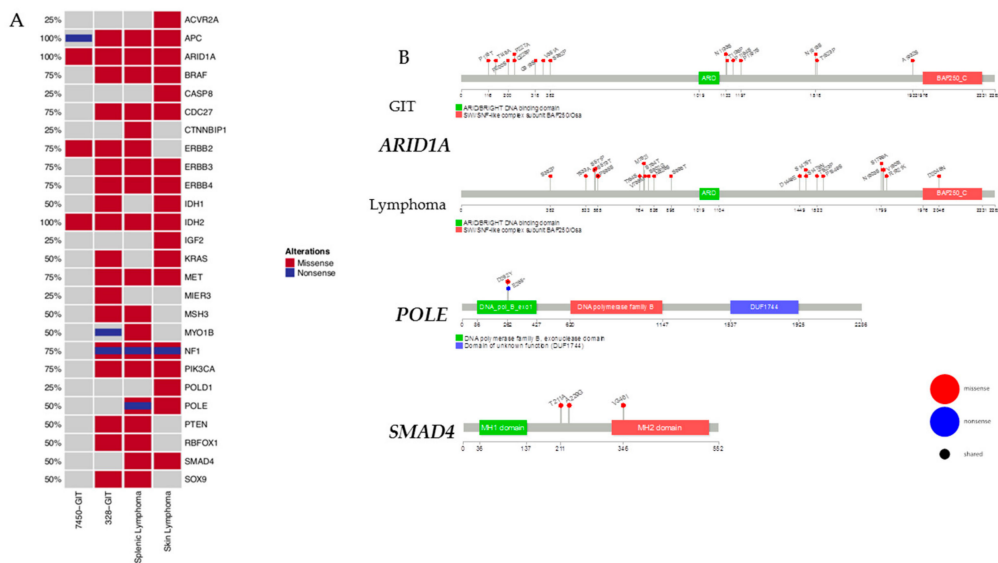


Figure 4. Specific cancer-related mutations in GIT and lymphoma. **(A)** The oncoprint presents genes with tumor-suppressive function as well as relevance for tumor initiation and progression. The analysis was done on primary GIT and lymphomas. This visualization provides an overview of the non-synonymous alterations in particular genes (rows) affecting particular individual samples (columns). Red bars indicate missense and blue bars nonsense mutations. **(B)** Prevalence and hotspot regions of GIT and lymphomas in *ARID1A*, *POLE*, and *SMAD4*. Known gene/protein domains are shown in color, other regions in dark grey. Non-synonymous mutations are depicted as nonsense by a blue lollipop and missense by a red lollipop, and a black circle in a lollipop indicates a shared locus of GIT and lymphomas entities. Each lollipop label shows the amino acid change and its location in the amino acid sequence.

2.4. Mutational Frequencies and Types of Alterations for Selected Gene Sets Based on Prior Knowledge in Mouse and Human

Next, we focused on selected genes with known relevance for tumor initiation and progression as well as tumor-suppressive function (Figure 4A). In this analysis and in accordance with the human counterpart, *MLH1*^{-/-}-associated tumors harbor additional mutations in *PIK3CA*, *BRAF* and/or *KRAS* and *ERBB3* [15]. Therefore, we created an oncoprint of the non-synonymous alterations in this selected gene set (Figure 4A). A missense mutation was found for *PIK3CA*, *BRAF*, *KRAS*, and *ERBB3*. Events associated with these genes show alterations distributed in a nearly mutually exclusive way across all samples. Unexpected are the alterations in the *NF1* gene, a negative regulator of the *RAS* signal transduction pathway that is modified in #328 (primary and cell line). These *NF1* mutations were shown to be concomitant with mutations in the oncogenes *BRAF* or *KRAS* [16,17]. Here, the same samples that had *NF1* mutations displayed *BRAF* mutations as well, but only the skin lymphoma and the GIT (#328) had additional *KRAS* mutations. In the case of *APC*, a negative regulator of the *Wnt* signaling pathway [18], we observed acquired additional missense mutations in the #7450-derived cell line that had not been detected in the primary.

The *PI3K* pathway is affected by *PIK3CA*. Further interactions of *PIK3CA* are with *AKT* and *mTOR* pathways, mutated in one GIT (#328) and both lymphomas. In gastric cancer, *PIK3CA* mutations are associated with a higher aggressiveness [19], while for the lymphomas a strong correlation with *PTEN* mutations is described [20]. Both genes were found to be mutated in the GIT (#328) and in splenic lymphoma, respectively. The same was also true for *EGFR* signaling members *ERBB2* and *ERBB3*.

SMAD4 and *POLE* were exclusively detectable in lymphomas (Figure 4A). Of note, the detected somatic *POLE* mutation was located within the exonuclease domain of this gene, executing a proofreading function to decrease the mutation rate during DNA replication. In conjunction with germline MMR-D, *POLE* mutations strongly increase the number of gene mutations in affected

tumor cells. Additionally, *SMAD4* acts as a tumor suppressor and inhibits transforming growth factor- β -mediated signaling. Mutations are associated with poor outcome and high metastatic potential. With respect to *MLH1*^{-/-} tumors, *SMAD4* mutations were constrained to lymphomas.

2.5. Pairwise Exclusively SNV in GIT and Lymphomas

The above data identify different and common mutations among *MLH1*^{-/-} cancers. We examined the pairwise exclusive mutations among GIT or lymphomas in-depth (Figure 3C). While the #328 GIT had fewer mutations per gene, this tumor had generally more genes affected such as *APC*, *PTEN*, *ERBB4*, *IDH1*, *PIK3CA*, *MET*, *BRAF*, *KRAS*, *ERBB3*, *NF1*, *CDC27*, *SOX9*, *MSH3*, *MIER3*, and *RBFOX*. Remarkable is that although both primary tumors have shared genes affected by mutations, the positions of the SNV are exclusive for each of the tumors (shown in red with the corresponding annotation containing the position and the gene, all SNV are in yellow), likely implicating random distribution during oncogenesis.

A pairwise comparison of the splenic and skin lymphomas showed that they share the mutated genes *APC*, *SMAD4*, *PTEN*, *ERBB4*, *PIK3CA*, *ARID1A*, *BRAF*, *MET*, *IDH2*, *ERBB3*, *NF1*, and *CDC27*. However, exclusive changes were also observed. Each of the lymphomas has SNV in six exclusive genes. Therefore, we analyzed in detail the processes and functions the exclusively mutated genes are involved in. The splenic lymphoma exhibits SNV in *MYO1B*, *CTNBL1*, *ERBB2*, *SOX9*, *MSH3*, and *RBFOX1*, involved in cell migration, the spliceosome, cellular responses, chondrocytes differentiation, and MMR. Exclusive SNV for the skin lymphoma were found in the following genes: *IDH1*, *CASP8*, *ACVR2A*, *KRAS*, *IGF2*, and *POLD1*. All of these genes are part of the electron acceptor, cell apoptosis, oncogene, growth, replication, and repair.

2.6. Prevalence and Hotspot Regions of GIT and Lymphomas in *ARID1A*, *POLE*, and *SMAD4*

Mutational hotspots provide more insights into the underlying biology together with the mutations in the domains, such as binding domains. Mutations of major interest for the GIT and lymphomas are appearing in *ARID1A* (Figure 4B). Significant differences between the GIT and lymphomas are the loci of the mutations as well as that more loci are affected in the lymphomas. One amino acid change, D2046N, is located in the SWI/SNF-like complex subunit BAF250/Osa, the ATP-dependent chromatin remodeling complex that regulates gene expression through nucleosome remodeling. Furthermore, the splenic and skin lymphomas share the amino acid changes S571P and P585S. All other mutated loci are found in the putative disordered regions of the *ARID1A* gene.

Figure 4B illustrates the mutational hotspots of the mutually exclusive lymphoma-associated genes *POLE* and *SMAD4*. In *POLE*, we found the shared amino acid change D262Y for the splenic and skin lymphoma. Furthermore, the splenic lymphoma has a nonsense mutation, E265*, in the DNA polymerase B exonuclease domain that adopts a ribonuclease H type fold [21]. Interestingly, the lymphomas show an amino acid change, V346I, in the MH2 domain implicated in the control of cell growth [22] of *SMAD4*. Additional loci are in the putative disordered regions, T211A and A229G.

2.7. Somatic Mutation Distribution and Nucleotide Changes in Chromosome X

In addition to the TMB calculations, we analyzed the mutational load per chromosome. This mutational load was normalized based on the size of the respective chromosome, and the mutation rate per MB plotted for all four entities (Figure S1). Therefore, this presentation of the chromosomal distribution of SNV shows that for all entities, chromosome 11 accumulated the highest numbers of somatic mutations per MB. Given the fact of differences in gender and a possible hyper-mutation of chromosome X, we investigated chromosome X further for the nucleotide changes in pyrimidines C and T (Figure S2). C is also responsible for the deamination of the X-chromosome. The coloring of the bars represent the proportions of the six possible nucleotide changes of the pyrimidines (C > A, C > G, C > T, T > A, T > C, and T > G) for chromosome X. Comparing the GIT, we see that T > C is less frequent in the #328 than in the #7450. In turn, the changes C > A and T > G are more frequent. The

lymphomas show only minor differences in the nucleotide changes for C > T and T > C, whereby the skin lymphoma had slightly lower changes in this latter than the splenic lymphoma.

2.8. Microsatellite Instability (MSI) Pathway—Prognostic, Predictive and Therapeutic Implications

To identify target genes for neoantigen-based vaccination approaches, we analyzed the SNVs in genes of the MSI pathway [23,24] as well as genes associated with the MSI status [25] and survival (based on hazard ratio in the human counterpart) (Tables 2 and 3). The SNVs shared among all cancer entities were found in *ARID1A*, *BCL2*, *PIK3CA*, and *SMAD4*. *ARID1A* shows more silent than missense SNVs for the 7450-GIT, whereby all the other show missense SNVs. These genes are associated with overall survival (OS), disease-free survival (DFS), and disease-specific survival (DSS), having SNVs, the prognosis gets worse. The same holds true for *PIK3CA*, but it is only associated with recurrence-free survival. *SMAD4*, however, ensures a slightly better prognosis, since it endeavors only silent SNVs. *BCL2* shows more clearly the differences between the female and male mice since it has SNVs that are not classifiable, whereby the other show silent with or without additional missense SNVs. It would ensure at least an event-free survival.

Table 2. Shared SNVs in *MLH1*^{-/-} cancer among the MSI pathway and potential association with survival.

Gene	7450-GIT	328-GIT	Splenic Lymphoma	Skin Lymphoma	Survival		
					OS	DFS	DSS
<i>ARID1A</i>	SILENT/MISSENSE	MISSENSE	MISSENSE	MISSENSE	OS	DFS	DSS
<i>BCL2</i>	NONE	SILENT/MISSENSE	SILENT	SILENT/MISSENSE			
<i>PIK3CA</i>	SILENT	SILENT/MISSENSE	SILENT/MISSENSE	SILENT/MISSENSE		DFS	
<i>SMAD4</i>	SILENT	SILENT	SILENT	SILENT/MISSENSE	OS		DSS

Based on COX *p*-value < 0.05; OS—overall survival; DFS—disease-free survival; DSS—disease-specific survival.

Table 3. Shared SNVs in *MLH1*^{-/-} lymphomas among the MSI pathway and potential association with survival.

Gene	Splenic Lymphoma	Skin Lymphoma	Survival		
			OS	DFS	DSS
<i>AKT2</i>	SILENT/MISSENSE	SILENT/MISSENSE		DFS	DSS
<i>CASP4</i>	NA	NA	OS		
<i>CASP9</i>	MISSENSE	NA			
<i>CERS1</i>	NA	NA		DFS	
<i>ECHS1</i>	SILENT	NA			
<i>EGR1</i>	SILENT	SILENT			
<i>INADL</i>	NA	NA	OS	DFS	DSS
<i>MAPK1</i>	SILENT	SILENT			
<i>RAC1</i>	SILENT	SILENT			DSS
<i>RHOA</i>	SILENT	SILENT			

Based on COX *p*-value < 0.05; OS—overall survival; DFS—disease-free survival; DSS—disease-specific survival.

Poorer prognosis is expected for the lymphomas, due to the identification of exclusive genes within the MSI pathway, MSI status and survival associated genes. The same SNV classes have been identified for *AKT2*, *EGR1*, *INADL*, and *MAPK1* lowering a good prognosis for survival due to their association with OS, RFS, and EFS. Besides come same SNV classes in *CASP4* (OS), *CERS1* (RFS), and *MAPK1* (OS and RFS). Remarkable are the different SNV classes for *CASP9* and *ECHS1*, not associated with survival yet. For *CASP9*, there is a change from a missense SNV to a not classified impact, and the same goes for *ECHS1*, whereby the change is from silent SNVs to a not classified impact.

2.9. Ongoing Mutations in *MLH1*^{-/-} Tumors

To pursue the ongoing event of acquiring novel mutations directly, we subsequently compared the mutational profile of selected primary tumors and their corresponding cell line upon in vitro culture. As anticipated and in conjunction with observations from clinical CMMR-D samples, numbers of somatic mutations continually increased (Figure 5A). As for the GIT case #7450, which harbors the lowest mutation numbers, TMB more than doubled in the derived cell line. The same was true for coding microsatellite (cMS) mutations, showing a gradual increase during in vitro passage (Figure 5B).

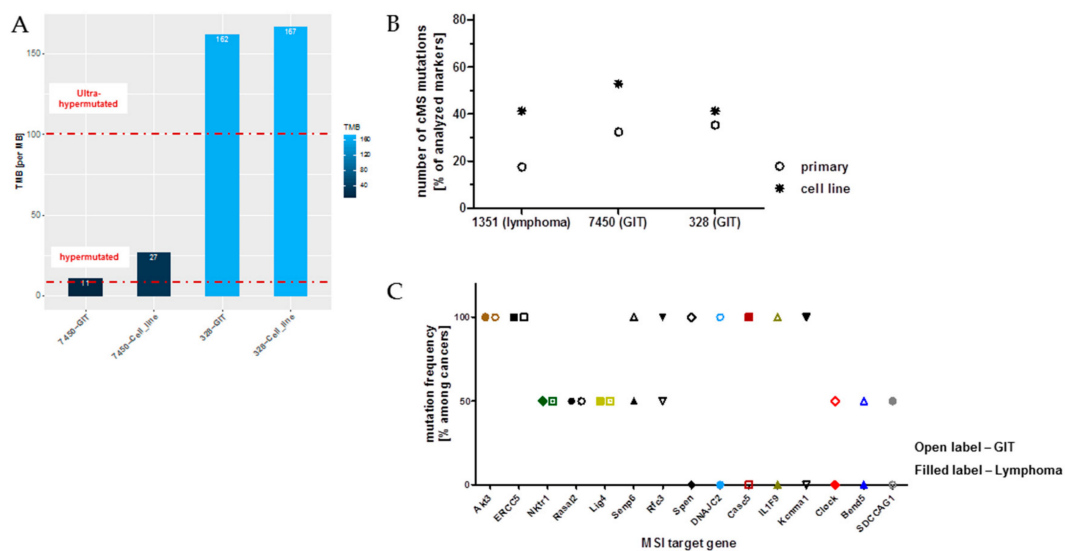


Figure 5. Mutational load and the ratio of non-synonymous mutations compared to all mutations in GIT versus the corresponding cell line. (A) The barplot shows the tumor mutational burden (TMB) per MB for the GIT and the corresponding cell lines, revealing ultra-hypermutation for GIT #328 and its corresponding cell line (TMB > 100). Furthermore, in the heatmap, gender, cancer type, as well as transitions and transversions are shown in percent. (B) Depicted are the percentages of cMS mutations in a panel of MSI target genes originally described in [9]. (C) Mutation frequency of individual cMS markers among GIT and lymphomas included in this study. Shared mutations were detected in 5/26 totally analyzed markers. Open label—GIT; filled label—lymphoma.

Finally, the cMS mutational profile was examined on a panel of primary tumors, i.e., lymphomas and GIT ($n = 10$ cases each). With this analysis, mutations were detectable in 15/26 markers, with 5 markers being shared among both cancer types (Figure 5C). Of note and of particular relevance, only two genes had the same mutation in their cMS marker of an MSI target gene, thus being in conjunction with our WES data of minor entity-overlapping mutational events. These are *AKT3*, a RAC-gamma serine/threonine-protein kinase with relevance in the maintenance of cellular homeostasis (balancing survival and apoptosis) and the endonuclease *ERCC5* (Excision Repair 5), involved in DNA excision repair. Mutations in this gene increase susceptibility for skin cancer development. Given the biological function of these two genes, their applicability as a target structure is an issue worth being determined.

2.10. In Vivo Vaccination Approach

Finally, mice with established GIT were included in a vaccination trial. Based on our previous study on minor entity-overlapping therapeutic potential between GIT and lymphomas [3], we here focused on cellular lysates made of two different GIT cell lines, namely *MLH1*^{-/-} A7450 and 328.

Repetitive application of the lysate prolonged survival of both treatment groups (Figure 6). Direct comparison of the individual lysates revealed better outcome in the *MLH1*^{-/-} A7450 treatment group. Median survival was 8.3 weeks (vs. 4.0 in the control group) and thus statistically significant longer even compared to the 328-treatment arm. Although not analyzed in detail here, accompanying in vivo

imaging at day 28 of therapy revealed disease control only in $MLH1^{-/-}$ A7450 vaccinated mice, while tumors in 328 treatment arm tended to progress (*personal observation*).

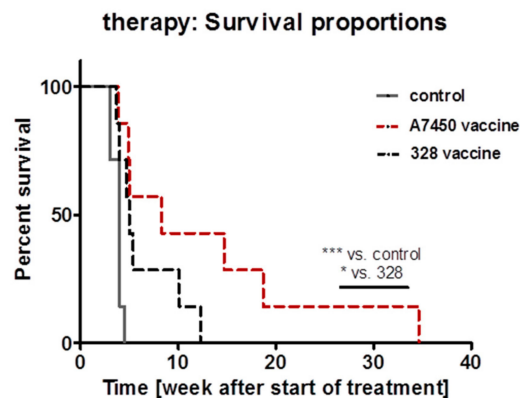


Figure 6. Kaplan–Meier survival curve analysis of vaccinated and control mice. Mice received repetitive local applications of the vaccine (10 mg/kg bw, s.c., $n = 7$ mice $MLH1^{-/-}$ A7450 and $n = 7$ mice $MLH1^{-/-}$ 328, respectively). Control mice were left untreated ($n = 7$ mice/group). *** $p < 0.01$ control vs. $MLH1^{-/-}$ A7450 Log-rank (Mantel-Cox) Test; * $p < 0.05$ $MLH1^{-/-}$ A7450 vs. $MLH1^{-/-}$ 328 Gehan–Breslow–Wilcoxon Test.

These findings confirm our molecular data on a biological level, in which the mutational signature predicts vaccination efficacy.

3. Discussion

Due to the biallelic germline MMR-D and the accordingly complex tumor spectrum, $MLH1^{-/-}$ mice are frequently considered a better model for CMMR-D, instead of Lynch Syndrome. In men, $MLH1$ has the highest frequency among monoallelic MMR mutation carriers and also the highest cumulative risk to develop any cancer [26]. Despite some differences, $MLH1^{-/-}$ mice combine several key features of both syndromes. The recently detected heterogeneity of Lynch-associated cancers, being highly immunogenic or not (i.e., “hot”, “warm” or “cold” depending on the mutational load and signature) [27], additionally argues in favor of using these mice as a model for Lynch Syndrome as well. All malignant human CMMR-D cancers are ultra-hypermuted. However, this is so far not known for tumors arising in the context of Lynch syndrome. Apart from this, there is growing evidence of individual differences between patients, likely being dependent on the underlying MMR defect [28]. In line with this broad heterogeneity among MMR-D-related malignancies, we here show that $MLH1^{-/-}$ tumors (I) are either hyper- or ultra-hypermuted, (II) have an entirely individual mutational profile and (III) harbor different mutational hotspots in affected genes.

Of note, along with the different sites of manifestation, a gender-specific prevalence is evident in mice, where thymic lymphomas primarily develop in females, and 80% of GIT manifest in males. Apart from hormonal-differences (i.e., estrogen-pathway activation vs. androgen-pathway activation), X-chromosome specific alterations may provide a rationale for the different prevalence. The X-chromosome of $MLH1^{-/-}$ mice is indeed highly mutated. However, we could not find evidence for a hypermutated X-chromosome. We hypothesize that the mutation rate as well as the type of mutations on the X chromosome is a cancer-specific feature for the lymphomas and represents a subgroup in the GIT primary tumors. This somatic mosaicism is the result of “self-promoting” mutations that favors (or hinders) expansion of the mutant clone and shapes their resulting genomic make-up in males or females. In addition to the gender, several biological characteristics argue in favor of this hypothesis, leaving technical artifacts as the sole or dominant explanation for the different mutational load very unlikely: (I) the age of onset; (II) the number of tumor-infiltrating lymphocytes in the primary and (III) the experimental follow-up study showing different vaccination efficacy in tumor-bearing $MLH1^{-/-}$

mice. In here, the GIT lysate from the #7450 case was much more effective than the tumor lysate of #328. Hence, the mutational signature likely predicts vaccination efficacy. Still and without any doubt, this has to be confirmed in a larger series on more samples per entity and gender.

In the only prior study on the mutational landscapes of *MLH1*^{-/-} murine tumors, spontaneous as well as irradiation-induced T cell lymphomas were analyzed. Results identified many genes commonly mutated but not previously implicated, such as genes involved in NOTCH (e.g., *Notch1*) and PI3K/AKT (e.g., *Pten*, *Akt2*) signaling pathways as well as frameshift mutations in mononucleotide repeat sequences [10]. Our in-depth sequencing analyses adds to these findings and broadens the knowledge on the mutational spectrum of MMR-D related malignancies apart from lymphomas.

By performing in-depth sequencing on lymphomas and GIT, most altered genes involved *APC*, *PTEN*, *PIK3CA*, *Her2*-related genes, *BRAF* and/or *KRAS*. However, it is often difficult to determine, which mutations arose first or whether their order is essential in driving tumorigenesis. While *MLH1*^{-/-} mice are inbred and all descendants harbor the same germline mutation, we expected a comparable mutational profile among cancers. This was, nevertheless, not the case with a comparably low number of shared mutations among analyzed cancer types. Indeed, entity-overlapping mutations were exclusively detectable in *ARID1A* and *IDH2*. In men, *IDH2* mutations are linked to brain tumors and acute myeloid leukemia, while *IDH1* mutations may drive tumor progression. IDH-mutated tumors are clinically and genetically distinct from their wildtype counterpart. *ARID1A* is a bona fide tumor suppressor and a frequent mutational hotspot in many cancers, including those of the breast, gastrointestinal tract and ovary [29]. *ARID1A* functions in MMR regulation, as its loss is associated with MSI. It is supposed to be a driver gene, occurring secondary to MMR-D in gastric cancers, representing an alternative oncogenic pathway to p53 alteration [30]. Neither of the tumors examined here showed *TP53* mutations, confirming this assumption. *ARID1A* mutations in *MLH1*^{-/-}-associated tumors are distributed along the entire length of the gene, thus distinguishing it from typical MSI target genes, a finding also described in the human counterpart [31].

Loss of *ARID1A* shortens time to cancer-specific mortality in men, where mutations are associated with both sporadic and hereditary MSI cancers [32]. Though it is still unknown whether CMMR-D cases harbor *ARID1A* mutations as well, this observation fits with our finding on *MLH1*^{-/-} mice having a significantly shortened life expectancy.

Comparable analysis done by others on human cancers revealed ten overlapping mutations, all of which are InDels involving G7 or C6 repeats [33]. These mutations are potentially caused by MSI, and although not analyzed in detail here, we also observed *ARID1A* mutations in our own collection of MMR-D-associated clinical CRC cases (22% of cases harbor mutations in the G7 repeat and have additionally wildtype *TP53*). *ARID1A* is thus a likely target for personalized medicine in cancer treatment. *ARID1A*-deficiency-based synthetic lethal, targeted drugs are under clinical investigation to suppress cell growth and promote apoptosis [33] (see also clinicaltrials.gov). Whether this provides a real opportunity for *ARID1A*-deficient MMR-D-related cancer will need to be tested prospectively.

Another interesting finding of this study was the high prevalence of transitions instead of transversions among gene mutations. Although this is commonly seen in non-smoking associated cancers, we wish to stress the close link between transitions (mainly C-to-T) and MSI as well as *POLE/POLD1* mutations in men [34]. While these alterations are detectable in *MLH1*^{-/-}-associated tumors similar mechanisms can be assumed. As for the latter, the relevance of somatic *POLE/POLD1* mutations in the context of MMR-D has only begun to be enlightened. In clinical cases of CMMR-D, most tumors harbor *POLE* mutations affecting the exonuclease domain or domains important to the intrinsic proofreading activity [35]. These cells rapidly and catastrophically accumulate point mutations, yielding the ultra-hypermutated phenotype – quite similar to our observations. Cells mutate continuously, raising the question of how many mutations can be tolerated until a maximum threshold is reached that creates the Achilles' heel.

All mutations described so far may have prognostic, predictive and surveillance potential. Besides, a plethora of low-frequency variants whose functional consequences and clinical actionability are

unknown were found as well. We thus conclude that for $MLH1^{-/-}$ -driven oncogenesis, many mutations are mere passengers with no functional consequence.

A point worth mentioning is the fact that MMR-D associated lymphomas in men are rarely documented in the literature and most of them develop in patients suffering from CMMR-D. In here, however, hematological malignancies, are common and even in this small group of patients, mostly associated with founder *MLH1* and/or *MSH2* mutations [36]. A recent study described the occurrence of MSI/MMR-D in de novo Hodgkin lymphoma, but at a significantly higher frequency in Hodgkin's lymphoma survivors developing therapy-related colorectal cancers [37]. *MLH1* promoter hypermethylation, but also double somatic mutations are the leading causes of MMR-D [38]. Highlighting the relevance of *MLH1*, one study also demonstrates the failure of a specific immune-chemotherapy regimen in follicular lymphoma patients showing *MLH1* gene polymorphisms [39]. In addition to germline MMR-D, the frequency of MMR-D arising secondary to anticancer treatment, is thus supposedly higher than previously predicted, considering this a clinically relevant finding. This may finally witness a resurgence of interest to apply and develop alternative therapeutic regimens.

With respect to our study, some neoantigens are auspicious. These include AKT3, ERCC5 and maybe ARID1A as general MMR-D targets, but also POLE as lymphoma-specific neoantigen. The finding that this particular mutation was confined to lymphomas may not only explain why this type of malignancies arises earlier in mice than GIT but also represent a target structure worth being included.

Summarizing our findings, we identified a gender-specific aspect in the tumor spectrum, with resulting tumors displaying their own idiosyncratic mutational landscape. The identified shared and mutually exclusive mutations warrant further investigations for therapeutic approaches ideally combined with immune-stimulating and/or -restoring agents.

4. Critical Limitation of the Study

A major limitation of the current study is the low sample number of each tumor type included in WES analysis ($n = 1$ plus 2 sample-derived cell lines at very low passage). Nevertheless, this is a pilot study aiming to explore the mutational signature of $MLH1^{-/-}$ -associated murine tumors - comparable analyses are to the best of our knowledge not documented in the literature. With respect to cMS mutations, which were done on a larger cohort, quite similar heterogenic distribution was found between individual tumor samples. Another limitation is the fact that all mice are inbred and thus harbor the very same germline mutation in the *MLH1* gene. In men, the germline mutational spectrum is much more complex (i.e., *MLH1*, *MSH2* +/- *EpCAM*, *PMS2*, and *MSH6*) and heterogeneous (different hotspots, with some of them still being classified as variance of uncertain significance). As a result, patients present with a diverse clinical presentation/organ manifestation and age of onset. Therefore, a direct comparison with human samples is hardly realizable. Still, the heterogeneity observed between individual *MLH1*-associated cancers well reflects the human situation and underlines the individual character of malignancies. By identifying some overlap between the murine and human oncoprint these analyses will likely help to get a deeper understanding of the genetic or genomic make-up of MMR-D-related cancers.

5. Materials and Methods

5.1. In Vivo Mouse Model and Sample Acquisition

Homozygous *MLH1*-deficient mice were bred in the animal facilities (University of Rostock) under specified pathogen-free conditions. Trials were performed in accordance with the German legislation on protection of animals and the Guide for the Care and Use of Laboratory Animals (Institute of Laboratory Animal Resources, National Research Council; NIH Guide, vol.25, no.28, 1996; approval number: LALLF M-V/TSD/7221.3- 1.1-053/12 and 026/17, respectively). Tumor samples for sequencing analysis were obtained from four individual $MLH1^{-/-}$ mice upon spontaneous development. Two

mice suffered from a lymphoma (spleen and skin), and the remaining two mice had gastrointestinal tumors both located in the duodenum. All tumors were at an advanced stage at resection. Samples were snap frozen in liquid nitrogen and stored at -80°C . Isolation of gDNA was done by GATC (samples: 7450 primary GIT; A7450 P15 cell line) and CEGAT (samples: 328 primary; 328 cell line; 1351 lymphoma (spleen); 1444 lymphoma (skin)), respectively, prior to the sequencing analyses.

5.2. Whole-Exome Sequencing (WES) Analysis

Data preprocessing: First, the read quality of the raw sequencing data was investigated with FastQC v.0.11.5 [3] to adjust the adapter trimming. The sequencing adapters were trimmed using Skewer v.0.2.2 [40], for the reads after demultiplexing with Illumina bcl2fastq v.2.1919 (Illumina, San Diego, CA, USA). Reads with a Phred quality score ≥ 30 were considered as high quality. The whole-exome reads were mapped to the published mouse genome build ENSEMBL mm10.75 reference using Burroughs Wheeler Aligner (BWA) [41] with the bwa-mem algorithm v.0.7.2 [42]. Reads with multiple mappings and the same mapping score have been discarded from further analysis. Additionally, PCR duplicate reads have been removed by using Samtools v.0.1.18 [41] to prevent artificial coverage brought on by the PCR amplification step during library preparation and to avoid its impacts on identifying false positive mutations. SNV calling was performed using the VarScan v.2.4.2 [43], and the detected variants were annotated based on their gene context using snpEff v.4.3 [44]. To compare these data with data from our previously published paper [3], the mm9 SNV coordinates were converted to the corresponding mm10 coordinates with hgliftover v.30-10-2018 [45,46]. The cell line A7450 from the batch was processed likewise [3].

5.3. Data Visualization

For the $\text{MLH1}^{-/-}$ GIT primary tumor, derived cell line, spleen and skin lymphoma, the summarized mutational profiles with the corresponding mutation type were summarized as stacked bar plot with ggplot2 [47], as Venn diagram with VennDiagram [48] as well as whole-genome ideograms with the circlize [49] R package. Based on the mm10 mouse assembly, provided by the circlize package [49], the filtered mutational profiles are presented as ideograms.

All four different mutational profiles were filtered for the exclusive SNV for each condition separately. Additionally, mutation filters such as the mutation type (missense and nonsense) as well as those mutations occurring in known annotated genes have been applied to mutational profiles.

Additional details like TSTV (transition and transversion) ratio, allele frequency spectrum, base changes, and variant quality for each mutational profile were obtained using the vcf.iobio tool [50].

Furthermore, mapping the mutations and its statistics on a linear gene product (proteins of interest) was done with a ‘lollipop’ mutation diagram generator [51]. Based on the knowledge from the human MMR-D counterpart and general involvement in tumorigenesis, the genes for further analysis were chosen with high probability to mutate.

5.4. Hypermutation and TMB Calculation

We adopted the following definition for calculating the TMB [52–54]:

$$\frac{\# \text{ all non - synonymous SNVs on target}}{\text{covered exome by sequencing kit}} \quad (1)$$

Simply, we used intersectBed from the BEDTools suite v.2.26.0 [55] for ensuring the SNV being on target regions of the Agilent SureSelect Mouse All Exon Kit mm10 (Santa Clara, CA, USA) and then summing up the number of somatic mutations per MB of the transcriptome mm10 (48.3 MB).

We use the definition of hypermutation as follows [34,53,54]:

$$\text{per MB}_{\text{cancer entity}} = \begin{cases} \text{hypermutated} & \text{if non - synonymous SNVs} > 10 \\ \text{ultra - hypermutated} & \text{if non - synonymous SNVs} > 100 \end{cases} \quad (2)$$

where

$$\text{cancer entity} \in \{\text{GIT, derived cell line, lymphoma}\} \quad (3)$$

5.5. Coding Microsatellite (cMS) Frameshift Mutation Analysis

Fragment length analysis was done from multiplexed PCRs of gDNA from tumor and normal tissue as described [9]. To identify potential MLH1 target genes, a panel ($n = 26$) was screened. Primers were designed using Primer3 software to yield short amplicons (≤ 200 bp).

5.6. In Vivo Vaccination with MLH1^{-/-} Tumor Lysates

Tumor cell lines MLH1^{-/-} A7450 and 328 were cultured and harvested prior to lysis using repetitive freeze/thaw cycles ($n = 4$). Protein lysates were gamma irradiated (60 Gy) and frozen immediately in aliquots at -80 °C before in vivo application. Thereafter, mice with gastrointestinal tumors (GIT) confirmed were conducted to vaccination according to a protocol described before [3]. Briefly, mice received four weekly injections of the vaccine (10 mg/kg bw, s.c.) in the first phase. Vaccination was continued (2.5 mg/kg bw, biweekly) until tumors progressed. Control mice were left untreated.

6. Conclusions

This study describes the mutational spectrum of MLH1^{-/-}-associated tumors, spontaneously developing in mice. While these tumors arise at the bottom of the same germline mutation, the clinical presentations as well as resulting molecular alterations are heterogeneous, and, thus, likely being directly linked. By performing in-depth whole-exome sequencing analysis, we here identified, for the first time, a common mutational hotspot. *ARID1A* constitutes a potentially causative mutation, shared among different MLH1^{-/-}-associated tumors and, thus, irrespective of the origin. This finding is of particular relevance for subsequent therapeutic and—even more important—prophylactic vaccination approaches aiming at entity-overlapping treatment of MLH1^{-/-}-related tumors.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2072-6694/11/10/1485/s1>, Figure S1: Distribution of somatic mutations in GIT and lymphomas in genomes of female versus male mice, Figure S2: Distribution of somatic mutations in GIT and lymphomas on chromosome X of female versus male mice.

Author Contributions: Y.S.G.: participated in the outline of the manuscript, developed bioinformatics study analysis, performed sequencing data preprocessing, data analysis/interpretation, wrote the manuscript. L.W.: performed fragment length analysis, participated in in vivo data acquisition and data analysis. M.H.: directed the bioinformatics study analysis, contributed to the discussion and interpretation of the results and paper finalization. A.-M.M.: participated in literature review and drafted the manuscript. G.F.: participated in paper finalization and critically revised the manuscript. C.J.: participated in paper finalization and critically revised the manuscript. C.M.: designed study, the outline of the manuscript, performed in vivo experiments, data analysis/ interpretation, wrote the manuscript.

Funding: This work was supported by the German research foundation to CM [DFG grant number MA5799/2-1 and MA5799/2-2].

Acknowledgments: The authors gratefully thank Ilona Klamfuss for breeding MLH1 mice.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

CMMR-D—constitutional mismatch repair deficiency, cMS—coding microsatellite, MB—Megabase, MMR—mismatch repair, MMR-D—mismatch repair deficiency, TMB—tumor mutational burden, GIT—gastrointestinal tumor, SNV—single nucleotide variant, WES—whole-exome sequencing.

References

1. Gupta, S.; Chatterjee, S.; Mukherjee, A.; Mutsuddi, M. Whole exome sequencing: Uncovering causal genetic variants for ocular diseases. *Exp. Eye Res.* **2017**, *164*, 139–150. [[CrossRef](#)] [[PubMed](#)]
2. Roche, M.I.; Berg, J.S. Incidental Findings with Genomic Testing: Implications for Genetic Counseling Practice. *Curr. Genet. Med. Rep.* **2015**, *3*, 166–176. [[CrossRef](#)] [[PubMed](#)]

3. Maletzki, C.; Gladbach, Y.S.; Hamed, M.; Fuellen, G.; Semmler, M.-L.; Stenzel, J.; Linnebacher, M. Cellular vaccination of MLH1^{-/-} mice—An immunotherapeutic proof of concept study. *Oncoimmunology* **2018**, *7*, e1408748. [[CrossRef](#)]
4. Boland, C.R.; Goel, A. Microsatellite instability in colorectal cancer. *Gastroenterology* **2010**, *138*, 2073–2087. [[CrossRef](#)] [[PubMed](#)]
5. Ozcan, M.; Janikovits, J.; von Knebel Doeberitz, M.; Kloor, M. Complex pattern of immune evasion in MSI colorectal cancer. *Oncoimmunology* **2018**, *7*, e1445453. [[CrossRef](#)]
6. Kloor, M.; Von Knebel Doeberitz, M. The immune biology of microsatellite-unstable cancer. *Trends Cancer* **2016**, *2*, 121–133. [[CrossRef](#)]
7. Bakry, D.; Aronson, M.; Durno, C.; Rimawi, H.; Farah, R.; Alharbi, Q.K.; Alharbi, M.; Shamvil, A.; Ben-Shachar, S.; Mistry, M.; et al. Genetic and clinical determinants of constitutional mismatch repair deficiency syndrome: Report from the constitutional mismatch repair deficiency consortium. *Eur. J. Cancer* **2014**, *50*, 987–996. [[CrossRef](#)]
8. Wimmer, K.; Rosenbaum, T.; Messiaen, L. Connections between constitutional mismatch repair deficiency syndrome and neurofibromatosis type 1. *Clin. Genet.* **2016**, *91*, 507–519. [[CrossRef](#)]
9. Maletzki, C.; Beyrich, F.; Hühns, M.; Klar, E.; Linnebacher, M. The mutational profile and infiltration pattern of murine MLH1^{-/-} tumors: Concurrences, disparities and cell line establishment for functional analysis. *Oncotarget* **2016**, *7*, 53583–53598. [[CrossRef](#)]
10. Daino, K.; Ishikawa, A.; Suga, T.; Amasaki, Y.; Kodama, Y.; Shang, Y.; Hirano-Sakairi, S.; Nishimura, M.; Nakata, A.; Yoshida, M.; et al. Mutational landscape of T-cell lymphoma in mice lacking the DNA mismatch repair gene Mlh1: No synergism with ionizing radiation. *Carcinogenesis* **2019**, *40*, 216–224. [[CrossRef](#)]
11. Campbell, B.B.; Light, N.; Fabrizio, D.; Zatzman, M.; Fuligni, F.; de Borja, R.; Davidson, S.; Edwards, M.; Elvin, J.A.; Hodel, K.P.; et al. Comprehensive Analysis of Hypermutation in Human Cancer. *Cell* **2017**, *171*, 1042–1056. [[CrossRef](#)] [[PubMed](#)]
12. Martincorena, I.; Raine, K.M.; Gerstung, M.; Dawson, K.J.; Haase, K.; Van Loo, P.; Davies, H.; Stratton, M.R.; Campbell, P.J. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **2017**, *171*, 1029–1041. [[CrossRef](#)] [[PubMed](#)]
13. Maletzki, C.; Wiegele, L.; Nassar, I.; Stenzel, J.; Junghanss, C. Chemo-immunotherapy improves long-term survival in a preclinical model of MMR-D-related cancer. *J. Immunother. Cancer* **2019**, *7*, 8. [[CrossRef](#)] [[PubMed](#)]
14. Ayhan, A.; Mao, T.-L.; Seckin, T.; Wu, C.-H.; Guan, B.; Ogawa, H.; Futagami, M.; Mizukami, H.; Yokoyama, Y.; Kurman, R.J.; et al. Loss of ARID1A expression is an early molecular event in tumor progression from ovarian endometriotic cyst to clear cell and endometrioid carcinoma. *Int. J. Gynecol. Cancer* **2012**, *22*, 1310–1315. [[CrossRef](#)]
15. Fakhri, B.; Lim, K.H. Molecular landscape and sub-classification of gastrointestinal cancers: A review of literature. *J. Gastrointest. Oncol.* **2017**, *8*, 379–386. [[CrossRef](#)] [[PubMed](#)]
16. Zenonos, K.; Kyprianou, K. RAS signaling pathways, mutations and their role in colorectal cancer. *World J. Gastrointest. Oncol.* **2013**, *5*, 97–101. [[CrossRef](#)]
17. Hayashi, T.; Desmeules, P.; Smith, R.S.; Drilon, A.; Somwar, R.; Ladanyi, M. RASA1 and NF1 are preferentially co-mutated and define a distinct genetic subset of smoking-associated non-small cell lung carcinomas sensitive to MEK inhibition. *Clin. Cancer Res.* **2018**, *24*, 1436–1447. [[CrossRef](#)] [[PubMed](#)]
18. Sampson, E.M.; Haque, Z.K.; Ku, M.C.; Tevosian, S.G.; Albanese, C.; Pestell, R.G.; Paulson, K.E.; Yee, A.S. Negative regulation of the Wnt-β-catenin pathway by the transcriptional repressor HBP1. *EMBO J.* **2001**, *20*, 4500–4511. [[CrossRef](#)]
19. Kim, J.-W.; Lee, H.S.; Nam, K.H.; Ahn, S.; Kim, J.W.; Ahn, S.-H.; Park, D.J.; Kim, H.-H.; Lee, K.-W. PIK3CA mutations are associated with increased tumor aggressiveness and Akt activation in gastric cancer. *Oncotarget* **2017**, *8*, 90948–90958. [[CrossRef](#)]
20. Abubaker, J.; Bavi, P.P.; Al-Harbi, S.; Siraj, A.K.; Al-Dayel, F.; Uddin, S.; Al-Kuraya, K. PIK3CA mutations are mutually exclusive with PTEN loss in diffuse large B-cell lymphoma. *Leukemia* **2007**, *21*, 2368–2370. [[CrossRef](#)]
21. Wang, J.; Yu, P.; Lin, T.; Konigsberg, W.; Steitz, T. Crystal structures of an NH2-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions. *Biochemistry* **1996**, *35*, 8110–8119. [[CrossRef](#)] [[PubMed](#)]

22. Yingling, J.M.; Das, P.; Savage, C.; Zhang, M.; Padgett, R.W.; Wang, X.F. Mammalian dwarfins are phosphorylated in response to transforming growth factor beta and are implicated in control of cell growth. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 8940–8944. [[CrossRef](#)] [[PubMed](#)]
23. Slenter, D.N.; Kutmon, M.; Hanspers, K.; Riutta, A.; Windsor, J.; Nunes, N.; Mélius, J.; Cirillo, E.; Coort, S.L.; Digles, D.; et al. WikiPathways: A multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* **2018**, *46*, D661–D667. [[CrossRef](#)] [[PubMed](#)]
24. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)]
25. Yu, C.; Hong, H.; Zhang, S.; Zong, Y.; Ma, J.; Lu, A.; Sun, J.; Zheng, M. Identification of key genes and pathways involved in microsatellite instability in colorectal cancer. *Mol. Med. Rep.* **2019**, *19*, 2065–2076. [[CrossRef](#)] [[PubMed](#)]
26. Moller, P.; Seppala, T.; Bernstein, I.; Holinski-Feder, E.; Sala, P.; Evans, D.G.; Lindblom, A.; Macrae, F.; Blanco, I.; Sijmons, R.; et al. Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: First report from the prospective Lynch syndrome database. *Gut* **2017**, *66*, 464–472. [[CrossRef](#)]
27. Binder, H.; Hopp, L.; Schweiger, M.R.; Hoffmann, S.; Juhling, F.; Kerick, M.; Timmermann, B.; Siebert, S.; Grimm, C.; Nersisyan, L.; et al. Genomic and transcriptomic heterogeneity of colorectal tumours arising in Lynch syndrome. *J. Pathol.* **2017**, *243*, 242–254. [[CrossRef](#)] [[PubMed](#)]
28. Hwang, J.; Marshall, J.; Salem, M.; Grothey, A.; Goldberg, R.; Xiu, J.; Korn, W.; Shields, A.; Philip, P.; Lenz, H. O-025 Association between tumor mutation burden (TMB) and MLH1, PMS2, MSH2, and MSH6 alterations in 395 microsatellite instability-high (MSI-High) gastrointestinal (GI) tumors. *Ann. Oncol.* **2018**, *29*, mdy149. [[CrossRef](#)]
29. Li, M.; Hruban, R.H.; Velculescu, V.E.; Bigner, D.; Goggins, M.; Kinzler, K.W.; Schmidt, M.K.; Zhang, X.; Iacobuzio-Donahue, C.A.; Maitra, A.; et al. Somatic mutations in the chromatin remodeling gene ARID1A occur in several tumor types. *Hum. Mutat.* **2011**, *33*, 100–103.
30. Allo, G.; Bernardini, M.Q.; Wu, R.-C.; Shih, I.-M.; Kalloger, S.; Pollett, A.; Gilks, C.B.; Clarke, B.A. ARID1A loss correlates with mismatch repair deficiency and intact p53 expression in high-grade endometrial carcinomas. *Mod. Pathol.* **2014**, *27*, 255–261. [[CrossRef](#)]
31. Hänninen, U.A.; Gylfe, A.E.; Renkonen-Sinisalo, L.; Böhm, J.; Taipale, M.; Järvinen, H.; Kondelin, J.; Katainen, R.; Mecklin, J.-P.; Tanskanen, T.; et al. Exome sequencing reveals frequent inactivating mutations in ARID1A, ARID1B, ARID2 and ARID4A in microsatellite unstable colorectal cancer. *Int. J. Cancer* **2014**, *135*, 611–623.
32. Bosse, T.; ter Haar, N.T.; Seeber, L.M.; v Diest, P.J.; Hes, F.J.; Vasen, H.F.A.; Nout, R.A.; Creutzberg, C.L.; Morreau, H.; Smit, V.T. Loss of ARID1A expression and its relationship with PI3K-Akt pathway alterations, TP53 and microsatellite instability in endometrial cancer. *Mod. Pathol.* **2013**, *26*, 1525–1535. [[CrossRef](#)] [[PubMed](#)]
33. Caumanns, J.J.; Wisman, G.B.A.; Berns, K.; van der Zee, A.G.J.; de Jong, S. ARID1A mutant ovarian clear cell carcinoma: A clear target for synthetic lethal strategies. *Biochim. Biophys. Acta-Rev. Cancer* **2018**, *1870*, 176–184. [[CrossRef](#)] [[PubMed](#)]
34. Rizvi, N.A.; Hellmann, M.D.; Snyder, A.; Kvistborg, P.; Makarov, V.; Havel, J.J.; Lee, W.; Yuan, J.; Wong, P.; Ho, T.S.; et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **2016**, *348*, 124–129. [[CrossRef](#)]
35. Shlien, A.; Campbell, B.B.; De Borja, R.; Alexandrov, L.B.; Merico, D.; Wedge, D.; Van Loo, P.; Tarpey, P.S.; Coupland, P.; Behjati, S.; et al. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat. Genet.* **2015**, *47*, 257–262. [[CrossRef](#)]
36. Ripperger, T.; Schlegelberger, B. Acute lymphoblastic leukemia and lymphoma in the context of constitutional mismatch repair deficiency syndrome. *Eur. J. Med. Genet.* **2016**, *59*, 133–142. [[CrossRef](#)] [[PubMed](#)]
37. Cuceu, C.; Colicchio, B.; Jeandidier, E.; Junker, S.; Plassa, F.; Shim, G.; Mika, J.; Frenzel, M.; Jawhari, M.A.L.; Hempel, W.M.; et al. Independent mechanisms lead to genomic instability in Hodgkin lymphoma: Microsatellite or chromosomal instability. *Cancers* **2018**, *10*, 233. [[CrossRef](#)]

38. Rigter, L.S.; Snaebjornsson, P.; Rosenberg, E.H.; Atmodimedjo, P.N.; Aleman, B.M.; Ten Hoeve, J.; Geurts-Giele, W.R.; van Ravesteyn, T.W.; Hoeksel, J.; Meijer, G.A.; et al. Double somatic mutations in mismatch repair genes are frequent in colorectal cancer after Hodgkin's lymphoma treatment. *Gut* **2018**, *67*, 447–455. [[CrossRef](#)]
39. Rossi, D.; Brusca, A.; La Cava, P.; Galimberti, S.; Ciabatti, E.; Luminari, S.; Rigacci, L.; Tucci, A.; Pulsoni, A.; Bertoldero, G.; et al. The genotype of MLH1 identifies a subgroup of follicular lymphoma patients who do not benefit from doxorubicin: FIL-FOLL study. *Haematologica* **2015**, *100*, 517–524. [[CrossRef](#)]
40. Jiang, H.; Lei, R.; Ding, S.W.; Zhu, S. Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinform.* **2014**, *15*, 182. [[CrossRef](#)]
41. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
42. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
43. 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](#)] [[PubMed](#)]
44. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M.; Zhou, S.; et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **2012**, *6*, 80–92. [[CrossRef](#)] [[PubMed](#)]
45. Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, D. The human genome browser at UCSC. *Genome Res.* **2002**, *12*, 996–1006. [[CrossRef](#)]
46. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860. [[CrossRef](#)] [[PubMed](#)]
47. Wickham, H. *ggplot2*; Springer: New York, NY, USA, 2009; ISBN 978-0-387-98140-6.
48. Chen, H.; Boutros, P.C. VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinform.* **2011**, *12*, 35. [[CrossRef](#)]
49. Gu, Z.; Gu, L.; Eils, R.; Schlesner, M.; Brors, B. *circlize* implements and enhances circular visualization in R. *Bioinformatics* **2014**, *30*, 2811–2812. [[CrossRef](#)]
50. Miller, C.A.; Qiao, Y.; DiSera, T.; D'Astous, B.; Marth, G.T. Bam.iobio: A web-based, real-time, sequence alignment file inspector. *Nat. Methods* **2014**, *11*, 1189. [[CrossRef](#)]
51. Jay, J.J.; Brouwer, C. Lollipops in the Clinic: Information Dense Mutation Plots for Precision Medicine. *PLoS ONE* **2016**, *11*, e0160519. [[CrossRef](#)]
52. Snyder, A.; Makarov, V.; Merghoub, T.; Yuan, J.; Zaretsky, J.M.; Desrichard, A.; Walsh, L.A.; Postow, M.A.; Wong, P.; Ho, T.S.; et al. Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *N. Engl. J. Med.* **2014**, *371*, 2189–2199. [[CrossRef](#)] [[PubMed](#)]
53. Le, D.T.; Uram, J.N.; Wang, H.; Bartlett, B.R.; Kemberling, H.; Eyring, A.D.; Skora, A.D.; Luber, B.S.; Azad, N.S.; Laheru, D.; et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.* **2015**, *372*, 2509–2520. [[CrossRef](#)] [[PubMed](#)]
54. Van Allen, E.M.; Miao, D.; Schilling, B.; Shukla, S.A.; Blank, C.; Zimmer, L.; Sucker, A.; Hillen, U.; Foppen, M.H.G.; Goldinger, S.M.; et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **2015**, *350*, 207–211. [[CrossRef](#)] [[PubMed](#)]
55. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [[CrossRef](#)] [[PubMed](#)]

