

1 **Unraveling the heterogeneous mutational signature of spontaneously developing tumors**
2 **in MLH1^{-/-} mice**

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

6
7 ¹ Rostock University Medical Center, Institute for Biostatistics and Informatics in Medicine
8 and Ageing Research (IBIMA), Rostock, Germany

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

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

12 ⁴ Department of Internal Medicine, Medical Clinic III - Hematology, Oncology, Palliative
13 Medicine; Rostock University Medical Center, University of Rostock, 18057 Rostock,
14 Germany

15

16 Running title: Mutations in MLH1^{-/-} tumors

17 Corresponding author:

18 Claudia Maletzki, PhD

19 Department of Internal Medicine, Medical Clinic III - Hematology, Oncology, Palliative
20 Medicine; Rostock University Medical Center, University of Rostock, 18057 Rostock
21 Germany

22 Ernst-Heydemann-Straße 6, D-18057 Rostock, Germany

23 claudia.maletzki@med.uni-rostock.de

24 Telephone: ++49 381 494 5764

25 Fax: ++49 381 494 5898

Mutations in MLH1^{-/-} tumors

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

28

29 Abbreviations: CMMR-D – constitutional mismatch repair deficiency, cMS – coding
30 microsatellite, MB – Megabase, MMR – mismatch repair, MMR-D – mismatch repair
31 deficiency, TMB – tumor mutational burden, GIT – gastrointestinal tumor, SNV – single
32 nucleotide variant, WES – whole exome sequencing

33

34 Keywords: Tumor mutational burden, exclusive/ shared mutations, predicted tumor antigens

Mutations in MLH1^{-/-} tumors

35 **Abstract**

36 MLH1 knock out mice represent a preclinical model that resembles features of the human
37 counterpart. As these mice develop mismatch repair deficient (MMR-D) neoplasias in a
38 sequential twin-peaked manner (first lymphomas, then gastrointestinal tumors) we aimed at
39 identification of the underlying molecular mechanisms. Using whole-exome sequencing, we
40 focused on (I) shared and (II) mutually exclusive mutations and described the processes of
41 ongoing mutational events in tumor-derived cultures.

42 A heterogeneous genetic landscape was found, with few mutations shared among different
43 neoplasias (*ARID1A* and *IDH2*). Mutations in tumor suppressor genes *SMAD4* and *POLE*
44 were mutually exclusive in lymphomas, most likely contributing to a more aggressive *in vivo*
45 phenotype. Comparing the mutational profile of selected primary tumors and their
46 corresponding cell line upon *in vitro* culture revealed continuous increased numbers of
47 somatic gene mutations. The same was true for coding microsatellite mutations in selected
48 MMR-D target genes, showing a gradual increase during *in vitro* passage. With respect to this
49 latter type of mutations, partial overlap was detectable, yet recognizing shared vaccination
50 antigens. The two most promising candidates are *AKT3*, a RAC-gamma serine/threonine-
51 protein kinase with relevance in maintenance of cellular homeostasis and the endonuclease
52 *ERCC5* (Excision Repair 5), involved in DNA excision repair.

53 Novel results of a comparison between spontaneously developing lymphomas and
54 gastrointestinal tumors as models for MMR-D driven tumorigenesis are reported. In addition
55 to identification of *ARID1A* as a potentially causative mutation hotspot, our comprehensive
56 characterization of the mutational signature is a starting point for immune-based approaches
57 to therapy.

58

59

Mutations in MLH1^{-/-} tumors

60 **Author Summary**

61 This study describes the mutational spectrum of MLH1^{-/-}-associated tumors, spontaneously
62 developing in mice. While these tumors arise at the bottom of the same germline mutation,
63 the clinical presentations as well as resulting molecular alterations are heterogeneous, and
64 thus likely being directly linked. Highly aggressive lymphomas, developing early in life are
65 ultra-hypermutated and harbor mutations in tumor suppressor genes *SMAD4* and *POLE*.
66 Gastrointestinal tumors develop later in life and show different mutations. By performing in-
67 depth whole exome sequencing analysis, we here identified for the first time a common
68 mutational hotspot. *ARID1A* constitutes a potentially causative mutation, shared among
69 different MLH1^{-/-}-associated tumors and thus irrespective of the origin. Additional interesting
70 and identified candidate genes include AKT3, a RAC-gamma serine/threonine-protein kinase
71 and the endonuclease ERCC5. Both genes are bona fide tumor suppressors with significant
72 relevance in DNA excision repair and maintenance of cellular homeostasis. This finding is of
73 particular relevance for subsequent therapeutic and - even more important - prophylactic
74 vaccination approaches aiming at entity-overlapping treatment of MLH1^{-/-}-related tumors.

75

76 **1. Introduction**

77 Comprehensive genomic sequencing is an increasingly common practice in oncological
78 precision medicine. The technological improvements in analyzing diseases at the genomic,
79 transcriptomic and epigenetic levels allow for the identification of characteristic genetic
80 changes. To detect disease-causing variants (so-called driver mutations) and discover
81 therapeutic target genes, whole exome sequencing (WES) is a highly fruitful sequencing
82 method (1). By applying this method, the coding region of the genome is captured and
83 sequenced at high depth. The interrogation of the exome in clinical diagnosis raises
84 challenges of uncovering mutational profiles of heterogeneous tumors and may allow the
85 development of new immunotherapeutic approaches (2,3).

Mutations in *MLH1*^{-/-} tumors

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

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

107 *MLH1*^{-/-} knockout mice constitute an ideal preclinical model for hereditary tumor syndromes.
108 With a virtually 100 % penetrance, these mice develop a variety of malignancies, including
109 highly aggressive lymphomas in different organ sites, gastrointestinal tumors, and malignant
110 skin lesions (9,10). The sequential appearance of these neoplasias, often in a mutually

Mutations in MLH1^{-/-} tumors

111 exclusive fashion ("either-or" principle in terms of tumor type) raises the question of how the
112 biological and/or molecular mechanisms of these neoplasias relate to each other. To gain
113 insights into the underlying molecular alterations that affect tumor formation and thus act as
114 drivers, we herein analyzed the number and type of somatic mutations arising at the bottom of
115 a germline MMR-D. Exclusive as well as shared alterations were identified, explaining the
116 heterogeneous clinical presentation.

Mutations in MLH1^{-/-} tumors

117 2. Results

118 2.1. *In vivo* and *ex vivo* data

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

126 For our comprehensive analysis, we included a set of four tumor samples, mainly covering the
127 spectrum observed in MLH1^{-/-} mice (Table 1). Since all MLH1^{-/-}-associated tumors harbor
128 MHC class I on their surface, which is preserved in the resulting *in vitro* cell culture as well
129 as in the allograft (9), phenotyping focused on immune evasion markers, such as PD-L1,
130 CTLA-4, LAG-3, and TIM-3. As anticipated, lymphomas expressed all markers in high
131 abundance (Figure 1B). By contrast, PD-L1 and CTLA-4 surface expression on GIT was
132 heterogeneous and found on only 20-40 % of cells, respectively (Figure 1B; referring to P11
133 and P12, respectively). LAG-3 and TIM-3 expression was confined to tumor-infiltrating
134 lymphocytes and thus below 5 % (Figure 1B; referring to P14 and P15, respectively and (11)).
135 Generally, lymphocytic infiltrates in MLH1^{-/-}-associated GIT showed an equal distribution
136 between CD3⁺CD4⁺ and CD3⁺CD8⁺ T cells (Figure 1C, right bars with dashed lines).
137 However, marked differences were found between the two GIT cases #7450 and #328, nicely
138 reflecting the heterogeneity even among GIT with similar genetic background. As can be
139 taken from Figure 1C, the value of both cell types in #7450 (dotted plot) exceeds the mean
140 value, while in #328 (checkered plot) infiltrating cell numbers are far lower than on average.

Mutations in MLH1^{-/-} tumors

142 2.2. Mutational landscape of MLH1^{-/-} tumors

143 The different tumor spectra detectable in MLH1^{-/-} mice prompted us to perform a whole
144 exome sequencing analysis. This experiment confirmed our observations of high TMB. For
145 the confirmatory analysis, we initially considered all point mutations. TMB ranged from 39
146 mutations/Megabase (MB) (in GIT #7450) to 943 mutations/MB (in GIT #328). Breaking this
147 down to non-synonymous substitutions, i.e., missense and nonsense mutations, high
148 mutational burden was, by the definition of ref(12) (> 10 mut/MB), preserved, with 3/4
149 samples even being ultra-hypermutated (> 100 mut/MB) (Figure 2A).

150 Most mutations in cancer genomes, however, are ‘passengers’ and do not bear strong imprints
151 of selection. Apart from this, the status of hypermutation complicates the identification of
152 driver mutations simply due to the sheer abundance of somatic variants (13). As MLH1^{-/-}-
153 associated lymphomas showed a more aggressive *in vivo* growth behavior and develop
154 considerably earlier than GIT, we expected a different molecular make-up of the lymphomas
155 and possibly also a higher mutational burden. While the former was actually true, the
156 mutational load was not generally higher. Indeed, the lymphomas’ mutation rates were
157 highest, but the two GIT cases #7450 and #328 itself differed significantly from each other
158 (Figure 2). The #328s’ mutational load was mostly comparable with that of the splenic and
159 skin lymphomas (Figure 2). The very early formation of this GIT before the mean age of
160 onset (~33 weeks for GIT), may explain our finding.

161 Missense mutations were more frequent than nonsense mutations, and base changes were
162 mainly due to transitions (C>T; A>G) (Figure 2A). High levels of transversion are associated
163 with PD-L1 abundance that functions as a biomarker for immune-checkpoint inhibition. In a
164 previous study, PD-L1 abundance was analyzed on MLH1^{-/-}-tumors (14). In line with our
165 recent findings on high percentage numbers of transitions instead of transversions, we also

Mutations in MLH1^{-/-} tumors

166 detected only low-moderate PD-L1 expression (<30 %), and if detectable, this was mainly
167 confined to infiltrating immune cells, see ref. (14).

168

169 **2.3. Mutation types and shared mutations**

170 Comparative whole exome sequencing of the MLH1^{-/-} GIT #7450 and #328 with the
171 lymphoma revealed the mutanome of these samples (Figures 3 and 4). Considering the
172 number of mutations per type, marked differences were found in the #328 compared to the
173 #7450 GIT and both lymphomas (Figure 3A). As anticipated, the number of silent mutations
174 was high whereas the nonsense mutation rate was low in all four tumor samples. Focusing on
175 the non-synonymous mutations demonstrated similar amounts of these for the #328 GIT and
176 the splenic lymphoma, but slightly increased numbers compared to the skin lymphoma. In
177 fact, #328 from a male mouse had the highest mutation rate in all mutation types, compared to
178 the female counterpart GIT and to the other cancers. Furthermore, lymphomas and GIT
179 shared 640 mutations (Figure 3B), although a high amount of single nucleotide variations
180 (SNV) for each case was exclusive. Interestingly, the lymphomas shared a higher amount of
181 SNV with 684 vs. only 10 in the GIT.

182 In dissecting the mutational landscape (Figure 3C) we initially focused on shared mutations
183 among all cancers most likely functioning as drivers for MLH1-associated tumorigenesis.
184 With this analysis, a surprisingly low number of shared mutations were detectable. These
185 included mutations in the following genes: *AT-rich interaction domain 1A (ARID1A)* and
186 *isocitrate dehydrogenase 2 (IDH2)*. *ARID1A*, located on chromosome 4, encodes a protein of
187 the SWI/SNF family, contributing to the large ATP-dependent chromatin-remodeling
188 complex. It is an epigenetic modifier that functions as a tumor suppressor; gene mutations
189 have been reported in several malignancies (15). *IDH2* is different, usually being amplified

Mutations in MLH1^{-/-} tumors

190 and/or overexpressed in tumors. Mutations are associated with poor clinical outcome, well
191 reflecting the clinical presentation of biallelic MMR-D-driven malignancies.

192

193 **2.4. Mutational frequencies and types of alterations for selected gene sets based on prior 194 knowledge in mouse and human**

195 Next, we focused on selected genes with known relevance for tumor initiation and
196 progression as well as tumor suppressive function (Figure 4A). In this analysis and in
197 accordance with the human counterpart, MLH1^{-/-}-associated tumors harbor additional
198 mutations in *PIK3CA*, *BRAF* and/or *KRAS* and *ERBB3* (16). Therefore, we created an
199 oncoprint of the non-synonymous alterations in this selected gene set (Figure 4A). A missense
200 mutation was found for *PIK3CA*, *BRAF*, *KRAS*, and *ERBB3*. Events associated with these
201 genes show alterations distributed in a nearly mutually exclusive way across all samples.
202 Surprising are the alterations in the *NFI* gene, a negative regulator of the *RAS* signal
203 transduction pathway that is modified in #328 (primary and cell line). These *NFI* mutations
204 were shown to be concomitant with mutations in the oncogenes *BRAF* or *KRAS* (17,18). Here,
205 the same samples that had *NFI* mutations displayed *BRAF* mutations as well, but only the
206 skin lymphoma and the GIT (#328) had additional *KRAS* mutations. In the case of *APC*, a
207 negative regulator of the *Wnt* signaling pathway (19), we observed for the batch #7450 that
208 the cell line acquired additional missense mutations that had not been detected in the primary.
209 The *PI3K* pathway is affected by *PIK3CA*. Further interactions of *PIK3CA* are with *AKT* and
210 *mTOR* pathways, mutated in one GIT (#328) and both lymphomas. In gastric cancer, *PIK3CA*
211 mutations are associated with a higher aggressiveness (20), while for the lymphomas a strong
212 correlation with *PTEN* mutations is described (21). Both genes were found to be mutated in
213 the GIT (#328) and in splenic lymphoma, respectively. The same was also true for *EGFR*
214 signaling members *ERBB2* and *ERBB3*.

Mutations in MLH1^{-/-} tumors

215 *SMAD4* and *POLE* were exclusively detectable in lymphomas (Figure 4A). Of note, the
216 detected somatic *POLE* mutation was located within the exonuclease domain of this gene,
217 executing a proofreading function to decrease the mutation rate during DNA replication. In
218 conjunction with germline MMR-D, *POLE* mutations strongly increase the number of gene
219 mutations in affected tumor cells. Additionally, *SMAD4* acts as a tumor suppressor and
220 inhibits transforming growth factor- β -mediated signaling. Mutations are associated with poor
221 outcome and high metastatic potential. With respect to MLH1^{-/-} tumors, *SMAD4* mutations
222 were constrained to lymphomas.

223

224 **2.5. Pairwise exclusively SNV in GIT and lymphomas**

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

234 A pairwise comparison of the splenic and skin lymphomas showed that they share the mutated
235 genes *APC*, *SMAD4*, *PTEN*, *ERBB4*, *PIK3CA*, *ARID1A*, *BRAF*, *MET*, *IDH2*, *ERBB3*, *NF1*,
236 and *CDC27*. However, exclusive changes were also observed. Each of the lymphomas has
237 SNV in six exclusive genes. Therefore, we analyzed in detail the processes and functions the
238 exclusively mutated genes are involved in. The splenic lymphoma exhibits SNV in *MYO1B*,
239 *CTNNBL1*, *ERBB2*, *SOX9*, *MSH3*, and *RBFOX1*, involved in cell migration, the spliceosome,

Mutations in MLH1^{-/-} tumors

240 cellular responses, chondrocytes differentiation, and MMR. Exclusive SNV for the skin
241 lymphoma were found in the following genes: *IDHI*, *CASP8*, *ACVR2A*, *KRAS*, *IGF2*, and
242 *POLD1*. All of these genes are part of the electron acceptor, cell apoptosis, oncogene, growth,
243 replication, and repair.

244

245 **2.6. Prevalence and hotspot regions of GIT and lymphomas in *ARID1A*, *POLE* and** 246 ***SMAD4***

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

256 Figure 4B illustrates the mutational hotspots of the mutually exclusive lymphoma-associated
257 genes *POLE* and *SMAD4*. In *POLE*, we found the shared amino acid change D262Y for the
258 splenic and skin lymphoma. Furthermore, the splenic lymphoma has a nonsense mutation,
259 E265*, in the DNA polymerase B exonuclease domain that adopts a ribonuclease H type fold
260 (22). Interestingly, the lymphomas show an amino acid change, V346I, in the MH2 domain
261 implicated in the control of cell growth (23) of *SMAD4*. Additional loci are in the putative
262 disordered regions, T211A and A229G.

263

264 **2.7. Somatic mutation distribution and nucleotide changes in chromosome X**

Mutations in MLH1^{-/-} tumors

265 In addition to the TMB calculations, we analyzed the mutational load per chromosome. This
266 mutational load was normalized based on the size of the respective chromosome, and the
267 mutation rate per MB plotted for all four entities (supplementary Figure 1). Therefore, this
268 presentation of the chromosomal distribution of SNV shows that for all entities, chromosome
269 11 accumulated the highest numbers of somatic mutations per MB. Given the fact of
270 differences in gender and a possible hyper-mutation of chromosome X, we investigated
271 chromosome X further for the nucleotide changes in pyrimidines C and T (supplementary
272 Figure 2). C is also responsible for the deamination of the X-chromosome. The coloring of the
273 bars represent the proportions of the six possible nucleotide changes of the pyrimidines (C >
274 A, C > G, C > T, T > A, T > C, and T > G) for chromosome X. Comparing the GIT, we see
275 that T>C is less frequent in the #328 than in the #7450. In turn, the changes C>A and T>G are
276 more frequent. The lymphomas show only minor differences in the nucleotide changes for
277 C>T and T>C, whereby the skin lymphoma had slightly lower changes in this latter than the
278 splenic lymphoma.

279

280 **2.8. Ongoing mutations in MLH1^{-/-} tumors**

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

288 Finally, the cMS mutational profile was examined on a panel of primary tumors, i.e.
289 lymphomas and GIT (N = 10 cases each). With this analysis, mutations were detectable in

Mutations in MLH1^{-/-} tumors

290 15/26 markers, with 5 markers being shared among both cancer types (Figure 5C). Of note
291 and of particular relevance, only two genes had the same mutation in their cMS marker of an
292 MSI target gene, thus being in conjunction with our WES data of minor entity-overlapping
293 mutational events. These are *AKT3*, a RAC-gamma serine/threonine-protein kinase with
294 relevance in the maintenance of cellular homeostasis (balancing survival and apoptosis) and
295 the endonuclease *ERCC5* (Excision Repair 5), involved in DNA excision repair. Mutations in
296 this gene increase susceptibility for skin cancer development. Given the biological function of
297 these two genes, their applicability as a target structure is an issue still to be determined.

298

Mutations in MLH1^{-/-} tumors

299 3. Discussion

300 Due to the biallelic germline MMR-D and the accordingly complex tumor spectrum, MLH1^{-/-}
301 mice are frequently considered a better model for CMMR-D, instead of Lynch Syndrome. In
302 men, *MLH1* has the highest frequency among monoallelic MMR mutation carriers and also
303 the highest cumulative risk to develop any cancer (24). Despite some differences, MLH1^{-/-}
304 mice combine several key features of both syndromes. The recently detected heterogeneity of
305 Lynch-associated cancers, being highly immunogenic or not (25), additionally argues in favor
306 of using these mice as a model for Lynch Syndrome as well. All malignant human CMMR-D
307 cancers are ultra-hypermutated. However, this is so far not known for tumors arising in the
308 context of Lynch syndrome. Apart from this, there is growing evidence of individual
309 differences between patients, likely being dependent on the underlying MMR defect (26). In
310 line with this broad heterogeneity among MMR-D-related malignancies, we here show that
311 MLH1^{-/-} tumors (I) are either hyper- or ultra-hypermutated, (II) have an entirely individual
312 mutational profile and (III) harbor different mutational hotspots in affected genes.

313 Of note, along with the different sites of manifestation, a gender-specific prevalence is
314 evident in mice, where thymic lymphomas primarily develop in females, and 80 % of GIT
315 manifest in males. Apart from hormonal-differences, X-chromosome specific alterations may
316 provide a rationale for the different prevalence. The X-chromosome of MLH1^{-/-} mice is
317 indeed highly mutated. However, we could not find evidence for a hypermutated X-
318 chromosome. We could demonstrate that the mutations/MB on the X-chromosome are almost
319 twice compared to the mean mutation rate of the autosomes. Hence, this suggests that the
320 accumulated mutations are coming from the paternal, active X chromosome for the #328
321 primary tumor and the inactive X chromosome for the #7450 primary tumor. In consequence,
322 this would mean for the lymphomas, that the inheritance of the highly accumulated mutations
323 of chromosome X came from the maternal, active X chromosome. We hypothesize that the
324 high mutation rate on the X chromosome is a cancer-specific feature for the lymphomas and

Mutations in MLH1^{-/-} tumors

325 represents a subgroup in the GIT primary tumors. In addition to the gender, several biological
326 characteristics argue in favor of this hypothesis, leaving technical artifacts as the sole or
327 dominant explanation for the different mutational load very unlikely: (I) the age of onset and
328 (II) the number of tumor-infiltrating lymphocytes.

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

355 By performing in-depth sequencing on lymphomas and GIT, most altered genes involved
356 *APC*, *PTEN*, *PIK3CA*, *Her2*-related genes, *BRAF* and/or *KRAS*. However, it is often difficult
357 to determine, which mutations arose first or whether their order is essential in driving
358 tumorigenesis. While MLH1^{-/-} mice are inbred and all descendants harbor the same germline
359 mutation, we expected a comparable mutational profile among cancers. This was,
360 nevertheless, not the case with a comparably low number of shared mutations among
361 analyzed cancer types. Indeed, entity-overlapping mutations were exclusively detectable in
362 *ARID1A* and *IDH2*. In men, *IDH2* mutations are linked to brain tumors and acute myeloid
363 leukemia, while *IDH1* mutations may drive tumor progression. IDH-mutated tumors are
364 clinically and genetically distinct from their wildtype counterpart. *ARID1A* is a bona fide
365 tumor suppressor and a frequent mutational hotspot in many cancers, including those of the
366 breast, gastrointestinal tract and ovary (27). *ARID1A* functions in MMR regulation, as its loss
367 is associated with MSI. It is supposed to be a driver gene, occurring secondary to MMR-D in
368 gastric cancers, representing an alternative oncogenic pathway to p53 alteration (28). Neither
369 of the tumors examined here showed *TP53* mutations, confirming this assumption. *ARID1A*
370 mutations in MLH1^{-/-}-associated tumors are distributed along the entire length of the gene,

Mutations in MLH1^{-/-} tumors

351 thus distinguishing it from typical MSI target genes, a finding also described in the human
352 counterpart (29).

353 Loss of *ARID1A* shortens time to cancer-specific mortality in men, where mutations are
354 associated with both sporadic and hereditary MSI cancers (30). Though it is still unknown
355 whether CMMR-D cases harbor *ARID1A* mutations as well, this observation fits with our
356 finding on MLH1^{-/-} mice having a significantly shortened life expectancy.

357 Comparable analysis done by others on human cancers revealed ten overlapping mutations,
358 all of which are InDels involving G7 or C6 repeats (31). These mutations are potentially
359 caused by MSI, and although not analyzed in detail here, we also observed *ARID1A* mutations
360 in our own collection of MMR-D-associated clinical CRC cases (22 % of cases harbor
361 mutations in the G7 repeat and have additionally wildtype *TP53*). *ARID1A* is thus a likely
362 target for personalized medicine in cancer treatment. *ARID1A*-deficiency-based synthetic
363 lethal, targeted drugs are under clinical investigation to suppress cell growth and promote
364 apoptosis (31) (see also clinicaltrials.gov). Whether this provides a real opportunity for
365 *ARID1A*-deficient MMR-D-related cancer will need to be tested prospectively.

366 Another interesting finding of this study was the high prevalence of transitions instead of
367 transversions among gene mutations. Although this is commonly seen in non-smoking
368 associated cancers, we wish to stress the close link between transitions (mainly C-to-T) and
369 MSI as well as *POLE/POLD1* mutations in men (32). While these alterations are detectable in
370 MLH1^{-/-}-associated tumors similar mechanisms can be assumed. As for the latter, the
371 relevance of somatic *POLE/POLD1* mutations in the context of MMR-D has only begun to be
372 enlightened. In clinical cases of CMMR-D, most tumors harbor *POLE* mutations affecting the
373 exonuclease domain or domains important to the intrinsic proofreading activity (33). These
374 cells rapidly and catastrophically accumulate point mutations, yielding the ultra-hypermutated
375 phenotype – quite similar to our observations. Cells mutate continuously, raising the question

Mutations in MLH1^{-/-} tumors

376 of how many mutations can be tolerated until a maximum threshold is reached that creates the
377 Achilles' heel.

378 All mutations described so far may have prognostic, predictive and surveillance potential.
379 Besides, a plethora of low-frequency variants whose functional consequences and clinical
380 actionability are unknown were found as well. We thus conclude that for MLH1^{-/-}-driven
381 oncogenesis, many mutations are mere passengers with no functional consequence.

382 A point worth mentioning is the fact that MMR-D associated lymphomas in men are rarely
383 documented in the literature and most of them develop in patients suffering from CMMR-D.
384 In here, however, hematological malignancies, are common and even in this small group of
385 patients, mostly associated with founder *MLH1* and/or *MSH2* mutations (34). A recent study
386 described the occurrence of MSI/MMR-D in *de novo* Hodgkin lymphoma, but at a
387 significantly higher frequency in Hodgkin's lymphoma survivors developing therapy-related
388 colorectal cancers (35). *MLH1* promoter hypermethylation, but also double somatic mutations
389 are the leading causes of MMR-D (36). Highlighting the relevance of *MLH1*, one study also
390 demonstrates the failure of a specific immune-chemotherapy regimen in follicular lymphoma
391 patients showing *MLH1* gene polymorphisms (37). In addition to germline MMR-D, the
392 frequency of MMR-D arising secondary to anticancer treatment, is thus supposedly higher
393 than previously predicted, considering this a clinically relevant finding. This may finally
394 witness a resurgence of interest to apply and develop alternative therapeutic regimens.

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

Mutations in MLH1^{-/-} tumors

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

Mutations in MLH1^{-/-} tumors

404 4. Materials and Methods

405 4.1. *In vivo mouse model & sample acquisition*

406 Homozygous MLH1-deficient mice were bred in the animal facilities (University of Rostock)
407 under specified pathogen-free conditions. Trials were performed in accordance with the
408 German legislation on protection of animals and the Guide for the Care and Use of
409 Laboratory Animals (Institute of Laboratory Animal Resources, National Research Council;
410 NIH Guide, vol.25, no.28, 1996; approval number: LALLF M-V/TSD/7221.3- 1.1-053/12 and
411 026/17, respectively). Tumor samples for sequencing analysis were obtained from four
412 individual MLH1^{-/-} mice upon spontaneous development. Two mice suffered from a
413 lymphoma (spleen and skin), and the remaining two mice had gastrointestinal tumors both
414 located in the duodenum. All tumors were at an advanced stage at resection. Samples were
415 snap frozen in liquid nitrogen and stored at -80°C. Isolation of gDNA was done by GATC
416 (samples: 7450 primary GIT; A7450 P15 cell line) and CEGAT (samples: 328 primary; 328
417 cell line; 1351 lymphoma (spleen); 1444 lymphoma (skin)), respectively, prior to the
418 sequencing analyses.

419

420 4.2. *Whole exome sequencing (WES) analysis.*

421 Data preprocessing: First, the read quality of the raw sequencing data was investigated with
422 FastQC v. 0.11.5 (3) to adjust the adapter trimming. The sequencing adapters were trimmed
423 using Skewer v. 0.2.2 (38), for the reads after demultiplexing with Illumina bcl2fastq v. 2.19.
424 Reads with a Phred quality score ≥ 30 were considered as high quality. The whole-exome
425 reads were mapped to the published mouse genome build ENSEMBL mm10.75 reference
426 using Burroughs Wheeler Aligner (BWA) (39) with the bwa-mem algorithm v. 0.7.2(40).
427 Reads with multiple mappings and the same mapping score have been discarded from further
428 analysis. Additionally, PCR duplicate reads have been removed by using Samtools v. 0.1.18
429 (39) to prevent artificial coverage brought on by the PCR amplification step during library

Mutations in MLH1^{-/-} tumors

430 preparation and to avoid its impacts on identifying false positive mutations. SNV calling was
431 performed using the Varscan v. 2.4.2 (41), and the detected variants were annotated based on
432 their gene context using snpEff v. 4.3 (42). To compare these data with data from our
433 previously published paper (3), the mm9 SNV coordinates were converted to the
434 corresponding mm10 coordinates with hg1liftOver v. 30-10-2018 (43,44). The cell line A7450
435 from the batch was processed likewise (3).

436

437 *4.3. Data visualization*

438 For the MLH1^{-/-} GIT primary tumor, derived cell line, spleen and skin lymphoma, the
439 summarized mutational profiles with the corresponding mutation type were summarized as
440 stacked bar plot with ggplot2 (45), as Venn diagram with VennDiagram (46) as well as
441 whole-genome ideograms with the circlize (47) R package. Based on the mm10 mouse
442 assembly, provided by the circlize package (47), the filtered mutational profiles are presented
443 as ideograms.

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

448 Additional details like TSTV (transition and transversion) ratio, allele frequency spectrum,
449 base changes, and variant quality for each mutational profile were obtained using the
450 vcf.iobio tool (48).

451 Furthermore, mapping the mutations and its statistics on a linear gene product (proteins of
452 interest) was done with a ‘lollipop’ mutation diagram generator (49). Based on the knowledge

Mutations in MLH1^{-/-} tumors

453 from the human MMR-D counterpart and general involvement in tumorigenesis, the genes for
454 further analysis were chosen with high probability to mutate.

455 *4.4. Hypermutation and TMB calculation*

456 We adopted the following definition for calculating the TMB:

$$457 \frac{\# \text{ all non-synonymous SNVs on target}}{\text{covered exome by sequencing kit}} (50-52).$$

458 Simply, we used intersectBed from the BEDTools suite v. 2.26.0(53) for ensuring the SNV
459 being on target regions of the Agilent SureSelect Mouse All Exon Kit mm10 and then
460 summing up the number of somatic mutations per MB of the transcriptome mm10 (48.3MB).

461 We use the definition of hypermutation as follows (32,51,52):

$$462 \text{per MB}_{\text{cancer entity}} = \begin{cases} \text{hypermutated} & \text{if non-synonymous SNVs} > 10 \\ \text{ultra-hypermutated} & \text{if non-synonymous SNVs} > 100 \end{cases}$$

463 where $\text{cancer entity} \in \{\text{GIT, derived cell line, lymphoma}\}$.

464

465 *4.5. Coding microsatellite (cMS) frameshift mutation analysis*

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

469

Mutations in MLH1^{-/-} tumors

470 Acknowledgments

471 The authors gratefully thank Mrs. Ilona Klamfuss for breeding MLH1 mice.

472

473 **Conflict of interest.** “No potential conflict of interest was reported by the authors.”

474 Author' contributions

475 YSG: participated in the outline of the manuscript, developed bioinformatics study analysis,

476 performed sequencing data preprocessing, data analysis/interpretation, wrote the manuscript

477 LW: performed fragment length analysis, participated in *in vivo* data acquisition and data

478 analysis

479 MH: directed the bioinformatics study analysis, contributed to the discussion and

480 interpretation of the results and paper finalization

481 AM: participated in literature review and drafted the manuscript

482 GF: participated in paper finalization and critically revised the manuscript

483 CJ: participated in paper finalization and critically revised the manuscript

484 CM: designed study, the outline of the manuscript, performed *in vivo* experiments, data

485 analysis/ interpretation, wrote the manuscript

486

487 All authors approved the final version of the paper.

488

Mutations in MLH1^{-/-} tumors

489 5. References

- 490 1. Gupta S, Chatterjee S, Mukherjee A, Mutsuddi M. Whole exome sequencing:
491 Uncovering causal genetic variants for ocular diseases. *Experimental Eye Research*.
492 2017.
- 493 2. Roche MI, Berg JS. Incidental Findings with Genomic Testing: Implications for
494 Genetic Counseling Practice. *Curr Genet Med Rep*. 2015;
- 495 3. Maletzki C, Gladbach YS, Hamed M, Fuellen G, Semmler M-L, Stenzel J, et al.
496 Cellular vaccination of MLH1^{-/-} mice – an immunotherapeutic proof of concept
497 study. *Oncoimmunology*. 2018;7(3):e1408748.
- 498 4. Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology*.
499 2010 Jun;138(6):2073-2087.e3.
- 500 5. Ozcan M, Janikovits J, von Knebel Doeberitz M, Kloost M. Complex pattern of
501 immune evasion in MSI colorectal cancer. *Oncoimmunology* [Internet]. 2018;7(7):1–
502 10. Available from: <https://doi.org/10.1080/2162402X.2018.1445453>
- 503 6. Kloost M, Von Knebel Doeberitz M. The immune biology of microsatellite-unstable
504 cancer. *Trends in Cancer* [Internet]. 2016;2(3):121–33. Available from:
505 <http://dx.doi.org/10.1016/j.trecan.2016.02.004>
- 506 7. Bakry D, Aronson M, Durno C, Rimawi H, Farah R, Alharbi QK, et al. Genetic and
507 clinical determinants of constitutional mismatch repair deficiency syndrome: Report
508 from the constitutional mismatch repair deficiency consortium. *Eur J Cancer* [Internet].
509 2014;50(5):987–96. Available from: <http://dx.doi.org/10.1016/j.ejca.2013.12.005>
- 510 8. Wimmer K, Rosenbaum T, Messiaen L. Connections between constitutional mismatch
511 repair deficiency syndrome and neurofibromatosis type 1. *Clin Genet* [Internet].
512 2016;(3):507–19. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27779754>
- 513 9. Maletzki C, Beyrich F, Hühns M, Klar E, Linnebacher M. The mutational profile and
514 infiltration pattern of murine MLH1^{-/-} tumors: concurrences, disparities and cell line

Mutations in MLH1^{-/-} tumors

515 establishment for functional analysis. *Oncotarget*. 2016 Aug;7(33):53583–98.

516 10. Daino K, Ishikawa A, Suga T, Amasaki Y, Kodama Y, Shang Y, et al. Mutational
517 landscape of T-cell lymphoma in mice lacking the DNA mismatch repair gene Mlh1:
518 no synergism with ionizing radiation. *Carcinogenesis*. 2019;40(2):216–24.

519 11. Maletzki C, Gladbach YS, Hamed M, Fuellen G, Semmler M-L, Stenzel J, et al.
520 Cellular vaccination of MLH1^{-/-} mice—an immunotherapeutic proof of concept study.
521 *Oncoimmunology*. 2018;7(3).

522 12. Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, et al.
523 Comprehensive Analysis of Hypermutation in Human Cancer. *Cell*. 2017;171(5):1042-
524 1056.e10.

525 13. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al.
526 Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*.
527 2017;171(5):1029-1041.e21.

528 14. Maletzki C, Wiegele L, Nassar I, Stenzel J, Junghanss C. Chemo-immunotherapy
529 improves long-term survival in a preclinical model of MMR-D-related cancer. *J
530 Immunother Cancer*. 2019;7(1):1–14.

531 15. Ayhan A, Mao T-L, Seckin T, Wu C-H, Guan B, Ogawa H, et al. Loss of ARID1A
532 expression is an early molecular event in tumor progression from ovarian
533 endometriotic cyst to clear cell and endometrioid carcinoma. *Int J Gynecol Cancer*.
534 2012 Oct;22(8):1310–5.

535 16. Fakhri B, Lim KH. Molecular landscape and sub-classification of gastrointestinal
536 cancers: A review of literature. *J Gastrointest Oncol*. 2017;8(3):379–86.

537 17. Zenonos K, Kyprianou K. RAS signaling pathways, mutations and their role in
538 colorectal cancer. *World J Gastrointest Oncol*. 2013 May;5(5):97–101.

539 18. Hayashi T, Desmeules P, Smith RS, Drilon A, Somwar R, Ladanyi M. RASA1 and
540 NF1 are preferentially co-mutated and define a distinct genetic subset of smoking-

Mutations in MLH1^{-/-} tumors

541 associated non–small cell lung carcinomas sensitive to MEK inhibition. *Clin Cancer*
542 *Res.* 2018;

543 19. Sampson EM, Haque ZK, Ku MC, Tevosian SG, Albanese C, Pestell RG, et al.
544 Negative regulation of the Wnt- β -catenin pathway by the transcriptional repressor
545 HBP1. *EMBO J.* 2001;

546 20. Kim J-W, Lee HS, Nam KH, Ahn S, Kim JW, Ahn S-H, et al. PIK3CA mutations are
547 associated with increased tumor aggressiveness and Akt activation in gastric cancer.
548 *Oncotarget.* 2017 Oct;8(53):90948–58.

549 21. Abubaker J, Bavi PP, Al-Harbi S, Siraj AK, Al-Dayel F, Uddin S, et al. PIK3CA
550 mutations are mutually exclusive with PTEN loss in diffuse large B-cell lymphoma.
551 *Leukemia.* 2007 Nov;21(11):2368–70.

552 22. Wang J, Yu P, Lin T, Konigsberg W, Steitz T. Crystal structures of an NH₂-terminal
553 fragment of T4 DNA polymerase and its complexes with single-stranded DNA and
554 with divalent metal ions. *Biochemistry.* 1996 Jun;35(25):8110–9.

555 23. Yingling JM, Das P, Savage C, Zhang M, Padgett RW, Wang XF. Mammalian
556 dwarfins are phosphorylated in response to transforming growth factor beta and are
557 implicated in control of cell growth. *Proc Natl Acad Sci.* 1996 Aug;93(17):8940–4.

558 24. Moller P, Seppala T, Bernstein I, Holinski-Feder E, Sala P, Evans DG, et al. Cancer
559 incidence and survival in Lynch syndrome patients receiving colonoscopic and
560 gynaecological surveillance: first report from the prospective Lynch syndrome
561 database. *Gut.* 2017 Mar;66(3):464–72.

562 25. Binder H, Hopp L, Schweiger MR, Hoffmann S, Juhling F, Kerick M, et al. Genomic
563 and transcriptomic heterogeneity of colorectal tumours arising in Lynch syndrome. *J*
564 *Pathol.* 2017 Oct;243(2):242–54.

565 26. Hwang J, Marshall J, Salem M, Grothey A, Goldberg R, Xiu J, et al. O-025 Association
566 between tumor mutation burden (TMB) and MLH1, PMS2, MSH2, and MSH6

Mutations in MLH1^{-/-} tumors

567 alterations in 395 microsatellite instability-high (MSI-High) gastrointestinal (GI)
568 tumors. Ann Oncol [Internet]. 2018;29(suppl_5). Available from:
569 <https://doi.org/10.1093/annonc/mdy149.024>

570 27. Li M, Hruban RH, Velculescu VE, Bigner D, Goggins M, Kinzler KW, et al. Somatic
571 mutations in the chromatin remodeling gene ARID1A occur in several tumor types.
572 Hum Mutat. 2011;33(1):100–3.

573 28. Allo G, Bernardini MQ, Wu R-C, Shih I-M, Kalloger S, Pollett A, et al. ARID1A loss
574 correlates with mismatch repair deficiency and intact p53 expression in high-grade
575 endometrial carcinomas. Mod Pathol an Off J United States Can Acad Pathol Inc.
576 2014 Feb;27(2):255–61.

577 29. Hänninen UA, Gylfe AE, Renkonen-Sinisalo L, Böhm J, Taipale M, Järvinen H, et al.
578 Exome sequencing reveals frequent inactivating mutations in ARID1A, ARID1B,
579 ARID2 and ARID4A in microsatellite unstable colorectal cancer. Int J Cancer.
580 2014;135(3):611–23.

581 30. Bosse T, ter Haar NT, Seeber LM, v Diest PJ, Hes FJ, Vasen HFA, et al. Loss of
582 ARID1A expression and its relationship with PI3K-Akt pathway alterations, TP53 and
583 microsatellite instability in endometrial cancer. Mod Pathol an Off J United States Can
584 Acad Pathol Inc. 2013 Nov;26(11):1525–35.

585 31. Caumanns JJ, Wisman GBA, Berns K, van der Zee AGJ, de Jong S. ARID1A mutant
586 ovarian clear cell carcinoma: A clear target for synthetic lethal strategies. Biochim
587 Biophys acta Rev cancer. 2018 Dec;1870(2):176–84.

588 32. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al.
589 Mutational landscape determines sensitivity to PD-1 blockade in non – small cell lung
590 cancer. 2016;348(6230):124–9.

591 33. Shlien A, Campbell BB, De Borja R, Alexandrov LB, Merico D, Wedge D, et al.
592 Combined hereditary and somatic mutations of replication error repair genes result in

Mutations in MLH1^{-/-} tumors

593 rapid onset of ultra-hypermutated cancers. *Nat Genet* [Internet]. 2015;47(3):257–62.

594 Available from: <http://dx.doi.org/10.1038/ng.3202>

595 34. Ripperger T, Schlegelberger B. Acute lymphoblastic leukemia and lymphoma in the
596 context of constitutional mismatch repair deficiency syndrome. *Eur J Med Genet*
597 [Internet]. 2016;59(3):133–42. Available from:
598 <http://dx.doi.org/10.1016/j.ejmg.2015.12.014>

599 35. Cuceu C, Colicchio B, Jeandidier E, Junker S, Plassa F, Shim G, et al. Independent
600 mechanisms lead to genomic instability in Hodgkin lymphoma: Microsatellite or
601 chromosomal instability. *Cancers (Basel)*. 2018;10(7).

602 36. Rigter LS, Snaebjornsson P, Rosenberg EH, Atmodimedjo PN, Aleman BM, Ten
603 Hoeve J, et al. Double somatic mutations in mismatch repair genes are frequent in
604 colorectal cancer after Hodgkin’s lymphoma treatment. *Gut*. 2018 Mar;67(3):447–55.

605 37. Rossi D, Bruscaggin A, La Cava P, Galimberti S, Ciabatti E, Luminari S, et al. The
606 genotype of MLH1 identifies a subgroup of follicular lymphoma patients who do not
607 benefit from doxorubicin: FIL-FOLL study. *Haematologica*. 2015;100(4):517–24.

608 38. Jiang H, Lei R, Ding SW, Zhu S. Skewer: A fast and accurate adapter trimmer for next-
609 generation sequencing paired-end reads. *BMC Bioinformatics*. 2014;

610 39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
611 Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.

612 40. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler
613 transform. *Bioinformatics*. 2009 Jul;25(14):1754–60.

614 41. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2:
615 Somatic mutation and copy number alteration discovery in cancer by exome
616 sequencing. *Genome Res*. 2012;

617 42. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for
618 annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*

Mutations in MLH1^{-/-} tumors

619 (Austin). 2012;

620 43. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The
621 human genome browser at UCSC. *Genome Res.* 2002;12(6):996–1006.

622 44. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial
623 sequencing and analysis of the human genome. *Nature.* 2001;

624 45. Wickham H. *ggplot2. Elegant Graphics for Data Analysis.* 2009. 221 p.

625 46. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-
626 customizable Venn and Euler diagrams in R. *BMC Bioinformatics.* 2011;12(1):35.

627 47. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular
628 visualization in R. *Bioinformatics.* 2014;30(19):2811–2.

629 48. Miller CA, Qiao Y, DiSera T, D'Astous B, Marth GT. bam.iobio: a web-based, real-
630 time, sequence alignment file inspector. *Nat Meth.* 2014 Dec;11(12):1189.

631 49. Jay JJ, Brouwer C. Lollipops in the Clinic: Information Dense Mutation Plots for
632 Precision Medicine. *PLoS One.* 2016;

633 50. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic
634 Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *N Engl J Med.* 2014
635 Dec;371(23):2189–99.

636 51. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1
637 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med.* 2015
638 Jun;372(26):2509–20.

639 52. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic
640 correlates of response to CTLA-4 blockade in metastatic melanoma. *Science.* 2015
641 Oct;350(6257):207–11.

642 53. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic
643 features. *Bioinformatics.* 2010;26:841–2.

644

Mutations in MLH1^{-/-} tumors

645 6. Figure legends

646 **Figure 1: General distribution of tumors in MLH1^{-/-} mice and tumor phenotyping of**
647 **immunoregulatory markers.** **(A)** Type and frequency of tumors in MLH1^{-/-} mice; data were
648 taken from a total of 90 mice in which spontaneous tumorigenesis was observed. Females
649 were more likely to develop thymic lymphomas, while males were prone to GIT. Generalized
650 splenic lymphomas were found in both genders. **(B)** Flow cytometric phenotyping of MLH1^{-/-}
651 tumors reveals a high abundance of immune-checkpoint molecules on lymphomas. The
652 phenotype of GIT is different with only a few cells expressing these exhaustion markers.
653 Shown data refer to % numbers of cells measured in a live gate on total white blood cells. **(C)**
654 Numbers of tumor-infiltrating lymphocytes in GIT as determined by flow cytometry. Left bar
655 – #7450; middle bar - #328; right bar – mean value of tumor-infiltrating lymphocytes in GIT
656 of MLH1^{-/-} mice. Values are given as percentage of lymphocytes measured from 50.000
657 events in a live gate. Mean + SD, n=10 mice.

658

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

666

667 **Figure 3: Mutational landscape and shared mutations in GIT versus lymphoma.** **(A)** The
668 statistics for each SNV type (missense, nonsense, and silent) are shown for the MLH1^{-/-} GIT
669 primary tumors and the lymphomas including the distribution of the exclusivity of the SNV
670 for each condition. **(B)** The GIT and lymphomas are showing 640 shared SNV in addition to

Mutations in MLH1^{-/-} tumors

671 their exclusive mutations. **(C) Pairwise comparative mutational analysis of MLH1^{-/-} GIT**
672 **and lymphomas.** Ideogram plots show the genomic distributions of the missense mutations
673 occurring in the annotated/known genes for the MLH1^{-/-} GIT (#7450 and #328, left side). The
674 outermost track (track 0) is the cytoband of the mm10 assembly, followed by four tracks that
675 visualize the SNV belonging to a specific condition (MLH1^{-/-} GIT primary tumors (#7450 and
676 #328, spleen and skin lymphoma). All missense SNV for MLH1^{-/-} GIT primary tumors are
677 blue (#7450, track 1) and yellow (#328, track 3) respectively with their corresponding
678 coordinate on the mouse reference genome mm10 cytoband. The exclusive missense SNV for
679 the MLH1^{-/-} GIT primary tumors are red (tracks 2 and 4) and annotated with a corresponding
680 ID containing the SNV position and the affected gene name. Analogously to for the GIT, the
681 splenic lymphoma (blue) and the skin lymphoma (yellow) are compared for all missense SNV
682 in a pairwise fashion (right side). Red are the exclusive SNV respectively.

683

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

695

696

Mutations in MLH1^{-/-} tumors

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

Mutations in MLH1^{-/-} tumors

706 **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

707 GIT – gastrointestinal tumor.

Mutations in MLH1^{-/-} tumors

708 **Supplementary Figure 1: Distribution of somatic mutations in GIT and lymphomas in**
709 **genomes of female versus male mice.** The graph shows the mutational load of non-
710 synonymous mutations on the respective chromosome in the female sample of GIT (#7450)
711 and in the male sample (#328). The mutational load is normalized for the size of the
712 respective chromosome. The coloring of bars indicates the female GIT sample, the male GIT
713 sample and the lymphomas (female, splenic and skin).

714

715 **Supplementary Figure 2: Distribution of somatic mutations in GIT and lymphomas on**
716 **chromosome X of female versus male mice.** Mutational load of non-synonymous mutations
717 on chromosome X in the female sample of GIT (#7450) and in the male sample (#328). The
718 mutational load of the GIT was compared further to the two female samples of the
719 lymphomas (splenic and skin). Coloring of bars represents the ratio of the six nucleotide
720 changes (C > A, C > G, C > T, T > A, T > C, and T > G) that are most affected by
721 substitutions in the X-chromosome.

722

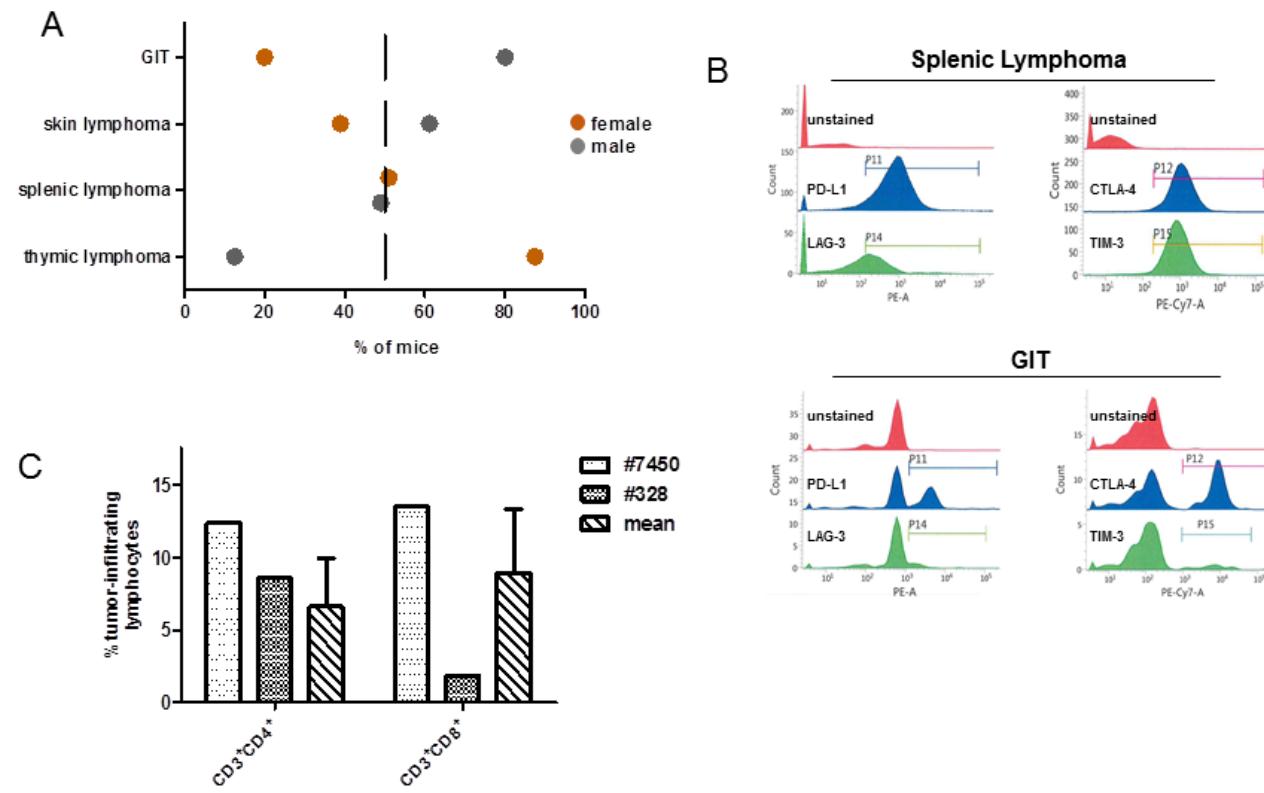


Figure 1

Mutations in MLH1^{-/-} tumors

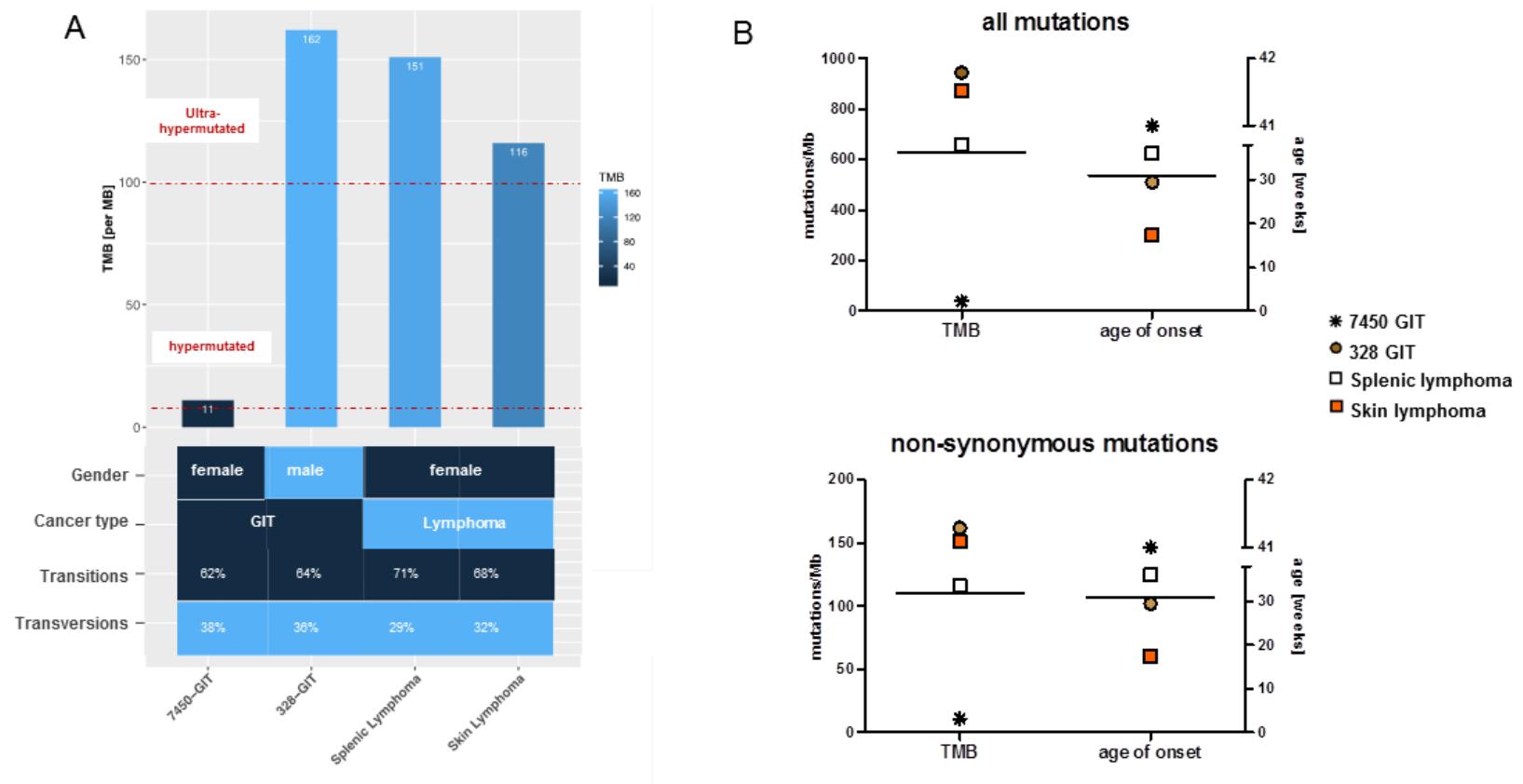


Figure 2

Mutations in MLH1^{-/-} tumors

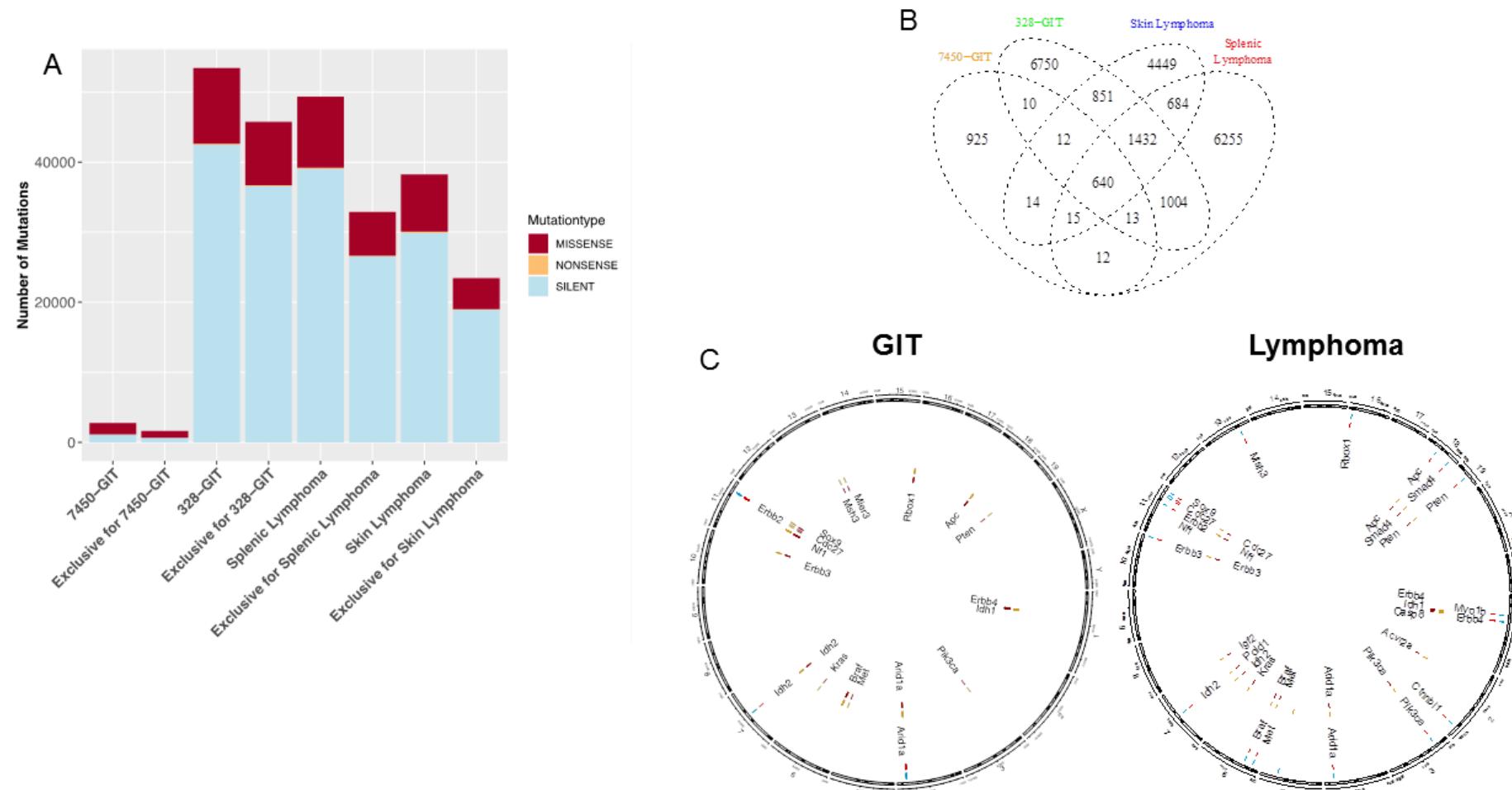


Figure 3

725

Mutations in $MLH1^{-/-}$ tumors

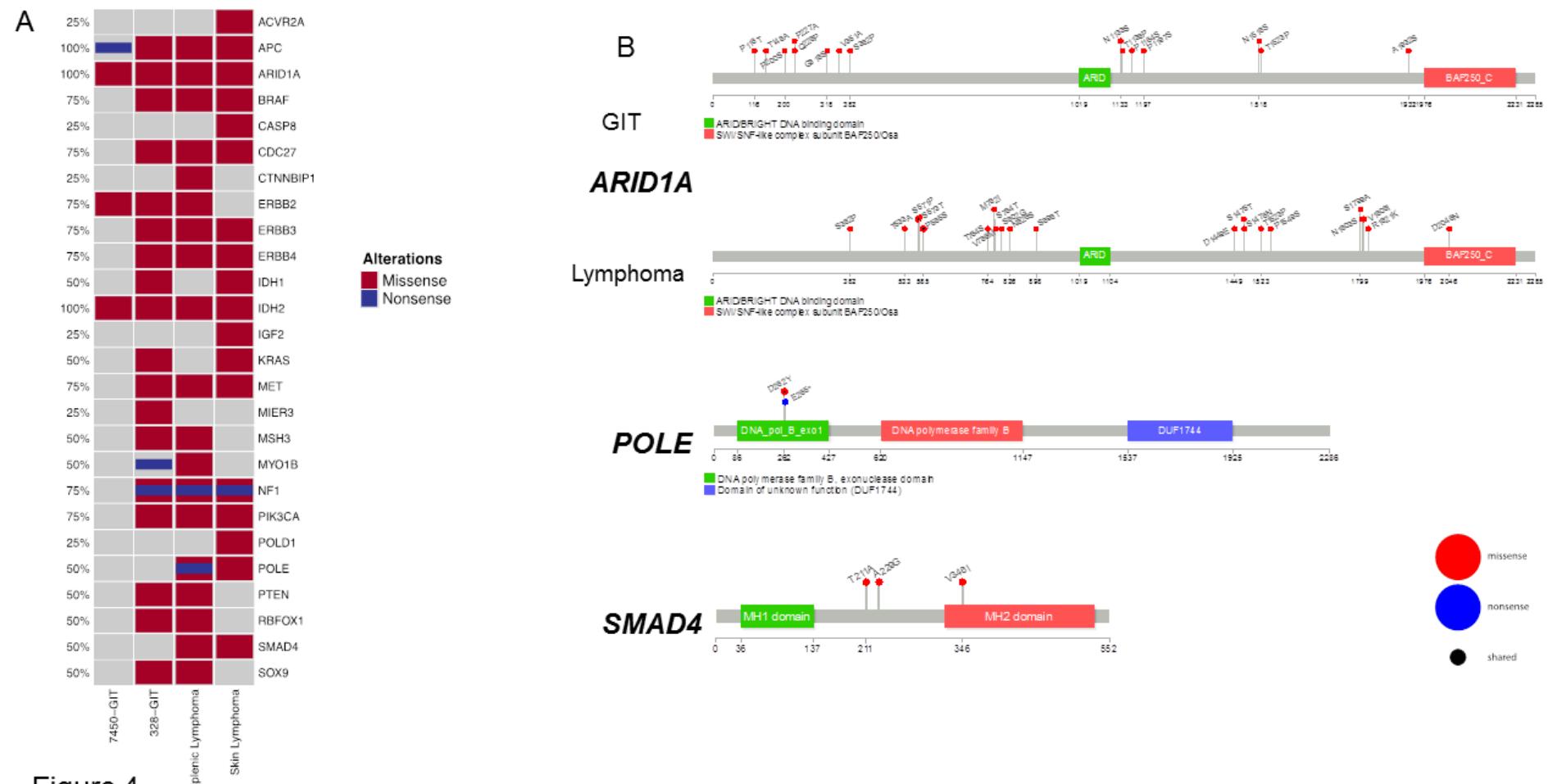


Figure 4

Mutations in $MLH1^{-/-}$ tumors

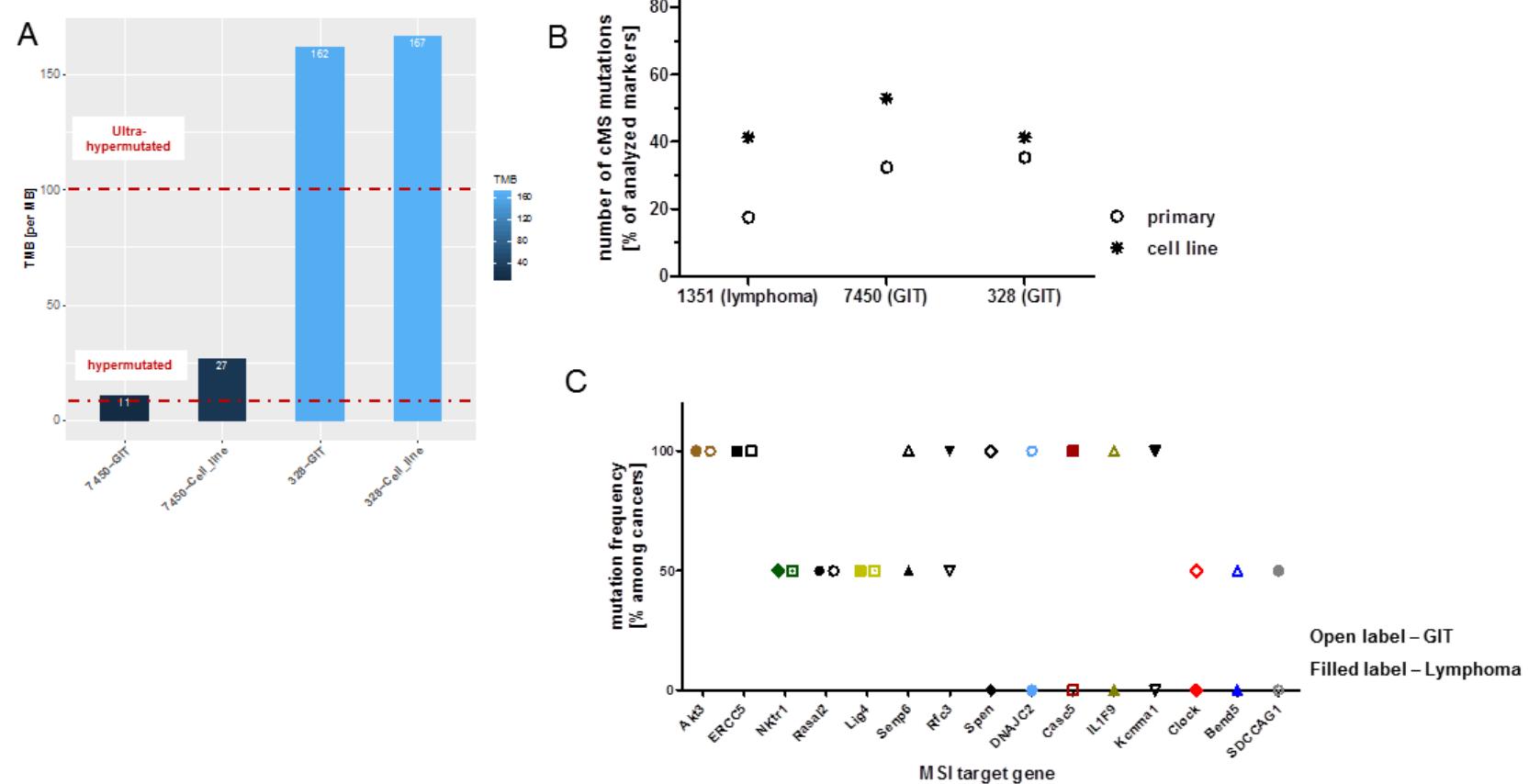
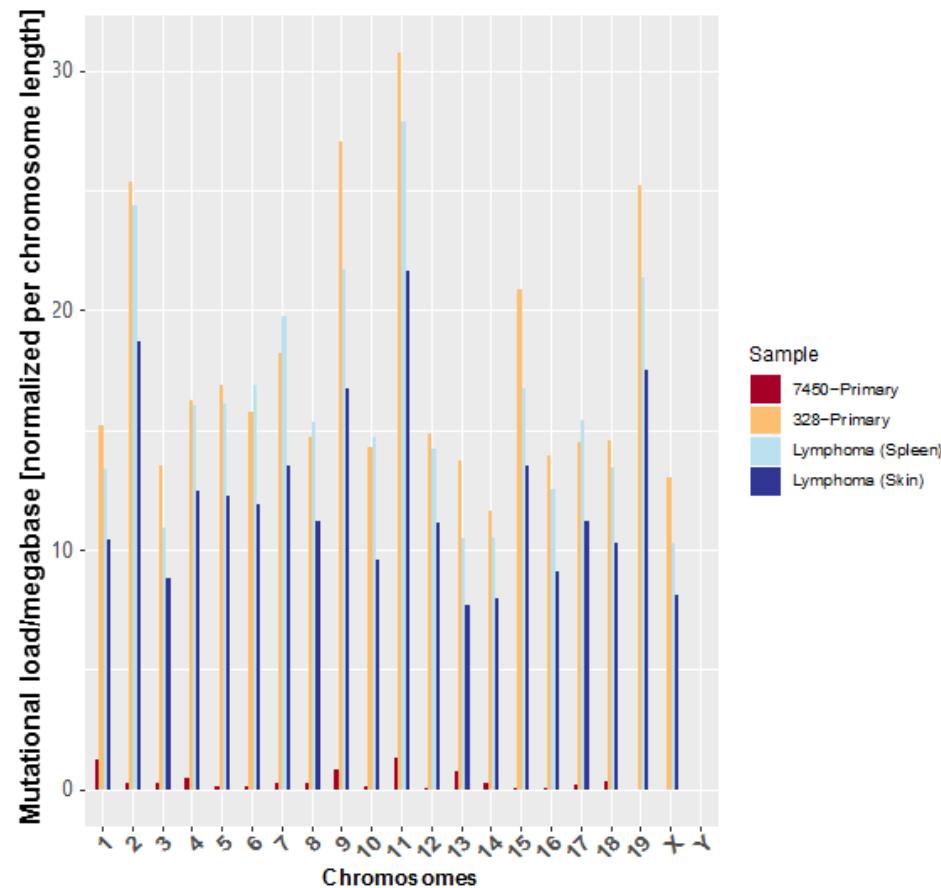


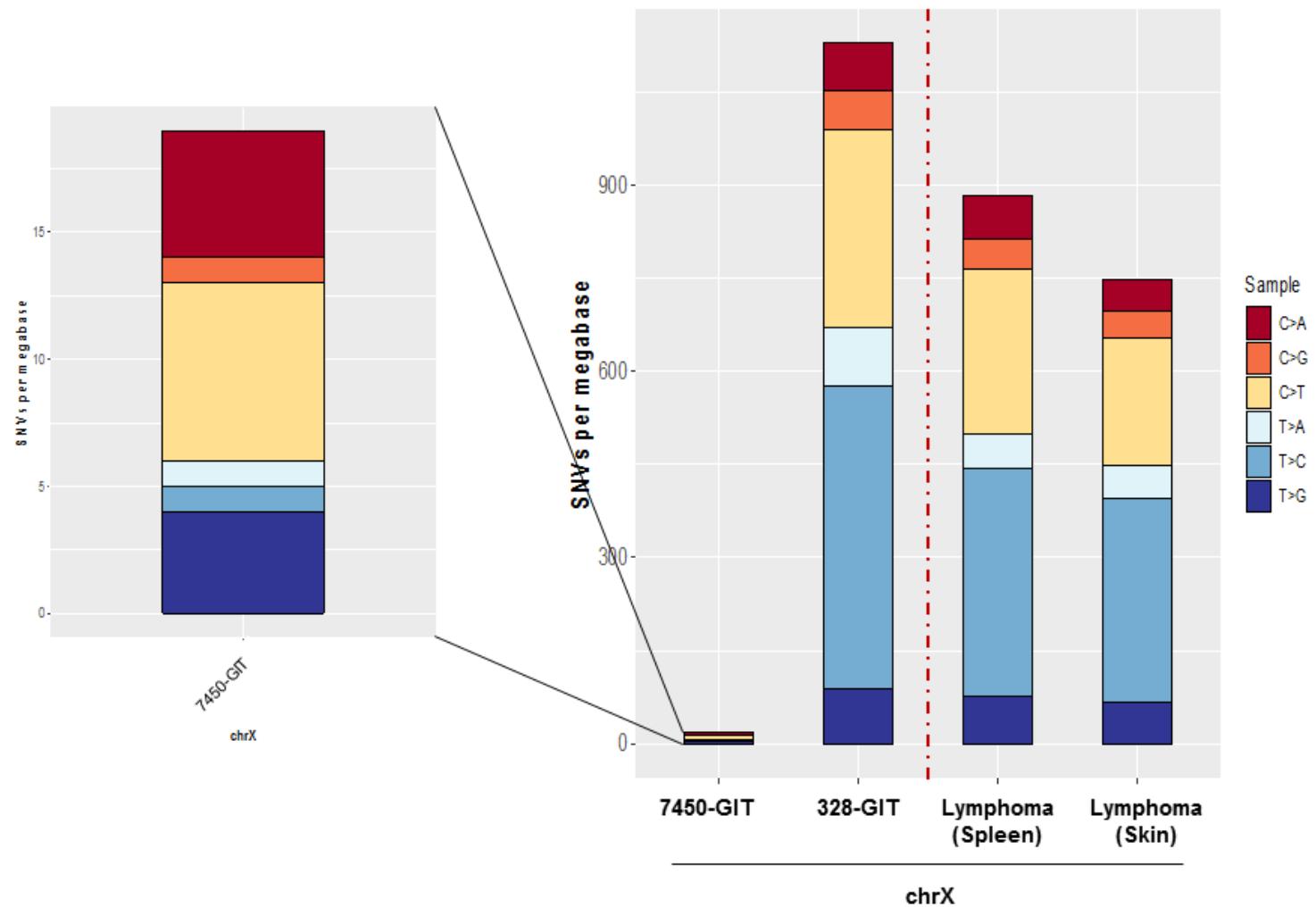
Figure 5

Mutations in $MLH1^{-/-}$ tumors



Supplementary Figure 1

Mutations in MLH1^{-/-} tumors



Supplementary Figure 2