bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

March 23, 2019

**Title**: Clinical and molecular characterization of virus-positive and virus-negative Merkel cell carcinoma

**Authors**:

Gabriel J. Starrett[1], Manisha Thakuria[2,3], Tianqi Chen[4], Christina Marcelus[5], Jingwei Cheng[5,6], Jason Nomburg[5], Aaron R. Thorner[7], Michael K. Slevin[7], Winslow Powers[7], Robert T. Burns[7], Caitlin Perry[8], Adriano Piris[2], Frank C. Kuo[9], Guilherme Rabinowits[3,5,10], Anita Giobbie-Hurder[4], Laura E. MacConaill[7,9], James A. DeCaprio[3,5,6,11]

1. Laboratory of Cellular Oncology, CCR, NCI, NIH, Bethesda, MD

2. Department of Dermatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

3. Merkel Cell Carcinoma Center of Excellence, Dana-Farber/Brigham Cancer Center, Boston, MA

4. Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA

5. Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA

6. Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

7. Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, MA

8. Department of Informatics & Analytics, Dana-Farber Cancer Institute, Boston, MA

9. Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

10. Current address: Miami Cancer Institute, Miami, FL

11. Corresponding author

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



**Abstract**

Merkel cell carcinoma (MCC) is a highly aggressive neuroendocrine carcinoma of the skin mediated by the integration of Merkel cell polyomavirus (MCPyV) and expression of viral T antigens or by ultraviolet induced damage to the tumor genome from excessive sunlight exposure. An increasing number of deep sequencing studies of MCC have identified significant differences between the number and types of point mutations, copy number alterations, and structural variants between virus-positive and virus-negative tumors. In this study, we assembled a cohort of 71 MCC patients and performed deep sequencing with OncoPanel, a next-generation sequencing assay targeting over 400 cancer-associated genes. To improve the accuracy and sensitivity for virus detection compared to traditional PCR and IHC methods, we developed a hybrid capture baitset against the entire MCPyV genome. The viral baitset identified integration junctions in the tumor genome and generated assemblies that strongly support a model of a hybrid, virus-host, circular DNA intermediate during integration that promotes focal amplification of host DNA. Using the clear delineation between virus-positive and virus-negative tumors from this method, we identified recurrent somatic alterations common across MCC and alterations specific to each class of tumor, associated with differences in overall survival. Comparing the molecular and clinical data from these patients revealed a surprising association of immunosuppression with virus-negative MCC and significantly shortened overall survival. These results demonstrate the value of high-confidence virus detection for identifying clinically important features in MCC that impact patient outcome.

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



## Introduction

Merkel cell carcinoma (MCC) is a highly aggressive neuroendocrine carcinoma of the skin. Risk factors for developing MCC include advanced age, light skin color with excessive sunlight exposure, and a variety of immunocompromised conditions (1). In 2008, Merkel cell polyomavirus (MCPyV) was first detected by Southern blot in some but not all MCC tumors with integration of viral DNA occurring at several different chromosomal sites. Importantly, an identical clonal integration pattern was detected in one tumor and corresponding metastatic lymph node (2). This important insight implied that integration of the viral DNA was an early if not initiating event in virus-positive MCC oncogenesis. MCPyV infects most people, typically at an early age, and results in an asymptomatic and lifelong infection indicated by the presence of antibodies to the viral coat protein VP1 (3, 4). Although MCPyV DNA can be readily detected on the skin, the cell types where the virus replicates *in vivo* have not been determined (5).

Since the original discovery of MCPyV, it has become increasingly clear that virus-positive MCC has a different etiology than virus-negative MCC (1). While virus-positive MCC expresses the viral oncogenes Large T antigen (LT) and Small T antigen (ST), the tumor genome usually contains very few mutations in cellular oncogenes and tumor suppressor genes. In contrast, studies using whole exome or targeted hybrid capture sequencing have revealed that virus-negative MCC has an exceptionally high somatic mutation load predominated by UV-mediated mutations with frequent mutations in *RB1*, *TP53*, *NOTCH1*, and *FAT1* (6, 7). Whole genome sequencing (WGS) of MCC confirmed virus-positive MCC exhibits a globally lower, non-UV-mediated, mutation burden as well as few somatic copy number amplifications, deletions, and rearrangements compared to virus-negative MCC, while providing new insights into the structure and mechanism of virus integration (8).

Accurate detection of the presence of MCPyV and distinguishing between virus-positive and virus-negative MCC is important for understanding the oncogenesis and cell-of-origin and potentially for therapeutic options. Currently, there is no routine clinical effort to distinguish

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



between virus-positive MCC and virus-negative MCC. Several recent studies have hinted at differences between virus-positive MCC and virus-negative MCC in presentation, age, and response to immunotherapy (9-15). However, current techniques for determining viral status have yielded either inaccurate or ambiguous results. Although WGS provides much more genetic information on the tumor and viral genome compared to targeted approaches, it remains impractical for clinical evaluation of MCC.

The most common methods for detection of MCPyV in MCC include PCR amplification of MCPyV DNA from DNA isolated from MCC tumors or immunohistochemistry (IHC) staining for MCPyV LT using monoclonal antibodies CM2B4 and Ab3 (16, 17). However, both PCR and IHC have been shown to be unreliable in distinguishing between virus-positive from virus-negative MCC. For example, a recent study of 282 cases of MCC evaluated virus-positivity by IHC with CM2B4 and Ab3 and PCR with a previously validated primer set (18). Notably, there was concordance for all three assays in only 167 of 282 (59.2%) cases with an additional 62 cases positive for 2 of the 3 tests. The remaining 53 (18.8%) were positive for 1 test or none. This study assigned the MCC to be virus-positive if 2 or 3 tests were positive; therefore, detection of viral DNA by PCR alone was not sufficient for a tumor to be called virus-positive MCC. Furthermore, because of the sensitivity of PCR in detecting DNA, a lower limit of 0.01 copy of MCPyV DNA per tumor cell was called virus-positive MCC. Tumors containing < 0.01 viral copies/cell were called virus-negative. A different study using RNA-FISH to detect mRNA specific for MCPyV LT and ST found this method to be as sensitive as qPCR when using two primer sets and the viral copy number was set to $\geq$ 0.004/cell (19). The AMERCK test detects circulating antibodies against the MCPyV ST (20). The sensitivity of this test is low for detection of virus-positive MCC but, when positive, can be used as a biomarker for disease status (20).

The high somatic mutation burden in virus-negative MCC is predicted to result to yield more tumor neoantigens than melanomas or non-small cell lung cancers (median of 173, 65, and 111 neoantigens/sample, respectively) (21) (22). As observed for other tumor types, the high

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



neoantigen burden in virus-negative MCC corresponds with a higher degree of tumor infiltrating lymphocytes in some tumors, but these tumors also express PD-L1 rendering these lymphocytes ineffective (7). Despite the numerous observed differences in mutation rate and number of predicted neoantigens, both virus-positive MCC and virus-negative MCC tumors have shown high response rates to PD-L1 and PD1 checkpoint blockade therapy (14, 15).

For further advancements to be made in understanding MCC, especially for patients not responsive to current therapies, clear and accurate determination of the MCPyV virus status and actionable variants in these tumors are required. In this study, we developed a viral hybrid capture next-generation sequencing (NGS) method to detect the presence of integrated MCPyV DNA in FFPE clinical specimens. This approach was combined with targeted sequencing of several hundred cancer-related genes to assess oncogenic changes in the tumor genome. We compared this viral hybrid capture approach to PCR detection of viral DNA, IHC for detection of MCPyV LT, and synoptic assessment of MCC pathology.

## Results

### Summary of patient cohort

A total of 71 patients diagnosed with MCC were included in this study (Table 1). The median (95% CI) follow-up duration from initial diagnosis of MCC was 47 (95% CI: 38 - 60) months based on inverse Kaplan-Meier estimation. Overall, 69 enrolled patients were white and two were black. Forty (56%) patients were male. The median age was 70 years (range: <50 to 93). The initial site of MCC presentation was in the head and neck (27%), upper extremity (20%), lower extremity (21%) and trunk (32%). The seventh edition TNM staging system of the American Joint Committee on Cancer (AJCC) was used to classify the initial presentation of MCC with 27% presenting at stage I, 14% Stage II, 42% Stage III, and 17% Stage IV.

### Somatic Variant Analysis of Targeted Sequencing

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



All 71 patients underwent OncoPanel analysis (23). Genomic studies were performed using DNA isolated from tumors obtained at the time of initial diagnosis (50) or upon relapse (21). The total number of mutations ranged from 0 to 73 corresponding to a Tumor Mutational Burden (TMB; mutations/megabase) from 0 to 38.89 with four cases containing no detectable mutations (Figure 1A, Table S1). From this mutation data, patients were binned into TMB-high (>=20), TMB-intermediate (>6<20), and TMB-low (<=6) . A limited set of mutation signatures could be identified (see Methods). The UV mutational signature (Signature 7) was detected in 24 cases, corresponding to the TMB-high patients (24). Additional mutational signatures identified included Aging (Signature 1; 3 cases), APOBEC (Signatures 2 & 13; 4 cases with 3 that also had an UV signature), and Signature 5 (one case) (Figure 1A, Table S1). TMB had some correlation with the number of copy number altered genes (Figure 1B). Several genes including *RB1*, *TP53*, *KMT2D*, *NOTCH1*, *NOTCH2,* and *FAT1* were highly enriched for missense and truncating mutations (Figure 1C, Figure S1). Single and dinucleotide substitutions in *RB1* and *TP53* revealed that most were likely mediated by UV damage (CC>TT, C>T; Figure 1D).

Copy number variants (CNVs) were examined individually as well against each other and other likely functional somatic changes for significant co-occurrence or mutual exclusivity (Table S2). Clusters of significantly co-occurrent CNVs were determined via network analysis (Figure 2A, Table S3). From these analyses, two chromosomal regions were found to be altered in more than 36% of cases (Figure 2B & C). Chromosome 10 (cluster 14) had frequent copy number loss with many tumors showing complete loss of the entire chromosome (Figure 2B) (25). Some cancer-relevant genes on chromosome 10 include PTEN and SUFU, negative regulators of PI3K and Hedgehog signaling respectively, with deletions reported in prior studies of MCC (25, 26). A region of Chr1q (cluster 13) was amplified in 28 cases. This region includes MDM4 (also known as MDMX), whose protein product cooperates with MDM2 to promote the ubiquitination and subsequent degradation of p53 (Figure 2B) (27, 28). In addition, we observed

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



a focal amplification of MYCL in Chr1p (cluster 4), which was reported in an earlier study of MCC (29).

The CNV clusters were observed at nearly equal frequencies in both TMB-high and TMB-low cases (Figure 2B & C). Many of the other CNV clusters were strongly associated with UV signature and high TMB (Figure 2C). However, cluster 5 that included an amplification of chromosome 6 was more than twice as frequent in TMB-low tumors than TMB-high tumors (Figure 2C). Interestingly, 35% of metastatic tumors carried cluster 5 and all but one of these metastatic tumors were TMB-low MCC. Furthermore, CNV cluster 5 was 2.5 times more frequent in TMB-low (22.2%) than TMB-high (8.7%) tumors in primary tumors. Both TMB-low and TMB-high patients with amplification of CNV cluster 5 had significantly improved overall survival compared to wild-type carrying patients that had a median survival time of 41.65 months ($p=0.0027$). Restricting this analysis to only primary tumors, revealed that there were no deaths at the time of this study in patients carrying this amplification ($p<0.001$) (Figure 2D & E). The recurrent copy number events on chromosomes 1 and 10 were compared within TCGA for similarities to other tumor types (Figure 2F-G). This analysis revealed that the chromosome 1 amplification was also frequently observed in ovarian, breast, and bladder cancers; whereas, the chromosome 10 loss was most frequently seen in prostate cancer.

**Analysis of Viral Sequences in Tumors**

Of the 71 tumors analyzed by Oncopanel, 48 were re-analyzed by Oncopanel (Profile/OncoPanel version 3, POPv3) combined with a hybrid-capture probe bait set targeting the entire genome of MCPyV (VB2). For the 48 cases, the number of MCPyV reads ranged from 0-21,095,751 with only a single case having zero MCPyV reads (Figure 3A). In total, 28 cases had substantial reads (>6,800) mapping to the MCPyV genome that also supported integration of the virus into the host genome. For the remaining 20 cases without evidence for integration, the number of viral reads ranged from 0 to 971. Generally, these cases had reads that covered

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



less than 10% of the viral genome with the normalized coverage less than two logs compared to samples with evidence for virus integration (Figure 3B & C, Table S3). Concordantly, the viral reads from most of these cases were unable to be assembled into larger viral contigs. Two cases, MCC011 and MCC015, had 212 and 177 MCPyV reads that could be assembled into nine and five contigs each smaller than 761 base pairs, respectively. Case MCC007 had the most reads of any likely virus-negative sample and could be assembled into a single 5343 bp contig. However, analysis of the point and deletion variants in these aforementioned viral contigs revealed that they were identical to the virus sequence from patient MCC037 indicating that the viral reads resulted from low level contamination (<0.005% of MCC037 MCPyV reads were detected in other samples).

For the 28 cases with evidence for integration of the viral DNA into the tumor, the number of reads mapping to the viral genome ranged from 6,824-21,095,751 (Median: 28,726). Consistent with previous reports, the integrated viral genome had undergone extensive mutagenesis with large deletions (>100 bp) particularly in the 3' half of LT (Figure 4). In 10 cases, approximately half of the total viral genome was deleted, 6 cases had approximately 25% of the viral genome deleted, while 12 cases had sequences corresponding to the entire or nearly complete genome (Figure 3C & 4). In all but one of the cases with a nearly complete coverage of the viral genome, there was a clonal point mutation which inserted a premature stop codon in LT resulting in truncated proteins between 208 and 771 amino acids (Figure 5A). In a single case (MCC054), LT was truncated by a 5bp deletion resulting in a frame shift putting a premature stop codon in frame. In all cases, the non-coding control region, the N-terminal 208 residues of LT, and an intact ST region of the viral genome were conserved. Beyond indels and nonsense mutations, LT also carried numerous clonal missense mutations (Figure 5A). In stark contrast, ST only had missense mutations at three residues, and the amino acid change A20S is consistent with a previously observed MCPyV strain difference (GenBank identical protein accession number: ACI25295.1). The other missense mutations occurred clonally at H41Y and

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



N100S once in the entire cohort (Figure 5B). Neither of these mutations are present in any of the ST sequences in GenBank.

The integration sites were mapped using the oncovirus tools suite (https://github.com/gstarrett/oncovirus_tools) (Figure 6A, Table 2). As previously reported, integrations primarily fell into two categories: either those that appear as a single integration event or as two events separated by >10 kilobases (kb) (8). Interestingly, two cases had integration events in non-identical but overlapping sites in chromosome 1 (Figure 6B). These represent the first reported cases of recurrent viral integration sites or hotspots in MCC.

Because of the limited number of capture targets of the OncoPanel platform, determining the extent of virus-mediated amplifications expected in the tumor genome was not possible. However, using the normalized viral coverage, the estimated number of viral genome copies ranged between 1 and 1881 copies (median: 7, interquartile range (IQR): 4-13) (Table 2). Additionally, based on previous observations, we can infer that the regions between viral integration sites were amplified (8). When annotating these regions, we observed that they frequently contain enhancer regions that may contribute to oncogenesis as seen in HPV-associated tumors (30). Patient MCC026 had integrations on chromosomes 9, 16, and 18, all of which had integration sites separated by 107.5 and 129.9 kbp. Using automated computational methods, we could not confidently determine an integration site for case MCC037 with the highest viral genome copy number in this study. Manually interrogating the human sequence hits from the assembly revealed that it matched a tandem repeat sequence flanked by MLT1H2 ERVL-MaLR elements. Based on the estimated copy number and the assembly graph, the viral component of this fusion DNA structure is likely larger than 10 Mbp (Figure S4).

With the high depth of coverage facilitated by the targeted NGS method, high resolution assemblies for the integrated virus were generated. Many integrations that appeared as a single linear contig contained a single copy of the viral genome flanked by the host genome (Figure 6C Figure S4). However, other integrations generated more complex assembly graphs with a

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



multiple contigs linked together in a “pigtails” conformation (Figure 6D, Figure S4). Based on coverage and conformation, this graph likely represents an integration event containing partially duplicated viral genomes fused to different segments of the human genome. For samples with distant integration sties, the directionality of the virus-host junctions strongly supports a circular virus-host DNA fusion intermediate prior to reintegration into the host chromosome. This model is further supported by assemblies in which one arm of the fusion contains sequences from both distant sites of the human genome (Figure 6E, Figure S4).

To address a possible mechanism for integration, we looked for microhomology between the human and MCPyV genomes at fusion junctions. We found significant enrichment for 4, 5, and 7 bp sequence microhomology at the site of integration compared to randomly selected sites in the human and MCPyV genomes (Figure 6F). There was no significant increase in overall homology between MCPyV and human DNA at integration sites versus randomly selected sites. Patient MCC027 had the integration site with the longest stretch of homology and MCC041 had both the integration site with the greatest overall homology on its 3’ end and lowest homology with no microhomology greater than 1bp on its 5’ end (Figure 6G). Additionally, we annotated integration sites for proximity to repeat elements, including LINES, SINES, LTR retrotransposons, and simple repeats in the human genome. No type of repeat element was significantly enriched, but all integration sites were within 1.5kb of a repeat element and there was a trend towards integrations near LTR retrotransposons and low-complexity regions (Figure 6H).

**Distinguishing virus-positive MCC from virus-negative MCC using somatic variants in comparison to Immunohistochemistry and PCR**

Given the striking differences in the number of mutations and mutational signature we observed in the VB2 dataset that strongly correlated with virus integration, we compared the data from the OncoPanel and POPv3VB2 datasets to determine the viral status of all 71 tumors

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



studied (Table 3). From the OncoPanel sequencing, we identified off-target reads for MCPyV in a total of 18/71 cases, ranging from 1 to 194 reads total. When compared to the VB2 data, there was a rough correlation between the number of off-target reads and the number of MCPyV reads in the VB2 dataset. There were 8 samples with MCPyV reads in the OncoPanel dataset that were not also analyzed by VB2. None of these 8 cases have any evidence for a UV mutational signature.

We assessed the total number of mutations, TMB, UV signature, and detection of MCPyV reads to characterize each tumor as virus-positive MCC or virus-negative MCC. Using these criteria, we called 25 tumors as virus-negative. All but one of the virus-negative MCC tumors had a UV mutational signature, had higher number of total mutations (18-73), higher TMB, and absence of integrated MCPyV compared to virus-positive MCC. The virus-negative MCC without a UV signature (MCC007) originally presented as a subcutaneous breast mass (31). A total of 46 MCC tumors of the 71 analyzed were characterized as virus-positive. These virus-positive MCC had an absence of UV mutational signature, a lower number of total mutations (0-16), and lower TMB than did any of the virus-negative MCC. The TMB-low and -high categories had perfect concordance with virus-positive and -negative MCC determined by sequencing, respectively. The TMB-intermediate samples were mostly virus-negative (7/9), but the lowest two TMB patients in this category are likely virus-positive based on VB2 sequencing and absence of UV mutation signature.

FFPE sections were available for 28 of the 71 cases to assess for MCPyV LT by IHC with antibodies CM2B4 and Ab3. For 8 of the virus-negative MCC, all were negative by IHC with both antibodies. For 20 virus-positive MCC cases, we observed 16 stained positive with both antibodies and 4 were negative (Table 3). In addition, DNA was tested by PCR with 5 primer sets for 15 cases. In 9 virus-positive MCC cases, all returned positive results with 2 to 5 primer sets (Table 3). For 6 virus-negative cases, PCR was negative for 5 primer sets and one was positive with one primer set. Interestingly, the virus-negative MCC (MCC007) with one PCR

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



primer set positive also ranked at the TMB borderline (9.58) between virus-negative and virus-positive and did not score as having a UV mutational signature; rather, the majority of mutations were classified as APOBEC-associated.

A synoptic review of dermatopathology was available for 19 cases (Table S4) (32). Criteria evaluated included procedure, laterality, site, size (mm), thickness (mm), lymphovascular invasion, tumor extension, mitotic rate, tumor infiltrating lymphocytes (TILS), growth pattern, neurotropism, and necrosis (%). TILS were largely absent in both virus-positive and virus-negative samples. An infiltrative growth pattern was observed in virus-positive MCC and nodular or nodular infiltrative observed in both forms of MCC. Neurotropism was present in three cases of virus-positive MCC and necrosis ranged from 0 to 40%.

**Statistical comparison of clinical and molecular characteristics**

Overall, 28 patients remained disease free after initial therapy and 43 developed one or more relapses or persisted as stage IV. According to the biopsy type and first recurrence status, patients could be grouped into primary biopsy with no further recurrence (N=30), primary biopsy with further recurrence (N=22), and recurrence biopsy (N=19). For all samples annotated with recurrence, the first recurrence occurred before the biopsy was obtained. Regardless of the biopsy type, patients were grouped into no relapse (N=30) and relapse (N=41). **Table 4** shows the association between relapse and genomic characteristics. Among 71 patients, 30 (42.3%) patients had no relapse and 41 (57.7%) had relapse after initial diagnosis. From the Fisher's exact test results, UV, RB1 status, TP53 status, and virus status were all not significantly associated with relapse (Table 4). If the OncoPanel data obtained after relapse (and prior treatment) was excluded and restricted to the 52 patients with primary biopsy, UV, RB1 status, TP53 status, and virus status were all not significantly associated with relapse (Table S5).

Consistent with known risk factors of MCC, 10 of the 71 cases had immunosuppression diagnosed prior to developing MCC. Remarkably, 8 of the 10 (80%) of the immunosuppressed

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



cases were identified as virus-negative MCC with relatively high TMB compared to the 28% virus-negative MCC in immunocompetent patients (Figure 7A, Table 5). Virus-negative MCC was present in three patients with solid organ transplantation, three with auto-immune disease including myasthenia gravis, rheumatoid arthritis, or granulomatosis with polyangiitis, one with monoclonal gammopathy of undetermined significance (MGUS) , and another with Waldenstrom's macroglubulinemia. In contrast, virus-positive MCC was identified in a patient with mantle zone lymphoma having been treated with Rituximab for 3 years and another with somatic mutations in NF1 and GATA2 (33). The median OS for patients with immunosuppression was 17.5 months (95% CI: 5.6-24.4) months, significantly shorter than patients without immunosuppression (48.5 months, 95% CI: 35.4-113.3, $p<0.01$) (Figure 7B, Table 5).

**Discussion**

We undertook this study to develop an assay to more accurately distinguish between virus-positive and virus-negative MCC. We built upon an NGS platform that has been instituted as a routine part of clinical care at the Dana-Farber Cancer Institute, Brigham and Women's Hospital, and Boston Children's Hospital. The viral hybrid capture assay, VB2, acquired a high number of MCPyV reads for many samples. Importantly, evidence for specific integration was associated with all cases with a high number of reads (>6,000). Spurious MCPyV reads were also detected in 19 of 20 MCC cases that were deemed to be virus-negative by TMB and UV mutations. There was no evidence for integration in these cases rather these reads could be traced to be extremely low level contamination from MCC037 during library preparation or sequencing. In contrast, true virus-positive MCCs have low TMB with clear assemblies of virus-host junctions with MCC-hallmark deletions in the MCPyV genome.

Integration sites were observed on 12 different chromosomes with the most occurring on chromosome 5. In addition, two fully overlapping integration sites from two different tumors were

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



observed on chromosome 1 separated by only 10-20kb. Based on the clonality of deletions and point mutations in the MCPyV genome, these events most likely occurred before or during integration as was similarly determined from another study on MCC cell lines (34). From previous publications, we can infer that these integration sites likely have amplified adjacent regions of the host genome in a tandem head-to-tail conformation (8). These integration sites frequently are within 50 kb of each other and flank or fully contain enhancer regions of the human genome. These observations are consistent with reports of HPV integration in cervical and head and neck cancers as well as cell lines (35). Particularly, HPV integration events can amplify enhancer regions to high copy number creating super enhancers that can drive expression of adjacent genes and tumor cell survival. This likely represents another important mechanism beyond the expression of the viral tumor antigen in promoting tumorigenesis.

A recurrent amplification of chromosome 6 has previously been observed for MCC; however, this observation predated the discovery of MCPyV and was not associated with morphology or outcome (36). In other cancers, such as basal cell carcinoma and ovarian cancer, this amplification is typically associated with worse outcome (36). Although the chromosome 6 amplification in this study was significantly associated with better overall survival, it was also more frequent in metastasis. This amplification co-occurring with MCPyV may represent a less fatal, but more metastatic subtype of MCC. Additionally, this result could be impacted by the immunotherapy administered to the patients with metastatic MCC.

Unexpectedly, we observed that 8 of 10 cases with immunosuppression were virus-negative MCC. While it was recognized in the early 1990s that individuals with hematologic malignancies that developed MCC had a poor prognosis (37), it was not until 1997 when a direct link between immunosuppression and MCC was postulated (38). At that time, a correlation was noted between medically induced immunosuppression with azathioprine and cyclosporine and the development and rapid spread of MCC. Early reports highlighted a prolonged period of immunosuppression prior to MCC development. The search for a viral

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



pathogen in MCC was initiated because of reports linking MCC with immunosuppression and with HIV-1/AIDS (2).

In the present report, three solid organ transplant recipients, three with chronic auto-immune diseases, and two with hematologic malignancies developed virus-negative MCC. The risk for developing MCC is increased in patients with chronic inflammatory disorders such as rheumatoid arthritis or medically induced immunosuppression for solid organ transplantation (38-41). Skin cancers account for 40-50% of all posttransplant malignancies with squamous cell carcinoma (SCC) and basal cell carcinoma (BCC) comprising 90-95% of these skin cancers (42). Importantly, some therapeutics used in organ transplantation increase risk for developing skin cancers. Azathioprine can sensitize cells to UV-induced damage through the incorporation of a metabolite into DNA that generates reactive oxygen species upon exposure to UV light (43). In patients with rheumatoid arthritis, methotrexate and anti-TNF drugs were associated with an increased risk of non-melanoma skin cancer (44). The increased risk for skin cancers in organ transplant recipients and rheumatoid arthritis is associated with UV-light induced mutagenesis for SCC and BCC. Therefore, the increased risk for UV-induced skin cancers may also extend to virus-negative MCC.

The AIDS-defining malignancies include Kaposi's sarcoma, driven by human herpes virus 8 (HHV-8) also known as Kaposi's sarcoma herpes virus (KSHV), non-Hodgkin lymphoma, often triggered by Epstein Barr virus (EBV), and cervical cancer, resulting from human papilloma virus (HPV). Of note, MCC was never categorized as an AIDS-defining malignancy likely due to the rarity of the malignancy even in individuals with profound immunosuppression. Given the association of virus-negative MCC with other forms of immunosuppression observed in this study, it should not be assumed that HIV-1/AIDS associated MCC will be virus-positive MCC.

Despite the significant differences in the TMB between virus-positive and virus-negative MCC, there were few phenotypic differences in the two types of MCC. Based on histopathological features alone, two subtypes of MCC can be recognized: pure neuroendocrine

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



tumors and combined tumors with neuroendocrine and divergent (mainly squamous) differentiation. Most pure tumors are MCPyV-positive and CK20-positive while combined tumors are uniformly MCPyV-negative and occasionally CK-20 negative (9, 45). Virus-negative MCC can also present as pure neuroendocrine type MCC.

While genomic sequencing has revealed that virus-negative MCC has evidence for a high degree of UV damage, this does not exclude a role for UV exposure in the development of virus-positive MCC. The relative lack of UV damaged DNA in virus-positive MCC indicates that the etiologies are clearly different, suggesting that the precursor to virus-negative MCC was a recipient of life-long intense UV exposure while the virus-positive MCC were not exposed to sunlight for the same degree or for as long. It was reported that the early promoter of MCPyV responds to UV exposure and that levels of ST mRNA increased in UV exposed skin from a healthy human volunteer (46). Transient UV exposure could affect the immune response to virus-negative and virus-positive MCC etiology. The effect of UV radiation in the pathogenesis of MCC has been suggested to be more likely a result of immune modulation than direct effects on DNA itself (47).

## Materials and Methods

### Sample collection

DNA was isolated from FFPE sections of MCC tumors corresponding to the 71 individuals summarized in Table 1. In addition, FFPE sections were sectioned for immunohistochemistry with antibodies CM2B4 and Ab3 (17). Sections stained with hematoxylin and eosin were evaluated by synoptic review (32).

### Nucleic acid isolation, library preparation and sequencing

Purified DNA was quantified using a Quant-iT PicoGreen dsDNA assay (ThermoFisher). Library construction was performed using 200 ng of DNA, which was first fragmented to ~250

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



bp using a Covaris LE220 Focused ultrasonicator (Covaris, Woburn, MA) followed by size-selected cleanup using Agencourt AMPureXP beads (Beckman Coulter, Inc. Indianapolis, IN) at a 1:1 bead to sample ratio. Fragmented DNA was converted to Illumina libraries using a KAPA HTP library kit using the manufacturer's recommendations (ThermoFisher). Adapter ligation was done using xGen dual index UMI adapters (IDT, Coralville, IA).

Samples were pooled in equal volume and run on an Illumina MiSeq nano flow cell to quantitate the amount of library based on the number of reads per barcode. All samples yielded sufficient library (> 250 ng) and were taken forward into hybrid capture. Libraries were pooled at equal mass (3 x 17-plex and 1 x 18-plex) to a total of 750 ng. Captures were done using the SureSelect$^{XT}$ Fast target enrichment assay (Agilent, Technologies, Sant Clara, CA) with VB2 with and without supplementation with the OncoPanel (v3) bait set. Captures were sequenced on an Illumina 2500 in rapid run mode (Illumina Inc., San Diego, CA). OncoPanel sequences are available through AACR GENIE (https://www.aacr.org/Research/Research/pages/aacr-project-genie.aspx).

**Sequence alignment and somatic variant calling**

Pooled samples were de-multiplexed and sorted using Illumina's bcl2fastq software (v2.17). Reads were aligned to the reference sequence b37 edition from the Human Genome Reference Consortium as well as viral genomes targeted by the Virus Capture Baitset v2 using bwa mem (http://bio-bwa.sourceforge.net/bwa.shtml) (48). The viral genomes and human genome were combined into one alignment reference so reads could map to the closest matching reference sequence.

Duplicate reads were identified using unique molecular indices (UMIs) and marked using the Picard tools. The alignments were further refined using the Genome Analysis Toolkit (GATK) for localized realignment around indel sites and base quality score recalibration (49, 50).

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Mutation analysis for single nucleotide variants (SNV) was performed using MuTect v1.1.4 (CEPH control was used as the “project normal”) and annotated by Variant Effect Predictor v 79 (VEP) (51, 52). We used the SomaticIndelDetector tool that is part of the GATK for indel calling. After initial identification of SNVs and indels by MuTect and GATK respectively, the variants were annotated using OncoAnnotate to determine what genes were impacted and their effect on the amino acid sequence. OncoAnnotate also applied additional filters using the Exome Sequencing Project (ESP) and gnomAD datasets to flag common SNPs.

Variants that affect protein coding regions underwent further filtering/classification based on frequency in the gnomAD, ESP, and COSMIC (version 80) databases. If the frequency of the variant was less than or equal to 1% in all gnomAD and ESP populations, the variant was flagged as "REVIEW_REQUIRED". If the frequency of the variant was greater than 1% and less than or equal to 10% in all gnomAD and ESP populations and present in "COSMIC" database at least 2 times, the variant was flagged as "REVIEW_REQUIRED". If the frequency of the variant was between 1% and less than or equal to 10% in all gnomAD and ESP populations and not present in "COSMIC" database at least 2 times, the variant is flagged as "NO_REVIEW_GERMLINE_FILTER". If the frequency of the variant was greater than 10% in any gnomAD and ESP populations, the variant was flagged as "NO_REVIEW_GERMLINE_FILTER". Variants with a frequency greater than 10% in any gnomAD or ESP population were considered to be a common SNP irrespective of presence in the COSMIC database.

Variants in the viral genomes were called using samtools mpileup and bcftools from the aligned bam files. Called variants were filtered to have a minimum coverage of 5 reads and minimum allele frequency of 1% of total reads covering that base in a single sample. Variants were annotated based on the NC_010277.2 reference sequence in GenBank using SnpEff (PMID: 22728672).

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



**Recurrent copy number analysis**

Copy number variant calling was performed using a combination of VisCap Cancer and CNVkit as previously described (23)(53). All resulting gene copy number variants from all patients were compared against each other with UV status and significant mutual exclusivity/co-occurrence was calculated using Fisher's exact test corrected by FDR for multiple comparisons in the R statistical environment. Using the network and iGraphs packages the significantly co-occurrent variants were clustered into networks. The genes belonging to each distinct network cluster with more than 5 member genes were then labeled and extracted. Using these gene lists as cluster definitions each patient was evaluated for presence or absence of each CNV cluster. Presence of a CNV cluster was determined if more than 50% of the member genes of that cluster were modified in the same patient. Code is available from https://github.com/gstarrett/oncovirus_tools.

**Viral integration analysis**

A custom perl script was written to extract, assemble, annotate, and visualize viral reads and determine viral integration sites. Viral reads and their mates were first identified and extracted by those that have at least one mate map to the viral genome. Additional reads containing viral sequence were identified by a bloom filter constructed of unique, overlapping 31bp k-mers of the MCPyV genome (54). The human genome positions for any read with a mate mapping to the viral genome were output into a bed file and the orientation of viral and human pairs was stored to accurately deconvolute overlapping integration sites. This bed file was then merged down into overlapping ranges based on orientation counting the number of reads overlapping that range. Skewdness in coverage of integration junctions was calculated by the difference in the fraction of virus-host read pairs overlapping the first and second halves of the aforementioned ranges. This skewdness value was used to determine the orientation of the viral-host junction (i.e. positive values, junction is on the 3' end of the range; negative values,

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



junction is on the 5' end of the range), which was validated from the results of de novo assembly. Integrated viral genomes were assembled from extracted reads using SPAdes with default parameters (55). The assembly graphs from SPAdes were annotated using blastn against hg19 and the MCPyV reference genome with an e-value cutoff of $1x10^{-10}$. Annotated assembly graphs were visualized using the ggraph R package. Code is available from https://github.com/gstarrett/oncovirus_tools.

Integrations sites confirmed by reference guided alignment and assembly data were analyzed for stretches of microhomology between the human and viral genomes by selecting 10bp upstream and downstream of the integration junction on the viral and human genomes. Within these sequences stretches of identical sequence at the same position longer than two base pairs were counted. Overall homology between the sequences was calculated by Levenshtein distance (). Three integration junctions with indeterminate DNA sequence ranging from 1-25bp inserted between viral and human DNA were excluded from analysis. Expected microhomology was calculated by randomly selecting 1000 20bp pairs of non-N containing sequence from the human and MCPyV genomes.

Integration site proximity to repeat elements were determined using bedtools closest and repeatmasker annotations acquired from the UCSC genome browser (56). Expected frequency of integration near repeat elements were determined by randomly selecting 1000 sites in the human genome. Sites within 2kb of a repeat element were counted as close proximity.

**Statistics**

The association between relapse and genomic characteristics are tested with Fisher's exact test. Overall survival (OS) is defined as the time from initial diagnosis to death, and patients who did not die are censored at the date last follow-up date. The 95% confidence intervals of the median OS times are estimated using log(-log(OS)) methodology. Multivariable Cox regression model is built to investigate the association between virus status and overall

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



survival adjusted by other covariates. And model is fit using backward, forward, and stepwise selection to assess model consistency. Hazard ratios are presented with 95% Wald confidence intervals. Statistical significance is defined as $p \leq 0.05$.

Associations between recurrent CNV, TMB, or viral copies and overall survival were calculated and graphed using Graphpad Prism 7. Fisher's exact test and Kaplan-Meier curves were computed with the R statistical environment. Significant enrichment of microhomology and repeat elements at integrations sites was determined using Fisher's exact test between observed and expected events.

**Human subjects**

This study was conducted according to Declaration of Helsinki principles and approved by the Dana-Farber Cancer Institute institutional review board. Written informed consent was received from participants prior to inclusion in the study.

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



**Author Contributions**

Conceptualization: G.J.S., L.E.M., J.A.D.

Methodology: G.J.S., T.C., L.E.M., J.A.D.

Software: G.J.S., A.R.T.

Validation: G.J.S., J.C.

Formal Analysis: G.J.S., T.C., J.N., R.T.B., J.A.D.

Investigation: M.T., C.M., J.C., A.R.T, W.P., M.S., A.P., F.K., J.A.D.

Resources: M.T., C.P., G.R., A.G.-H., L.E.M., J.A.D.

Data Curation: G.J.S., T.C., C.P. A.R.T., L.E.M., J.A.D.

Writing-Original Draft: G.J.S., T.C., A.R.T., J.A.D.

Writing-Reviewing & Editing: G.J.S., L.E.M., J.A.D.

Visualization: G.J.S.

Supervision: A.R.T., A.G.-H., L.E.M., J.A.D.

Project Administration: A.G.-H., L.E.M., J.A.D.

Funding Acquisition: J.A.D.

**Acknowledgements**

This work was supported in part by US Public Health Service grants (report.nih.gov) R01CA63113, R01CA173023 and P01CA203655 to JAD. JAD has received honoraria for participation in advisory board from Merck & Co., Inc. JAD has received research funding from Constellation Pharmaceuticals, Inc. Salary support for GJS comes from the NCI's cancer research training award. We thank Christopher B. Buck (NCI) for helpful comments and discussion.

The authors acknowledge the DFCI Oncology Data Retrieval System (OncDRS) for the aggregation, management, and delivery of the clinical and operational research data used in this project. The content is solely the responsibility of the authors.

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



## Figure Legends

**Figure 1**. **Somatic variants in Merkel cell carcinoma**. A. Tumor mutation burden (TMB) for each patient in descending order colored by mutation signature. B. Count of gene copy number alterations per patient. C. OncoPrint for the top 10 genes with the greatest number of point mutations in this MCC cohort. D. Distribution of point mutations in the cDNA of RB1 and TP53 from this MCC cohort. Functional domains of p53 and pRB are highlighted by colored boxes. Each type of base substitution is highlight by a different color lollipop and nonsense mutations are indicated by asterisks.

**Figure 2**. **Recurrent copy number variants in MCC**. A. Representative network analysis clusters of significantly co-modified genes in MCC on chromosomes 1 (red), 6 (yellow) and 10 (blue). B. Frequency of amplifications (red) and deletions (blue) for the genes comprising representative CNV clusters and their occurrence in each patient with UV, RB1, and TP53 status clustered by all variants. C. Counts of each CNV cluster colored by TMB-low (blue), -intermediate (grey), and -high (red) categories. Clusters that are nearly equivalent between TMB-low and -high (<2:1 ratio are highlighted by open triangles). The cluster that is more frequent in TMB-low than -high is highlighted by a black-filled triangle. D. Kaplan-Meier plot of overall survival stratified by chromosome 6 amplification for all patients. E. Kaplan-Meier plot of overall survival stratified by chromosome 6 amplification for primary tumors. F-G. Analysis of TCGA cancers for the two most abundant CNV clusters (13 & 14, respectively) in MCC.

**Figure 3**. **Detection of MCPyV via targeted capture and NGS**. A. Raw number of reads mapping to the MCPyV genome per patient from VB2 (n=48). B. Normalized count of MCPyV reads based on number of human reads and fraction of viral genome covered. C. Scatter plot of

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



genome coverage vs normalized MCPyV copies with virus-positive patients highlighted in red and virus-negative patients in black.

**Figure 4. MCPyV coverage and mutations from virus-positive cases**. Read coverage for MCPyV in gray and each plot represents a single patient with their ID in the upper left corner. Scales for the coverage plots are set from 0 to the maximum read coverage per patient. Point and insertion-deletion mutations are indicated by vertical lines located at the start point of the mutation colored by the type of base substitution. The effects of point mutations within LT antigen are indicated by a triangle (frameshift) or asterisk (stop gain) at the top of the vertical line of the mutation.

**Figure 5. Protein changes in Large and Small T antigens in MCC**. A. Lollipop plot of all LT missense mutations relative to the NC_010227.2 MCPyV reference with height reflecting the number of observations in our cohort and residue change labeled above the position. LT domains are highlighted by colored boxes. Below the LT diagram, MAFFT alignment of predicted LT sequences from all virus-positive cases colored by amino acids. B. Lollipop plot of all ST missense mutations relative to the NC_010227.2 MCPyV reference genome

**Figure 6**. **Characterization of MCPyV integration sites**. A. Location of integration events in the human genome labeled and colored by patient. B. Coverage of reads corresponding to predicted overlapping integration sites on chromosome 1. Direction of virus-to-host fusion is shown by black arrows. C-E. Representative assembly graphs for different types of viral integrations. Human DNA is a blue gradient and viral DNA is a red gradient representing different genomic segments. Human chromosome positions at the virus junctions are shown. Detailed assembly graphs for all virus-positive cases are in Figure S4. C. Representative single linear assembly graph for integrated MCPyV from case MCC001 on chromosome 3. D.

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Representative assembly graph of partially duplicated MCPyV genome integrated into the tumor genome of MCC025 on chromosome 1. Path for linearization of assembly graph shown by the dark grey line. E. Representative assembly graph of MCPyV genome integrated into chromosome 7 of MCC071 supporting a circular DNA intermediate diagrammed on the right. F. Barplot showing the frequency of microhomology lengths between two and seven basepairs. Expected values are in black and observed are in grey. Asterisks representing p-values from Fisher's exact test are represented above the bars (* <0.05, ** <0.01). G. Diagram of representative integration sites with viral sequence highlighted in yellow and host sequence in blue. Matching bases between host and virus are in red. H. Barplot showing the frequency of repetitive elements within 2 kb of integration sites. Expected values are in black and observed are in grey. P-values from Fisher's exact test are represented above the bars.

**Figure 7**. **Clinical outcome based on mutation signature, virus status, and immune suppression**. A. Pie charts representing the portion of patients that are virus-positive (VP, red) or virus-negative (VN, grey) and immunocompetent or immunosuppressed. B. Kaplan-Meier plot of overall survival of immunocompetent (black) and immunosuppressed (red) MCC patients.

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



## Tables

**Table 1. Patient Characteristics (N=71)**

| Characteristics | All (N=71) |
|---|---|
| **Age at initial diagnosis, yrs** | 70 (10-93) |
| **Age at initial diagnosis, yrs** | |
| <=70 | 36 (51%) |
| >70 | 35 (49%) |
| **Gender** | |
| Female | 31 (44%) |
| Male | 40 (56%) |
| **Race** | |
| Black or African American | 2 (3%) |
| White | 69 (97%) |
| **Initial site** | |
| Head | 19 (27%) |
| LE | 15 (21%) |
| Trunk | 23 (32%) |
| UE | 14 (20%) |
| **AJCC stage at initial diagnosis** | |
| I | 19 (27%) |
| II | 10 (14%) |
| III | 30 (42%) |
| IV | 12 (17%) |
| **Significant immunosuppression** | |
| No | 61 (86%) |
| Yes | 10 (14%) |
| **Prior chemotherapy or radiation** | |
| No | 53 (75%) |
| Yes | 18 (25%) |

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



**Table 2. MCPyV Integration sites**

| ID | Offtarget Reads | VB2 Reads | MCPyV Coverage | Normalized Coverage | Integration site(s) |
|---|---|---|---|---|---|
| **MCC001** | 0 | 7562 | 0.85 | 0.27197 | 3:181965781,181965770 |
| **MCC002** | 0 | 74 | 0.61 | 0.00013 | |
| **MCC003** | 4 | | | | |
| **MCC004** | 0 | | | | |
| **MCC005** | 4 | 87721 | 0.77 | 5.60786 | 5:20753360,33939328 |
| **MCC006** | 4 | 37150 | 0.77 | 2.50159 | 2:196945370,196945371 |
| **MCC007** | 0 | 971 | 0.97 | 0.00317 | |
| **MCC008** | 0 | 69634 | 1.00 | 1.73052 | 5:14961981,149709442 |
| **MCC009** | 0 | 3 | 0.05 | 0.00004 | |
| **MCC010** | 4 | 111147 | 1.00 | 5.37166 | 1:116791739,117025123 |
| **MCC011** | 0 | 212 | 0.87 | 0.00039 | |
| **MCC012** | 0 | | | | |
| **MCC013** | 78 | 113184 | 1.00 | 5.32306 | 6:36192882,36282634 |
| **MCC014** | 0 | 21307 | 0.70 | 1.45712 | 5:13820218,138511276 |
| **MCC015** | 0 | 177 | 0.77 | 0.00036 | |
| **MCC016** | 0 | | | | |
| **MCC017** | 0 | | | | |
| **MCC018** | 2 | | | | |
| **MCC019** | 0 | 13343 | 0.50 | 0.83000 | 10:63999700,64000021 |
| **MCC020** | 0 | 120 | 0.83 | 0.00029 | |
| **MCC021** | 0 | 8 | 0.07 | 0.00004 | |
| **MCC022** | 0 | 22715 | 0.93 | 0.72455 | 16:83581326,83890305 |
| **MCC023** | 0 | 1 | 0.02 | 0.00002 | |
| **MCC024** | 8 | 521049 | 1.00 | 24.81107 | |
| **MCC025** | 0 | | | | |
| **MCC026** | 0 | 48346 | 0.99 | 2.25473 | 9:76893837,77023700; 16:47914233,48036152; 18:1561377,1668866 |
| **MCC027** | 0 | 17748 | 0.36 | 1.84619 | 1:3582621,4107851 |
| **MCC028** | 0 | 0 | 0.00 | 0.00000 | |
| **MCC029** | 0 | 8808 | 0.53 | 0.51872 | 2:206984157,206984156 |
| **MCC030** | 0 | 102 | 0.77 | 0.00026 | |
| **MCC031** | 0 | | | | |
| **MCC032** | 14 | | | | |
| **MCC033** | 0 | | | | |
| **MCC034** | 0 | 119 | 0.76 | 0.00026 | |

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



| | | | | | |
|---|---|---|---|---|---|
| **MCC035** | 0 | 5 | 0.08 | 0.00002 | |
| **MCC036** | 0 | 21479 | 0.47 | 1.21591 | 15:57507670,57507677 |
| **MCC037** | 194 | 21095751 | 1.00 | ######## | |
| **MCC038** | 0 | | | | |
| **MCC039** | 0 | | | | |
| **MCC040** | 0 | 181807 | 0.96 | 7.46100 | 8:28408988,28457320 |
| **MCC041** | 0 | 12858 | 0.74 | 0.78054 | 9:11156335,111579165 |
| **MCC042** | 0 | 75177 | 1.00 | 1.99230 | 1:116797448,116797523 |
| **MCC043** | 1 | 31969 | 0.48 | 3.14034 | 9:13451094,13451103 |
| **MCC044** | 0 | 35415 | 1.00 | 1.39092 | 11:79113528,79113529 |
| **MCC045** | 4 | | | | |
| **MCC046** | 0 | 2 | 0.04 | 0.00003 | |
| **MCC047** | 0 | 6824 | 0.40 | 0.60372 | |
| **MCC048** | 0 | 2 | 0.04 | 0.00002 | |
| **MCC049** | 0 | 1 | 0.02 | 0.00003 | |
| **MCC050** | 0 | 34047 | 0.48 | 2.10949 | 6:51146411,51146421 |
| **MCC051** | 7 | | | | |
| **MCC052** | 8 | 74199 | 1.00 | 2.06071 | 8:113896842,114256794 |
| **MCC053** | 0 | | | | |
| **MCC054** | 0 | 103352 | 1.00 | 2.95039 | 5:8556313,34193826 (34349919-34349456) |
| **MCC055** | 0 | 265 | 0.96 | 0.00052 | |
| **MCC056** | 6 | 85232 | 0.80 | 5.65339 | 6:9659029,9659034 |
| **MCC057** | 0 | 7 | 0.09 | 0.00004 | |
| **MCC058** | 3 | | | | |
| **MCC059** | 0 | 1 | 0.02 | 0.00003 | |
| **MCC060** | 0 | | | | |
| **MCC061** | 7 | | | | |
| **MCC062** | 0 | 10735 | 0.38 | 0.66668 | 7:121478017,121478033 |
| **MCC063** | 0 | 4 | 0.07 | 0.00001 | |
| **MCC064** | 0 | | | | |
| **MCC065** | 0 | | | | |
| **MCC066** | 0 | | | | |
| **MCC067** | 10 | | | | |
| **MCC068** | 3 | | | | |
| **MCC069** | 0 | 25483 | 0.55 | 1.31799 | 1:76825442,76826185 |
| **MCC070** | 0 | 5 | 0.09 | 0.00001 | |
| **MCC071** | 0 | 19543 | 0.47 | 1.61800 | 7:1330002,1593035 |

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



**Table 3. Comparison of sequencing, PCR, and IHC for determination of tumor viral status**

| ID | TMB | TMB category | PCR # primer sets | IHC | VB2 | Virus-positive Evidence | Add'l VP Evidence | UV | VP or VN |
|---|---|---|---|---|---|---|---|---|---|
| **MCC001** | 8.5 | Intermediate | 2 | 1 | Positive | VB2 | PCR IHC | 0 | VP |
| **MCC002** | 30.4 | High | 0 | 0 | Negative | UV VB2 TM | | 1 | VN |
| **MCC003** | 0.0 | Low | | | | | | 0 | VP |
| **MCC004** | 6.6 | Intermediate | | | | | | 0 | VP |
| **MCC005** | 1.6 | Low | | 1 | Positive | VB2 | IHC | 0 | VP |
| **MCC006** | 3.7 | Low | | | Positive | VB2 | | 0 | VP |
| **MCC007** | 9.6 | Intermediate | 1 | 0 | Negative | VB2 | | 0 | VN |
| **MCC008** | 2.7 | Low | 5 | 1 | Positive | VB2 | PCR IHC | 0 | VP |
| **MCC009** | 28.2 | High | | 0 | Negative | UV VB2 TM | | 1 | VN |
| **MCC010** | 4.3 | Low | | 1 | Positive | | IHC | 0 | VP |
| **MCC011** | 11.7 | Intermediate | 0 | 1 | Negative | UV TM | | 1 | VN |
| **MCC012** | 0.0 | Low | | 0 | | | | 0 | VP |
| **MCC013** | 3.2 | Low | 5 | 1 | Positive | VB2 | PCR IHC | 0 | VP |
| **MCC014** | 3.7 | Low | | 1 | Positive | VB2 | | 0 | VP |
| **MCC015** | 27.7 | High | 0 | 0 | Negative | UV VB2 TM | | 1 | VN |
| **MCC016** | 23.0 | High | | | | UV TM | | 1 | VN |
| **MCC017** | 2.5 | Low | | | | | | 0 | VP |
| **MCC018** | 1.6 | Low | | | | | | 0 | VP |
| **MCC019** | 4.3 | Low | | | Positive | VB2 | | 0 | VP |
| **MCC020** | 27.7 | High | | 0 | Negative | UV TM | | 1 | VN |
| **MCC021** | 25.0 | High | | | Negative | UV VB2 TM | | 1 | VN |
| **MCC022** | 3.7 | Low | | | Positive | VB2 | | 0 | VP |
| **MCC023** | 19.2 | Intermediate | | | Negative | UV VB2 TM | | 1 | VN |
| **MCC024** | 3.2 | Low | | | Positive | VB2 | | 0 | VP |
| **MCC025** | 0.0 | Low | | | | | | 0 | VP |
| **MCC026** | 4.3 | Low | | | Positive | VB2 | | 0 | VP |
| **MCC027** | 2.7 | Low | 4 | 1 | Positive | VB2 | PCR IHC | 0 | VP |
| **MCC028** | 29.3 | High | | | Negative | UV VB2 TM | | 1 | VN |
| **MCC029** | 5.3 | Low | | 0 | | | | 0 | VP |
| **MCC030** | 38.9 | High | 0 | 0 | Negative | UV VB2 TM | | 1 | VN |
| **MCC031** | 1.6 | Low | | | | | | 0 | VP |
| **MCC032** | 1.1 | Low | | | | | | 0 | VP |

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **MCC033** | 5.3 | Low | | 1 | Positive | VB2 | IHC | 0 | VP |
| **MCC034** | 28.8 | High | 0 | 0 | Negative | UV VB2 TM | | 1 | VN |
| **MCC035** | 11.2 | Intermediate | | | Negative | UV VB2 TM | | 1 | VN |
| **MCC036** | 3.2 | Low | 4 | 1 | Positive | VB2 | PCR IHC | 0 | VP |
| **MCC037** | 4.8 | Low | | 1 | Positive | VB2 | IHC | 0 | VP |
| **MCC038** | 1.6 | Low | | | | | | 0 | VP |
| **MCC039** | 22.4 | High | | | | UV TM | | 1 | VN |
| **MCC040** | 3.7 | Low | | | Positive | VB2 | | 0 | VP |
| **MCC041** | 2.1 | Low | | | Positive | VB2 | | 0 | VP |
| **MCC042** | 2.1 | Low | | | Positive | VB2 | | 0 | VP |
| **MCC043** | 4.3 | Low | | | Positive | VB2 | | 0 | VP |
| **MCC044** | 5.3 | Low | 2 | 1 | Positive | VB2 | PCR IHC | 0 | VP |
| **MCC045** | 2.1 | Low | | 1 | Positive | VB2 | IHC | 0 | VP |
| **MCC046** | 20.8 | High | | | Negative | UV VB2 TM | | 1 | VN |
| **MCC047** | 2.1 | Low | | 1 | Positive | VB2 | IHC | 0 | VP |
| **MCC048** | 26.6 | High | | | Negative | UV VB2 TM | | 1 | VN |
| **MCC049** | 34.1 | High | | | Negative | UV VB2 TM | | 1 | VN |
| **MCC050** | 4.8 | Low | | | Positive | VB2 | | 0 | VP |
| **MCC051** | 3.2 | Low | | 1 | Positive | VB2 | IHC | 0 | VP |
| **MCC052** | 2.1 | Low | | | | | | 0 | VP |
| **MCC053** | 0.0 | Low | 3 | 0 | | | PCR | 0 | VP |
| **MCC054** | 3.7 | Low | 3 | 1 | Positive | VB2 | PCR IHC | 0 | VP |
| **MCC055** | 30.9 | High | | | Negative | UV VB2 TM | | 1 | VN |
| **MCC056** | 3.7 | Low | | | Positive | VB2 | | 0 | VP |
| **MCC057** | 29.8 | High | | | Negative | UV VB2 TM | | 1 | VN |
| **MCC058** | 1.1 | Low | | | | | | 0 | VP |
| **MCC059** | 19.2 | Intermediate | | | Negative | UV VB2 TM | | 1 | VN |
| **MCC060** | 1.1 | Low | | | | | | 0 | VP |
| **MCC061** | 2.1 | Low | | | | | | 0 | VP |
| **MCC062** | 4.8 | Low | 4 | 0 | Positive | VB2 | PCR | 0 | VP |
| **MCC063** | 27.2 | High | | | Negative | UV VB2 TM | | 1 | VN |
| **MCC064** | 25.6 | High | | | | UV TM | | 1 | VN |
| **MCC065** | 13.3 | Intermediate | | | | UV TM | | 1 | VN |
| **MCC066** | 18.6 | Intermediate | | | | UV TM | | 1 | VN |
| **MCC067** | 0.5 | Low | | | | | | 0 | VP |

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **MCC068** | 2.1 | Low | | | | | 0 | VP |
| **MCC069** | 4.3 | Low | 1 | Positive | VB2 | IHC | 0 | VP |
| **MCC070** | 29.8 | High | | Negative | UV VB2 TM | | 1 | VN |
| **MCC071** | 3.2 | Low | | | | | 0 | VP |

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



**Table 4. Association between Relapse and Genomic Sequencing (N=71)**

| Characteristics | All (N=71) | Relapse or not | | Fisher'e exact test p-value |
|---|---|---|---|---|
| | | No relapse (N=30) | Relapse (N=41) | |
| **UV** | | | | |
| Negative | 47 (66%) | 20 (67%) | 27 (66%) | >0.99 |
| Positive | 24 (34%) | 10 (33%) | 14 (34%) | |
| **pRB status** | | | | |
| Mutate | 32 (45%) | 13 (43%) | 19 (46%) | 0.81 |
| Wild type | 39 (55%) | 17 (57%) | 22 (54%) | |
| **p53 status** | | | | |
| Mutate | 31 (44%) | 13 (43%) | 18 (44%) | >0.99 |
| Wild type | 40 (56%) | 17 (57%) | 23 (56%) | |
| **Virus positive or negative** | | | | |
| VN | 25 (35%) | 10 (33%) | 15 (37%) | 0.81 |
| VP | 46 (65%) | 20 (67%) | 26 (63%) | |
| **pRB and p53** | | | | |
| pRB=M, p53=M | 24 (34%) | 10 (33%) | 14 (34%) | >0.99 |
| pRB=M, p53=W | 8 (11%) | 3 (10%) | 5 (12%) | |
| pRB=W, p53=M | 7 (10%) | 3 (10%) | 4 (10%) | |
| pRB=W, p53=W | 32 (45%) | 14 (47%) | 18 (44%) | |

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



**Table 5. Association between Patients Characteristics and Immunosuppression using Fisher's exact test**

| Characteristics | All (N=71) | Immunosuppression | | Fisher's exact test p-value |
|---|---|---|---|---|
| | | No (N=61) | Yes (N=10) | |
| **Gender** | | | | |
| Female | 31 (44%) | 29 (48%) | 2 (20%) | 0.17 |
| Male | 40 (56%) | 32 (52%) | 8 (80%) | |
| **Race** | | | | |
| Black or African American | 2 (3%) | 2 (3%) | 0 (0%) | >0.99 |
| White | 69 (97%) | 59 (97%) | 10 (100%) | |
| **Age at initial diagnosis, yrs** | | | | |
| <=70 | 36 (51%) | 32 (52%) | 4 (40%) | 0.51 |
| >70 | 35 (49%) | 29 (48%) | 6 (60%) | |
| **Initial site** | | | | |
| Head | 19 (27%) | 14 (23%) | 5 (50%) | 0.39 |
| LE | 15 (21%) | 14 (23%) | 1 (10%) | |
| Trunk | 23 (32%) | 21 (34%) | 2 (20%) | |
| UE | 14 (20%) | 12 (20%) | 2 (20%) | |
| **AJCC stage at initial diagnosis** | | | | |
| I | 19 (27%) | 16 (26%) | 3 (30%) | 0.96 |
| II | 10 (14%) | 9 (15%) | 1 (10%) | |
| III | 30 (42%) | 25 (41%) | 5 (50%) | |
| IV | 12 (17%) | 11 (18%) | 1 (10%) | |
| **Prior chemotherapy or radiation** | | | | |
| No | 53 (75%) | 46 (75%) | 7 (70%) | 0.71 |
| Yes | 18 (25%) | 15 (25%) | 3 (30%) | |
| **UV** | | | | |
| Negative | 47 (66%) | 45 (74%) | 2 (20%) | <0.01 |
| Positive | 24 (34%) | 16 (26%) | 8 (80%) | |
| **RB1 status** | | | | |
| Mutant | 32 (45%) | 24 (39%) | 8 (80%) | 0.04 |
| Wild type | 39 (55%) | 37 (61%) | 2 (20%) | |
| **TP53 status** | | | | |
| Mutant | 31 (44%) | 25 (41%) | 6 (60%) | 0.31 |
| Wild type | 40 (56%) | 36 (59%) | 4 (40%) | |
| **Virus positive or negative** | | | | |
| VN | 25 (35%) | 17 (28%) | 8 (80%) | <0.01 |
| VP | 46 (65%) | 44 (72%) | 2 (20%) | |

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



## References


1. Harms PW, Harms KL, Moore PS, DeCaprio JA, Nghiem P, Wong MKK, et al. The biology and treatment of Merkel cell carcinoma: current understanding and research priorities. *Nat Rev Clin Oncol.* 2018;15(12):763-76.
2. Feng H, Shuda M, Chang Y, and Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science.* 2008;319(5866):1096-100.
3. Tolstov YL, Pastrana DV, Feng H, Becker JC, Jenkins FJ, Moschos S, et al. Human Merkel cell polyomavirus infection II. MCV is a common human infection that can be detected by conformational capsid epitope immunoassays. *Int J Cancer.* 2009;125(6):1250-6.
4. Chen T, Hedman L, Mattila PS, Jartti T, Ruuskanen O, Soderlund-Venermo M, et al. Serological evidence of Merkel cell polyomavirus primary infections in childhood. *J Clin Virol.* 2011;50(2):125-9.
5. Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, and Buck CB. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. *Cell Host Microbe.* 2010;7(6):509-15.
6. Harms PW, Vats P, Verhaegen ME, Robinson DR, Wu YM, Dhanasekaran SM, et al. The Distinctive Mutational Spectra of Polyomavirus-Negative Merkel Cell Carcinoma. *Cancer Res.* 2015;75(18):3720-7.
7. Wong SQ, Waldeck K, Vergara IA, Schroder J, Madore J, Wilmott JS, et al. UV-Associated Mutations Underlie the Etiology of MCV-Negative Merkel Cell Carcinomas. *Cancer Res.* 2015;75(24):5228-34.
8. Starrett GJ, Marcelus C, Cantalupo PG, Katz JP, Cheng J, Akagi K, et al. Merkel Cell Polyomavirus Exhibits Dominant Control of the Tumor Genome and Transcriptome in Virus-Associated Merkel Cell Carcinoma. *MBio.* 2017;8(1).
9. Kuwamoto S, Higaki H, Kanai K, Iwasaki T, Sano H, Nagata K, et al. Association of Merkel cell polyomavirus infection with morphologic differences in Merkel cell carcinoma. *Hum Pathol.* 2011;42(5):632-40.
10. Schrama D, Peitsch WK, Zapatka M, Kneitz H, Houben R, Eib S, et al. Merkel cell polyomavirus status is not associated with clinical course of Merkel cell carcinoma. *J Invest Dermatol.* 2011;131(8):1631-8.
11. Iwasaki T, Matsushita M, Kuwamoto S, Kato M, Murakami I, Higaki-Mori H, et al. Usefulness of significant morphologic characteristics in distinguishing between Merkel cell polyomavirus-positive and Merkel cell polyomavirus-negative Merkel cell carcinomas. *Hum Pathol.* 2013;44(9):1912-7.
12. Lipson EJ, Vincent JG, Loyo M, Kagohara LT, Luber BS, Wang H, et al. PD-L1 expression in the Merkel cell carcinoma microenvironment: association with inflammation, Merkel cell polyomavirus and overall survival. *Cancer immunology research.* 2013;1(1):54-63.
13. Leroux-Kozal V, Leveque N, Brodard V, Lesage C, Dudez O, Makeieff M, et al. Merkel cell carcinoma: histopathologic and prognostic features according to the immunohistochemical expression of Merkel cell polyomavirus large T antigen correlated with viral load. *Hum Pathol.* 2015;46(3):443-53.
14. Nghiem PT, Bhatia S, Lipson EJ, Kudchadkar RR, Miller NJ, Annamalai L, et al. PD-1 Blockade with Pembrolizumab in Advanced Merkel-Cell Carcinoma. *N Engl J Med.* 2016;374(26):2542-52.

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



15. Kaufman HL, Russell J, Hamid O, Bhatia S, Terheyden P, D'Angelo SP, et al. Avelumab in patients with chemotherapy-refractory metastatic Merkel cell carcinoma: a multicentre, single-group, open-label, phase 2 trial. *The Lancet Oncology.* 2016;17(10):1374-85.
16. Shuda M, Arora R, Kwun HJ, Feng H, Sarid R, Fernandez-Figueras MT, et al. Human Merkel cell polyomavirus infection I. MCV T antigen expression in Merkel cell carcinoma, lymphoid tissues and lymphoid tumors. *Int J Cancer.* 2009;125(6):1243-9.
17. Rodig SJ, Cheng J, Wardzala J, DoRosario A, Scanlon JJ, Laga AC, et al. Improved detection suggests all Merkel cell carcinomas harbor Merkel polyomavirus. *J Clin Invest.* 2012;122(12):4645-53.
18. Moshiri AS, Doumani R, Yelistratova L, Blom A, Lachance K, Shinohara MM, et al. Polyomavirus-Negative Merkel Cell Carcinoma: A More Aggressive Subtype Based on Analysis of 282 Cases Using Multimodal Tumor Virus Detection. *J Invest Dermatol.* 2017;137(4):819-27.
19. Wang L, Harms PW, Palanisamy N, Carskadon S, Cao X, Siddiqui J, et al. Age and Gender Associations of Virus Positivity in Merkel Cell Carcinoma Characterized Using a Novel RNA In Situ Hybridization Assay. *Clin Cancer Res.* 2017.
20. Paulson KG, Lewis CW, Redman MW, Simonson WT, Lisberg A, Ritter D, et al. Viral oncoprotein antibodies as a marker for recurrence of Merkel cell carcinoma: A prospective validation study. *Cancer.* 2017;123(8):1464-74.
21. Goh G, Walradt T, Markarov V, Blom A, Riaz N, Doumani R, et al. Mutational landscape of MCPyV-positive and MCPyV-negative Merkel cell carcinomas with implications for immunotherapy. *Oncotarget.* 2016;7(3):3403-15.
22. Gonzalez-Vela MD, Curiel-Olmo S, Derdak S, Beltran S, Santibanez M, Martinez N, et al. Shared Oncogenic Pathways Implicated in Both Virus-Positive and UV-Induced Merkel Cell Carcinomas. *J Invest Dermatol.* 2017;137(1):197-206.
23. Sholl LM, Do K, Shivdasani P, Cerami E, Dubuc AM, Kuo FC, et al. Institutional implementation of clinical tumor profiling on an unselected cancer population. *JCI insight.* 2016;1(19):e87062.
24. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature.* 2013;500(7463):415-21.
25. Van Gele M, Leonard JH, Van Roy N, Cook AL, De Paepe A, and Speleman F. Frequent allelic loss at 10q23 but low incidence of PTEN mutations in Merkel cell carcinoma. *Int J Cancer.* 2001;92(3):409-13.
26. Cohen PR, and Kurzrock R. Merkel Cell Carcinoma with a Suppressor of Fused (SUFU) Mutation: Case Report and Potential Therapeutic Implications. *Dermatology and therapy.* 2015;5(2):129-43.
27. Park DE, Cheng J, Berrios C, Montero J, Cortes-Cros M, Ferretti S, et al. Dual inhibition of MDM2 and MDM4 in virus-positive Merkel cell carcinoma enhances the p53 response. *Proc Natl Acad Sci U S A.* 2019;116(3):1027-32.
28. Nomura K, Klejnot M, Kowalczyk D, Hock AK, Sibbet GJ, Vousden KH, et al. Structural analysis of MDM2 RING separates degradation from regulation of p53 transcription activity. *Nat Struct Mol Biol.* 2017;24(7):578-87.
29. Paulson KG, Lemos BD, Feng B, Jaimes N, Penas PF, Bi X, et al. Array-CGH reveals recurrent genomic changes in Merkel cell carcinoma including amplification of L-Myc. *J Invest Dermatol.* 2009;129(6):1547-55.
30. Hu Z, Zhu D, Wang W, Li W, Jia W, Zeng X, et al. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat Genet.* 2015;47(2):158-63.
31. Nambudiri VE, Vivero M, Watson AJ, Thakuria M, Ng A, Russell S, et al. Merkel Cell Carcinoma Presenting as Subcutaneous Breast Masses: An Uncommon Presentation of a Rare Neuroendocrine Neoplasm. *Breast J.* 2016;22(1):113-5.

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.




32. Rao P, Balzer BL, Lemos BD, Liegeois NJ, McNiff JM, Nghiem P, et al. Protocol for the examination of specimens from patients with merkel cell carcinoma of the skin. *Arch Pathol Lab Med.* 2010;134(3):341-4.
33. Crall C, Morley KW, Rabinowits G, Schmidt B, Broyles AD, and Huang JT. Merkel cell carcinoma in a patient with GATA2 deficiency: a novel association with primary immunodeficiency. *Br J Dermatol.* 2015.
34. Schrama D, Sarosi EM, Adam C, Ritter C, Kaemmerer U, Klopocki E, et al. Characterization of six Merkel cell polyomavirus-positive Merkel cell carcinoma cell lines: integration pattern suggest that Large T antigen truncating events occur before or during integration. *Int J Cancer.* 2019.
35. Gillison ML, Akagi K, Xiao W, Jiang B, Pickard RKL, Li J, et al. Human papillomavirus and the landscape of secondary genetic alterations in oral cancers. *Genome Res.* 2018.
36. Santos GC, Zielenska M, Prasad M, and Squire JA. Chromosome 6p amplification and cancer progression. *J Clin Pathol.* 2007;60(1):1-7.
37. Pitale M, Sessions RB, and Husain S. An analysis of prognostic factors in cutaneous neuroendocrine carcinoma. *Laryngoscope.* 1992;102(3):244-9.
38. Gooptu C, Woollons A, Ross J, Price M, Wojnarowska F, Morris PJ, et al. Merkel cell carcinoma arising after therapeutic immunosuppression. *Br J Dermatol.* 1997;137(4):637-41.
39. Lanoy E, and Engels EA. Skin cancers associated with autoimmune conditions among elderly adults. *Br J Cancer.* 2010;103(1):112-4.
40. Clarke CA, Robbins HA, Tatalovich Z, Lynch CF, Pawlish KS, Finch JL, et al. Risk of merkel cell carcinoma after solid organ transplantation. *J Natl Cancer Inst.* 2015;107(2).
41. Sahi H, Sihto H, Artama M, Koljonen V, Bohling T, and Pukkala E. History of chronic inflammatory disorders increases the risk of Merkel cell carcinoma, but does not correlate with Merkel cell polyomavirus infection. *Br J Cancer.* 2017;116(2):260-4.
42. Garrett GL, Blanc PD, Boscardin J, Lloyd AA, Ahmed RL, Anthony T, et al. Incidence of and Risk Factors for Skin Cancer in Organ Transplant Recipients in the United States. *JAMA dermatology.* 2017.
43. Wheless L, Jacks S, Mooneyham Potter KA, Leach BC, and Cook J. Skin cancer in organ transplant recipients: more than the immune system. *J Am Acad Dermatol.* 2014;71(2):359-65.
44. Scott FI, Mamtani R, Brensinger CM, Haynes K, Chiesa-Fuxench ZC, Zhang J, et al. Risk of Nonmelanoma Skin Cancer Associated With the Use of Immunosuppressant and Biologic Agents in Patients With a History of Autoimmune Disease and Nonmelanoma Skin Cancer. *JAMA dermatology.* 2016;152(2):164-72.
45. Busam KJ, Jungbluth AA, Rekthman N, Coit D, Pulitzer M, Bini J, et al. Merkel cell polyomavirus expression in merkel cell carcinomas and its absence in combined tumors and pulmonary neuroendocrine carcinomas. *Am J Surg Pathol.* 2009;33(9):1378-85.
46. Mogha A, Fautrel A, Mouchet N, Guo N, Corre S, Adamski H, et al. Merkel cell polyomavirus small T antigen mRNA level is increased following in vivo UV-radiation. *PLoS One.* 2010;5(7):e11423.
47. Dowlatshahi M, Huang V, Gehad AE, Jiang Y, Calarese A, Teague JE, et al. Tumor-specific T cells in human Merkel cell carcinomas: a possible role for Tregs and T-cell exhaustion in reducing T-cell responses. *J Invest Dermatol.* 2013;133(7):1879-89.
48. Li H, and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-60.
49. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491-8.

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



50. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-303.
51. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol.* 2013;31(3):213-9.
52. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, and Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics.* 2010;26(16):2069-70.
53. Talevich E, Shain AH, Botton T, and Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol.* 2016;12(4):e1004873.
54. Bloom BH. Space/Time Trade-offs in Hash Coding with Allowable Errors. *Comm ACM.* 1970;13(7):422-6.
55. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455-77.
56. Quinlan AR, and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841-2.

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

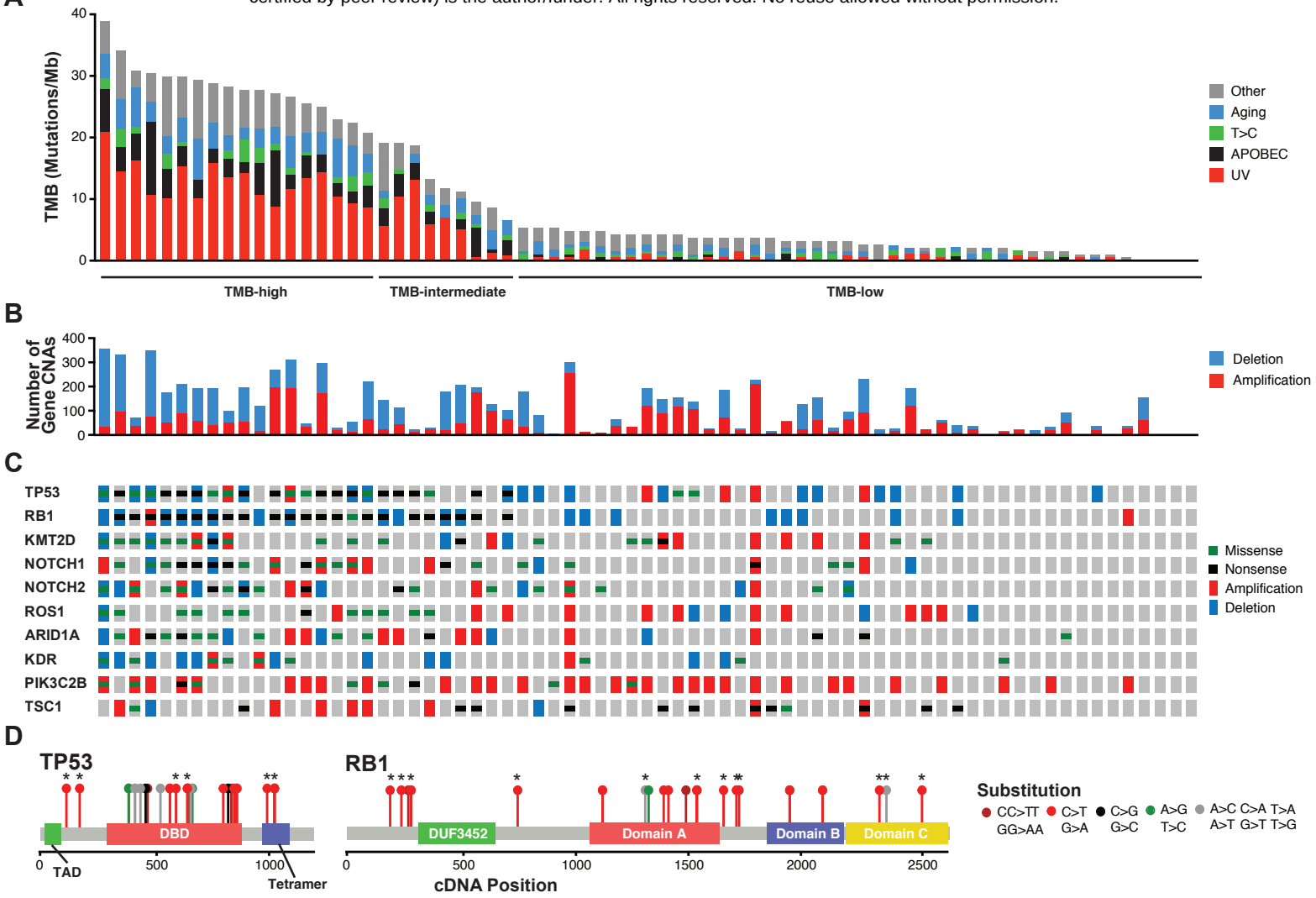


Figure 1

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

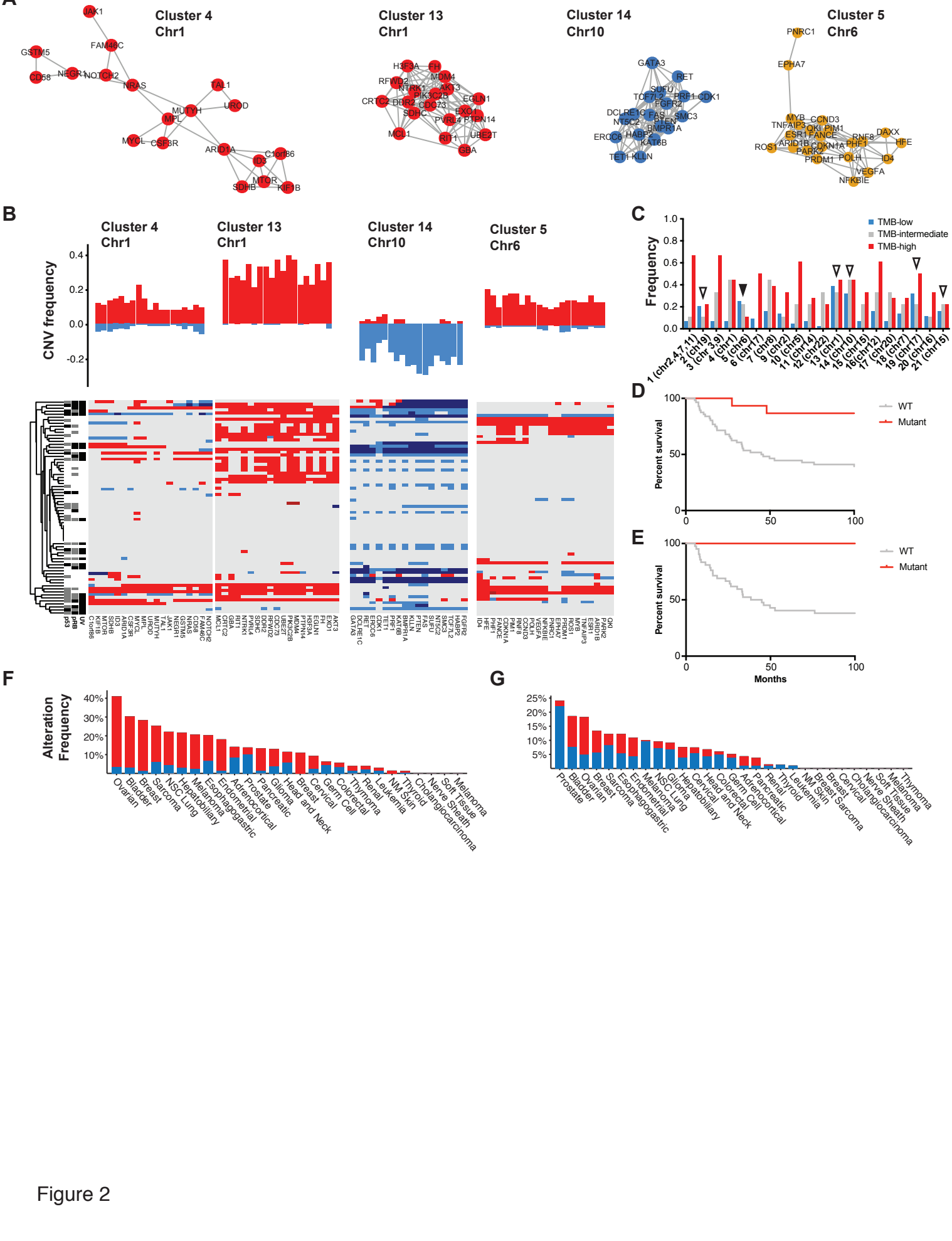


Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

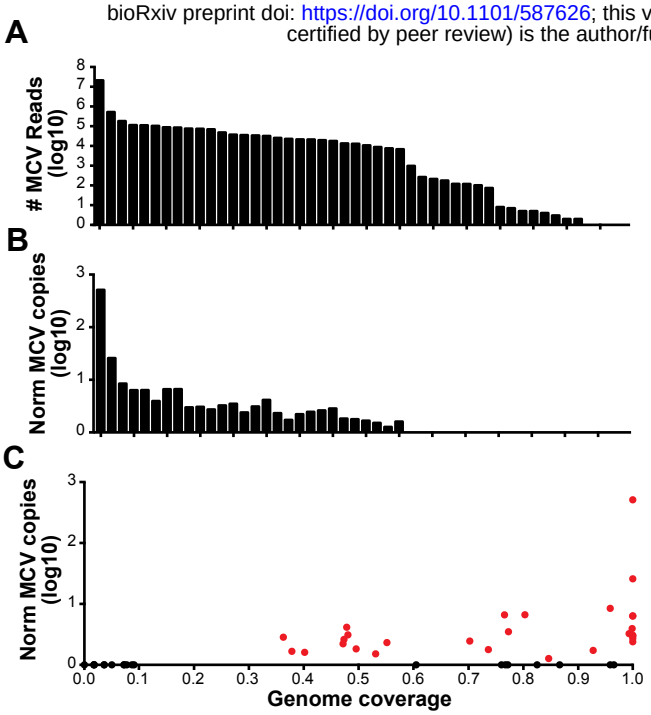


Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

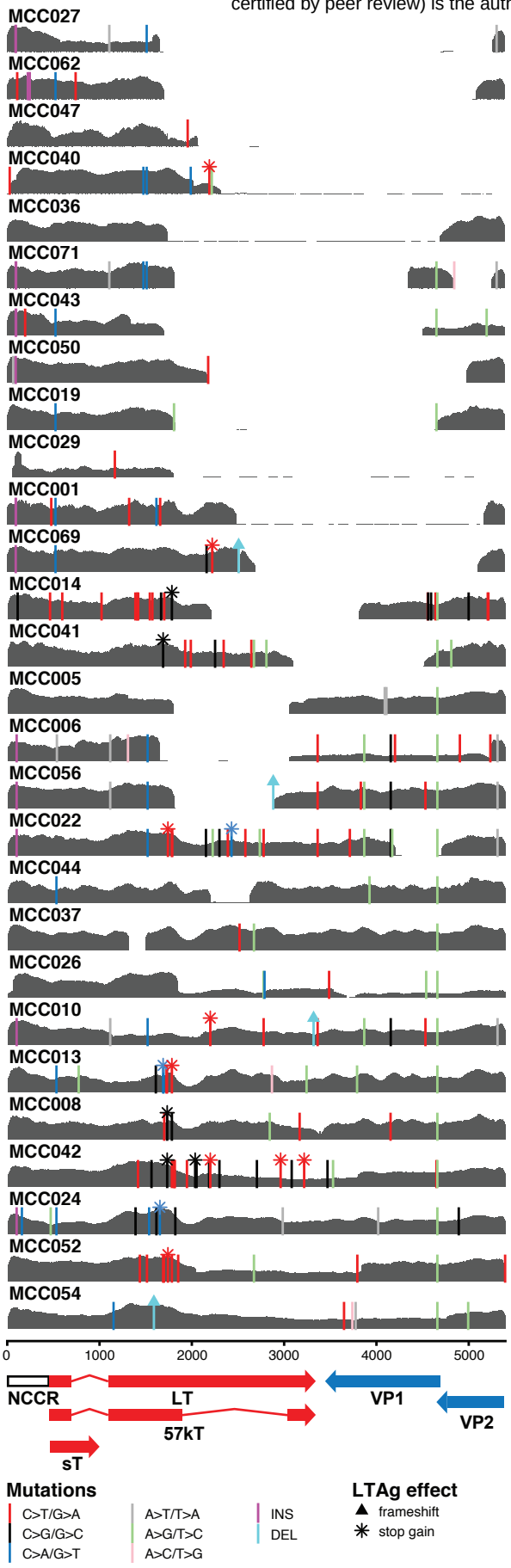


Figure 4

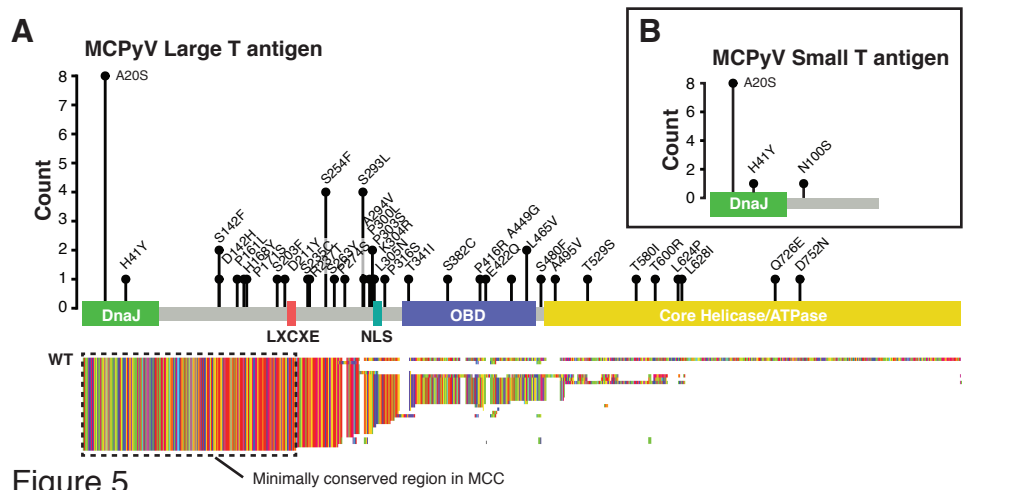
A
MCPyV Large T antigen
Count
A20S
H41Y
S142F
D142H
P161L
H168Y
P171S
S203F
D211Y
S254F
S293L
A294V
P300L
P303S
K304R
L305N
P316S
T341I
S382C
P416R
E422Q
A449G
L465V
S480F
A495V
T529S
T580I
T600R
L624P
L628I
Q726E
D752N
DnaJ
LXCXE
NLS
OBD
Core Helicase/ATPase
WT
Minimally conserved region in MCC
B
MCPyV Small T antigen
Count
A20S
H41Y
N100S
DnaJ

Figure 5

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

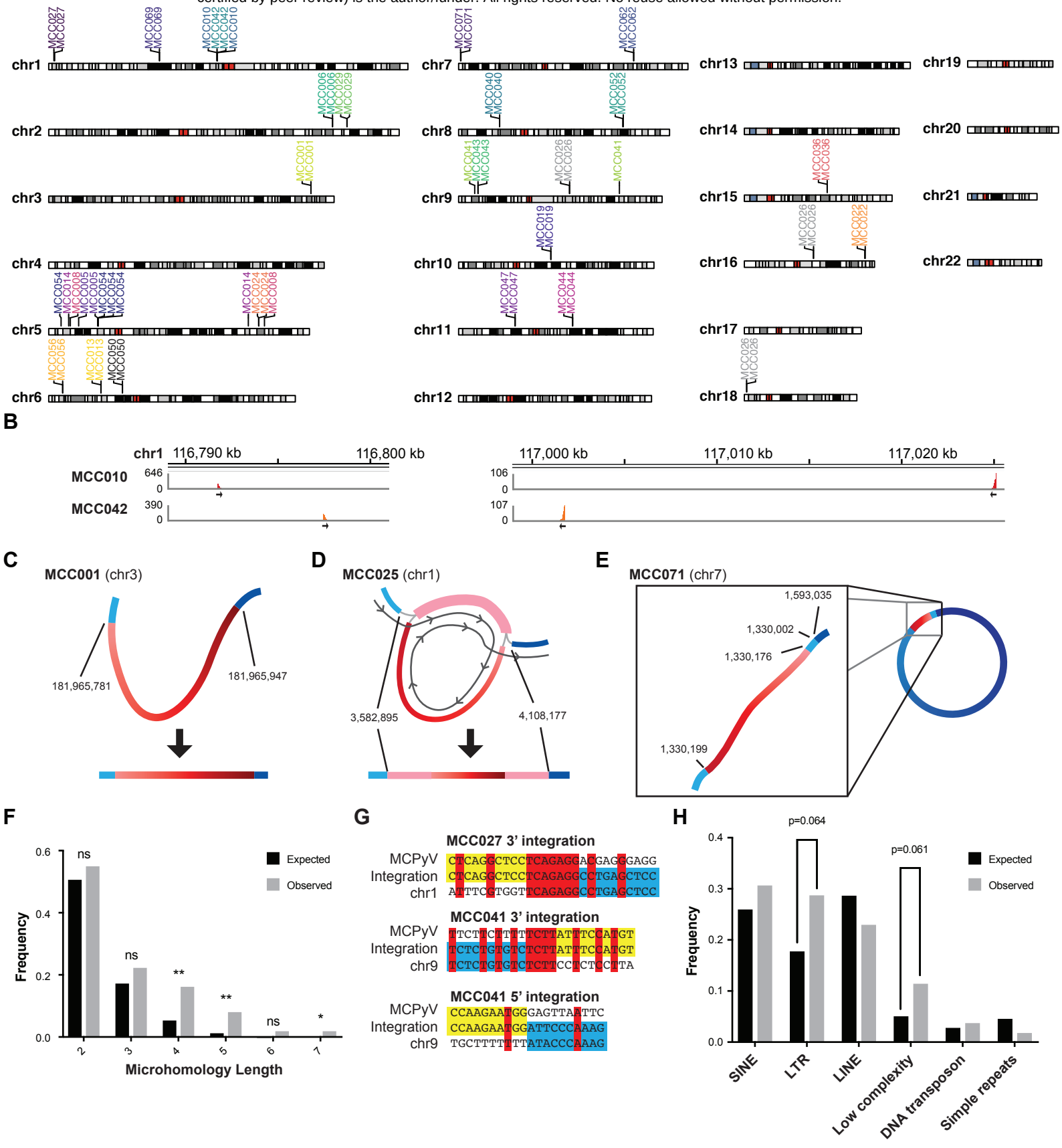


Figure 6

bioRxiv preprint doi: https://doi.org/10.1101/587626; this version posted March 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

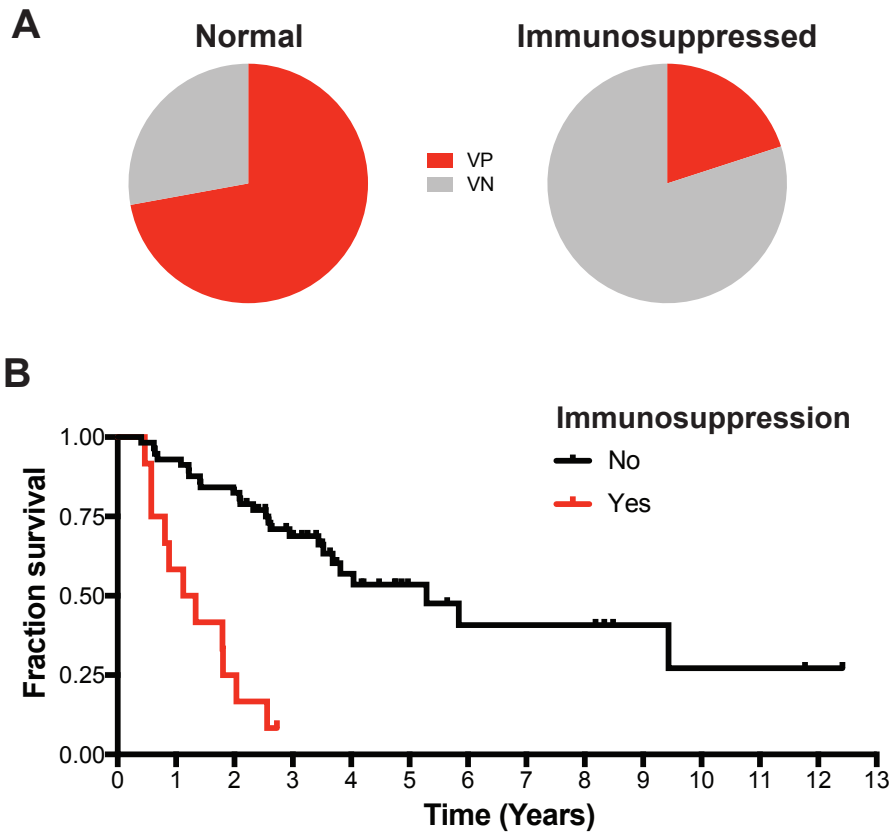


Figure 7