

RESEARCH

Open Access

# Clinical and molecular characterization of virus-positive and virus-negative Merkel cell carcinoma



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

## Abstract

**Background:** Merkel cell carcinoma (MCC) is a highly aggressive neuroendocrine carcinoma of the skin caused by either 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. However, it has been challenging to reliably distinguish between virus positive and UV damaged MCC.

**Methods:** In this study, we assembled a cohort of 71 MCC patients and performed deep sequencing with OncoPanel, a clinically implemented, 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 and software to detect integration sites and structure.

**Results:** Sequencing from this approach revealed distinct integration junctions in the tumor genome and generated assemblies that strongly support a model of microhomology-initiated hybrid, virus-host, circular DNA intermediate that promotes focal amplification of host and viral 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. Finally, comparing the molecular and clinical data from these patients revealed a surprising association of immunosuppression with virus-negative MCC and significantly shortened overall survival.

**Conclusions:** These results demonstrate the value of high-confidence virus detection for identifying molecular mechanisms of UV and viral oncogenesis in MCC. Furthermore, integrating these data with clinical data revealed features that could impact patient outcome and improve our understanding of MCC risk factors.

**Keywords:** Cancer genomics, Polyomavirus, Integration, Somatic variants, Mutagenesis

\* Correspondence: [James\\_DeCaprio@dfci.harvard.edu](mailto:James_DeCaprio@dfci.harvard.edu)

<sup>3</sup>Merkel Cell Carcinoma Center of Excellence, Dana-Farber/Brigham Cancer Center, Boston, MA, USA

<sup>5</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, Mayer 440, 450 Brookline Avenue, Boston, MA 02215, USA

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

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 primary 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 persistent 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, UV-associated, MCC [1]. Virus-positive MCC expresses the viral oncogenes large T antigen (LT) and small T antigen (ST) and 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 insight into the oncogenesis, cell-of-origin, and therapeutic options. Currently, there is no routine clinical effort to distinguish between virus-positive MCC and virus-negative MCC. Several recent studies have suggested 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 monoclonal antibodies CM2B4 and Ab3 or by 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 two of the three tests. The remaining 53 (18.8%) were positive for one test or none. This study assigned the MCC to be virus-positive if two or three tests were positive, implying that 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-ISH 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  $\geq 0.004/\text{cell}$  [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 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 neoantigen burden in virus-negative MCC corresponds to 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 for routine use in a clinical setting. This approach was combined with targeted sequencing of several hundred cancer-related genes to assess oncogenic changes in the tumor genome. Lastly, we compared the sensitivity and accuracy of this viral hybrid capture approach to more traditional approaches, PCR detection of viral DNA, IHC

for detection of MCPyV LT, and synoptic assessment of MCC pathology.

## Methods

### Study design and participants

This study included all patients ( $n = 71$ ) with a reported diagnosis of MCC at Dana-Farber/Brigham and Women's Cancer Center who underwent comprehensive genomic profiling by OncoPanel between May 2013 and April 2018. OncoPanel version 3 (POPv3) is a custom hybrid capture assay targeting the exons of 447 genes and 191 regions across 60 genes commonly rearranged in cancer [23, 24]. A retrospective chart review collected demographic, clinical, disease, treatment, and outcome variables on all 71 patients. For 40 patients, sufficient DNA remained from the initial OncoPanel profiling or from additional FFPE tumor specimens to perform POPv3/ViroPanel. When available, FFPE sections were sectioned for immunohistochemistry with antibodies CM2B4 and Ab3 [17]. Sections stained with hematoxylin and eosin were evaluated by synoptic review [25].

### Nucleic acid isolation, library preparation and sequencing

To perform ViroPanel with and without supplementation with the OncoPanel (v3) bait set, purified DNA was quantified using a Quant-iT PicoGreen dsDNA assay (Thermo Fisher). Library construction was performed using 200 ng of DNA, which was first fragmented to  $\sim 250$  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 (Thermo Fisher). 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 ( $3 \times 17$ -plex and  $1 \times 18$ -plex) to a total of 750 ng. Captures were done using the SureSelect<sup>XT</sup> Fast target enrichment assay (Agilent, Technologies, Santa Clara, CA) with ViroPanel 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).

### 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>) [26]. 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 [27, 28].

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) [29, 30]. 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 two 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 two 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 [31].

### Recurrent copy number analysis

Copy number variant calling was performed using a combination of VisCap Cancer and CNVkit as previously described [32, 33]. 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 five 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. Copy number variants from TCGA were retrieved from cBioPortal (<https://www.cbioportal.org/>) and plotted using ggplot2 in the R statistical environment [34, 35].

### 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 31 bp k-mers of the MCPyV genome [36]. 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. Skewness 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 skewness 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, 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 [37]. The assembly graphs from SPAdes were annotated using blastn against hg19 and the MCPyV reference genome with an e-value cutoff of  $1 \times 10^{-10}$ . Annotated assembly graphs were visualized using the ggraph R package.

Integrations sites confirmed by reference guided alignment and assembly data were analyzed for stretches of microhomology between the human and viral genomes by selecting 10 bp 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 to 25 bp inserted between viral and human DNA were excluded from analysis. Expected microhomology was calculated by randomly selecting 1000 20 bp 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 [38]. Expected frequency of integration near repeat elements was determined by randomly selecting 1000 sites in the human genome. Sites within 2 kb of a repeat element were counted as close proximity.

Functional annotation of somatic mutations and viral integration events was performed using PANTHER ([www.pantherdb.org](http://www.pantherdb.org)) [39].

### Statistics

The association between relapse and genomic characteristics are tested with Fisher's exact test using all patient sequencing data regardless of primary or recurrence biopsy. Overall survival (OS) is defined as the time from initial diagnosis to death, and patients who did not die are censored at the last follow-up date. The 95% confidence intervals of the median OS times are estimated using  $\log(-\log(\text{OS}))$  methodology. 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 the 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.

## 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

**Table 1** Patient characteristics (N = 71)

Characteristics	All (N = 71)
Age at initial diagnosis, years	70 (10–93)
Age at initial diagnosis, years	
< =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%)

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

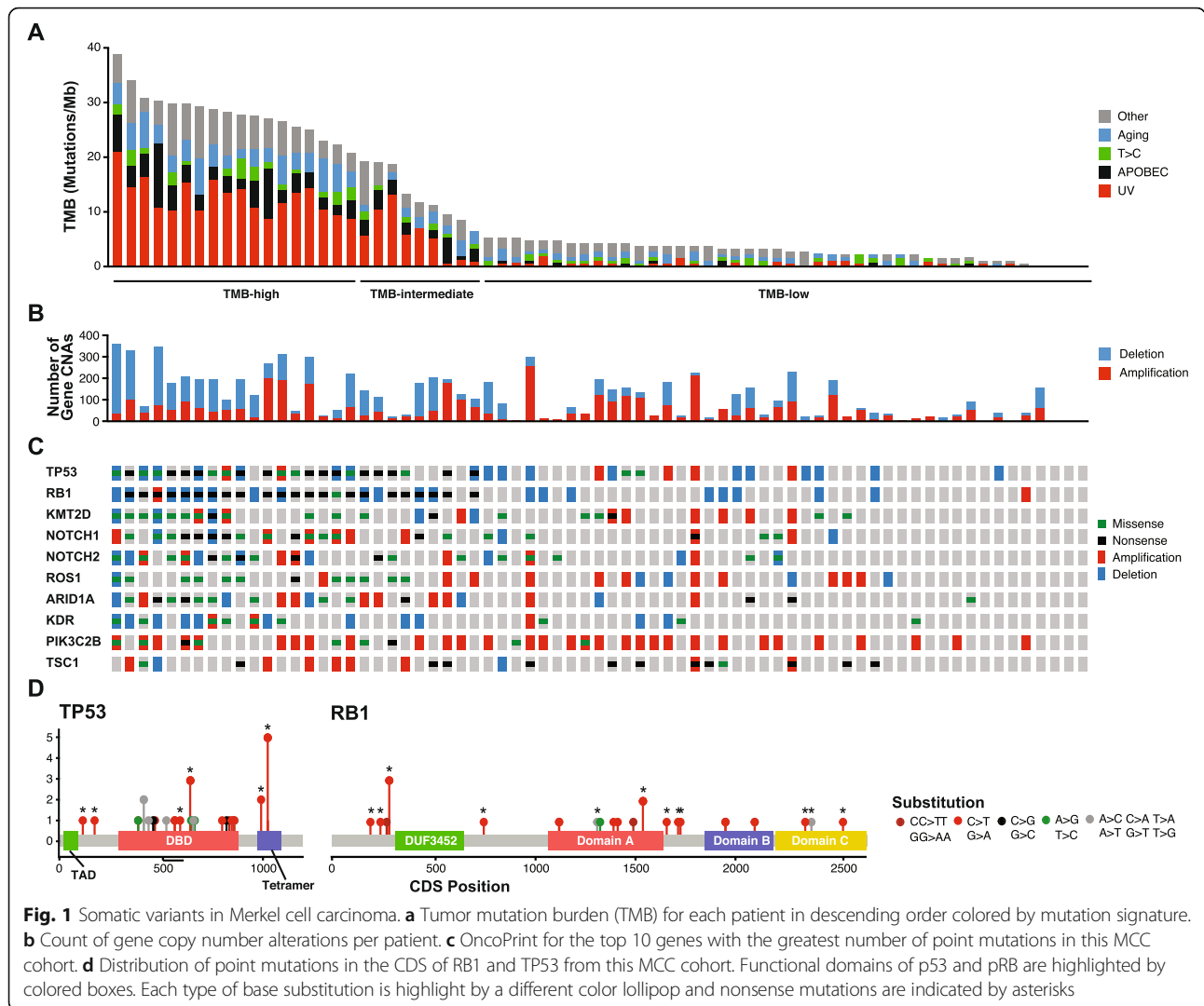
All 71 patients underwent OncoPanel analysis [32]. Genomic studies were performed using DNA isolated from tumors obtained at the time of initial diagnosis (n=50) or upon relapse (n=21). The total number of mutations ranged from 0 to 73 corresponding to a tumor mutational burden (TMB; mutations/megabase) from 0 to 38.89 with four cases containing no detectable mutations (Fig. 1a, Additional file 1: Table S1). From this mutation data, patients were binned into TMB-high ( $\geq 20$ ), TMB-intermediate ( $> 6 < 20$ ), and TMB-low ( $\leq 6$ ). 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 [40]. Additional mutational signatures identified included Aging (Signature 1; 3 cases), APOBEC (Signatures 2 and 13; 4

cases with 3 that also had an UV signature), and Signature 5 (one case) (Fig. 1a, Additional file 1: Table S1). TMB had some correlation with the number of copy number altered genes (Fig. 1b). Several genes including *RBI*, *TP53*, *KMT2D*, *NOTCH1*, *NOTCH2*, and *FAT1* were highly enriched for missense and truncating mutations (Fig. 1c, Additional file 2: Fig. S1). Single and dinucleotide substitutions in *RBI* and *TP53* revealed that most were likely mediated by UV damage (CC > TT, C > T; Fig. 1d).

Copy number variants (CNVs) were examined individually as well as against each other and other likely functional somatic changes for significant co-occurrence or mutual exclusivity (Additional file 1: Table S2). Clusters of significantly co-occurrent CNVs were determined via network analysis (Fig. 2a, Additional file 2: Fig. S2 & Fig. S3). From these analyses, two distinct CNV clusters were each found to be altered in more than 36% of cases (Fig. 2b, c). Chromosome 10 (cluster 14) had frequent copy number loss with 26 tumors showing heterozygous or homozygous loss of the chromosome (Fig. 2b) [41]. 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 [41, 42]. 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 (Fig. 2b) [43, 44]. In addition, we observed a focal amplification of *MYCL* within a greater amplification of Chr1p (cluster 4), which was reported in an earlier study of MCC [45].

CNV clusters 13 and 14 were observed at nearly equal frequencies in both TMB-high and TMB-low cases (Fig. 2b, c). Six other CNV clusters were strongly associated with UV signature and high TMB (Fig. 2c). Functional annotation of the clusters revealed that the two largest UV-associated CNV clusters (1 and 3) had significant enrichment for genes related broadly to DNA damage response and S-phase DNA damage checkpoint likely enhancing tolerance for UV mutagenesis.

Cluster 5, corresponding to 6p22.3 to 6q26 and likely representing a gain of the entire chromosome 6, was the only cluster more than twice as frequent in TMB-low tumors than TMB-high tumors (Fig. 2c). Interestingly, 33.3% (6/18) 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 (25%, 11/44) than TMB-high (11%, 2/18) 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 ( $p = 0.005$ ). Restricting this analysis to only primary tumors, revealed that there were no deaths at the time of this study in patients carrying this amplification ( $p = 0.007$ )



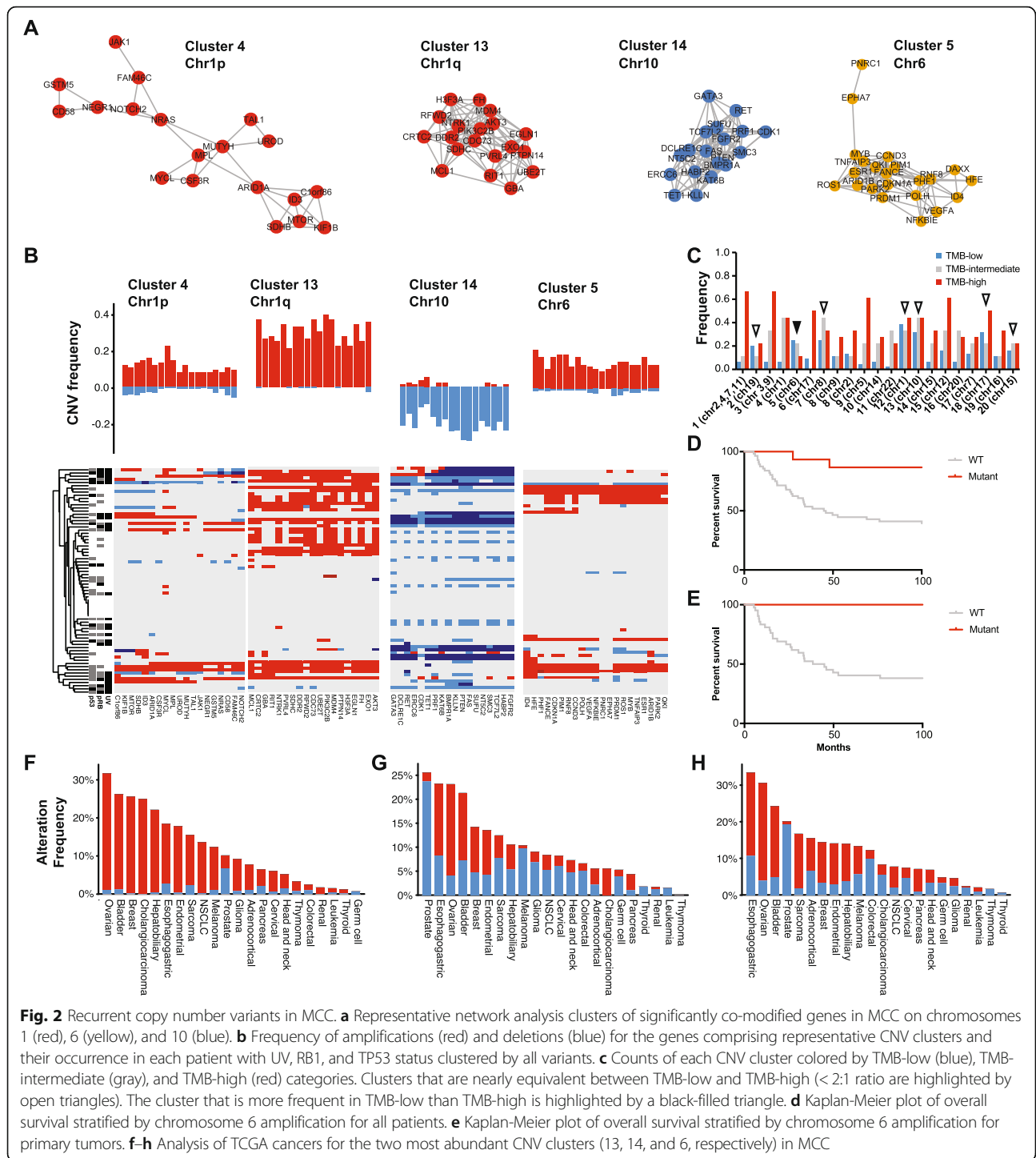
(Fig. 2d, e). Unsurprisingly, considering the prevalence of this CNV event in metastatic patients, there was no difference in recurrence-free survival (RFS, Additional file 2: Fig. S4).

The recurrent copy number events on chromosomes 1, 6, and 10 were compared within The Cancer Genome Atlas (TCGA) for similarities to other tumor types (Fig. 2f–h). This analysis revealed that the chromosome 1 (cluster 13) amplification was also frequently observed in ovarian, breast, and bladder cancers, whereas the chromosome 10 (cluster 14) loss was most frequently seen in prostate cancer. Gain of chromosome 6 (cluster 5) was most frequently seen in ovarian, bladder, and esophago-gastric cancers.

**Analysis of viral sequences in tumors**

Of the 71 tumors analyzed by OncoPanel, 48 with sufficient remaining material were re-analyzed by OncoPanel (Profile/OncoPanel version 3, POPv3) combined with a

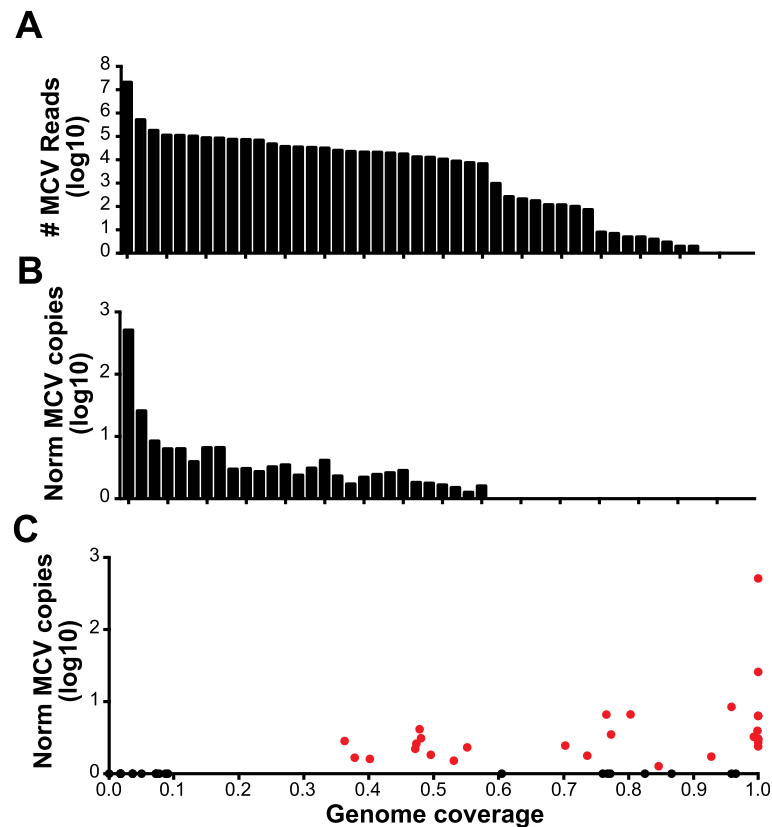
hybrid-capture probe bait set targeting the entire genome of MCPyV and other known oncogenic viruses (ViroPanel). For the 48 cases, the number of MCPyV reads ranged from 0 to 21,095,751 with only a single case having zero MCPyV reads (Fig. 3a). In total, 28 cases had substantial reads (> 6800) mapping to the MCPyV genome that also supported integration of the virus into the host genome through reads and read pairs that span integration junctions. For the remaining 20 cases without evidence of integration, the number of viral reads ranged from 0 to 971. Generally, these cases had reads that covered less than 10% of the viral genome with the normalized coverage less than two logs compared to samples with evidence for virus integration (Fig. 3b, c). 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 6824 to 21,095,751 (median 28,726). Consistent with previous reports, the integrated viral genome had undergone extensive mutagenesis



**Fig. 3** Detection of MCPyV via targeted capture and NGS. **a** Raw number of reads mapping to the MCPyV genome per patient from ViroPanel ( $n = 48$ ). **b** Normalized count of MCPyV reads based on number of human reads and fraction of viral genome covered. **c** Scatter plot of genome coverage vs normalized MCPyV copies with virus-positive patients highlighted in red and virus-negative patients in black

with large deletions (> 100 bp) particularly in the 3' half of LT as well as in the viral coat protein genes VP1 and VP2 (Fig. 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 (Figs. 3c and 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 (Fig. 5a) similar to what has previously been seen in MCC cell lines and clinical cohorts. In a single case (MCC054), LT was truncated by a 5-bp deletion resulting in a frame shift that introduced 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.

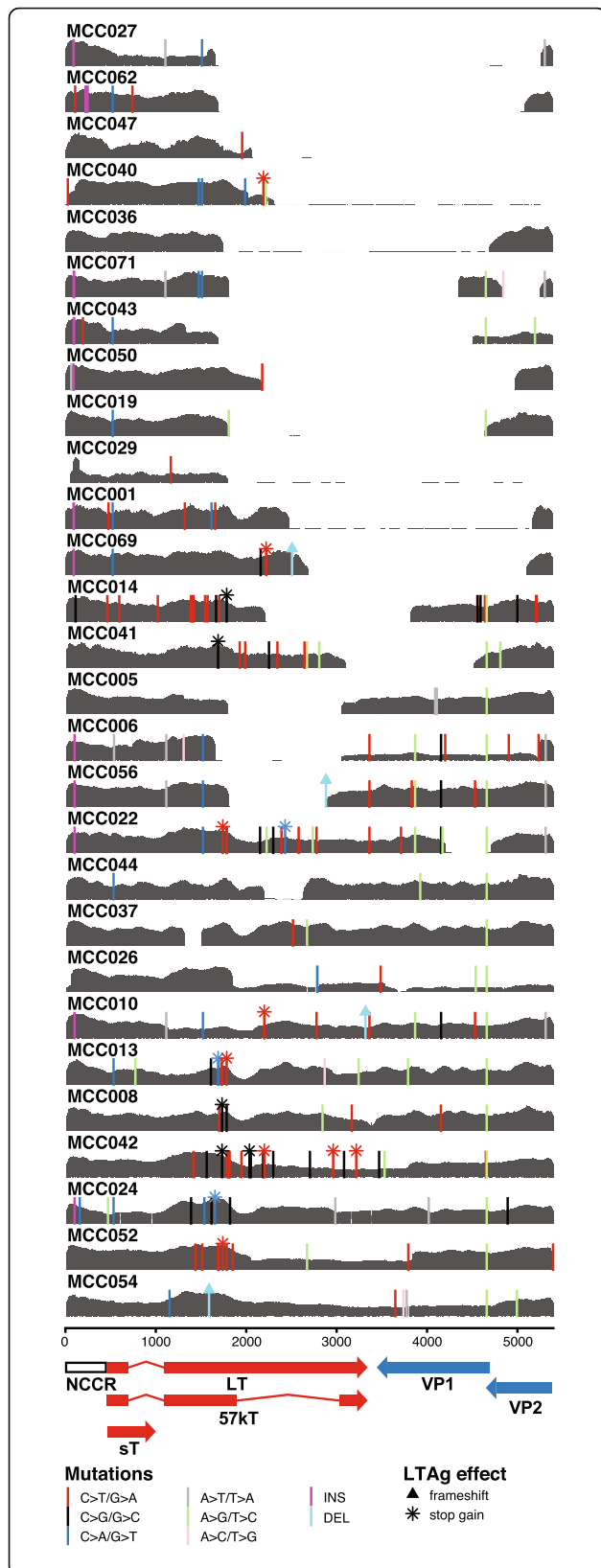
Beyond indels and nonsense mutations, LT also carried numerous novel clonal missense mutations (Fig. 5a) unique to the patients in this cohort. In stark contrast, ST only had missense mutations at three residues, and the amino acid change A20S is consistent with a previously observed MCPyV strain difference (GenBank

identical protein accession number: ACI25295.1). The other missense mutations occurred clonally at H41Y and N100S once in the entire cohort (Fig. 5b). Neither of these mutations are present in any of the ST sequences in GenBank and have not been previously reported.

The integration sites were mapped using the oncovirus tools suite ([https://github.com/gstarrett/oncovirus\\_tools](https://github.com/gstarrett/oncovirus_tools)) (Fig. 6a, Table 2) [46]. 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 (Fig. 6b). These represent the first reported cases of recurrent viral integration sites in MCC.

Based on previous MCC WGS, MCPyV integration sites frequently coincide with focal amplifications in the human genome. We can therefore infer that the regions between distant (> 10 kb) viral integration sites were amplified; however, no targeted exon were within these regions [8]. Because of the limited number of capture targets sequence by the OncoPanel platform, determining the exact boundaries of the expected virus-mediated amplifications in cases with junctions < 10 kb apart was





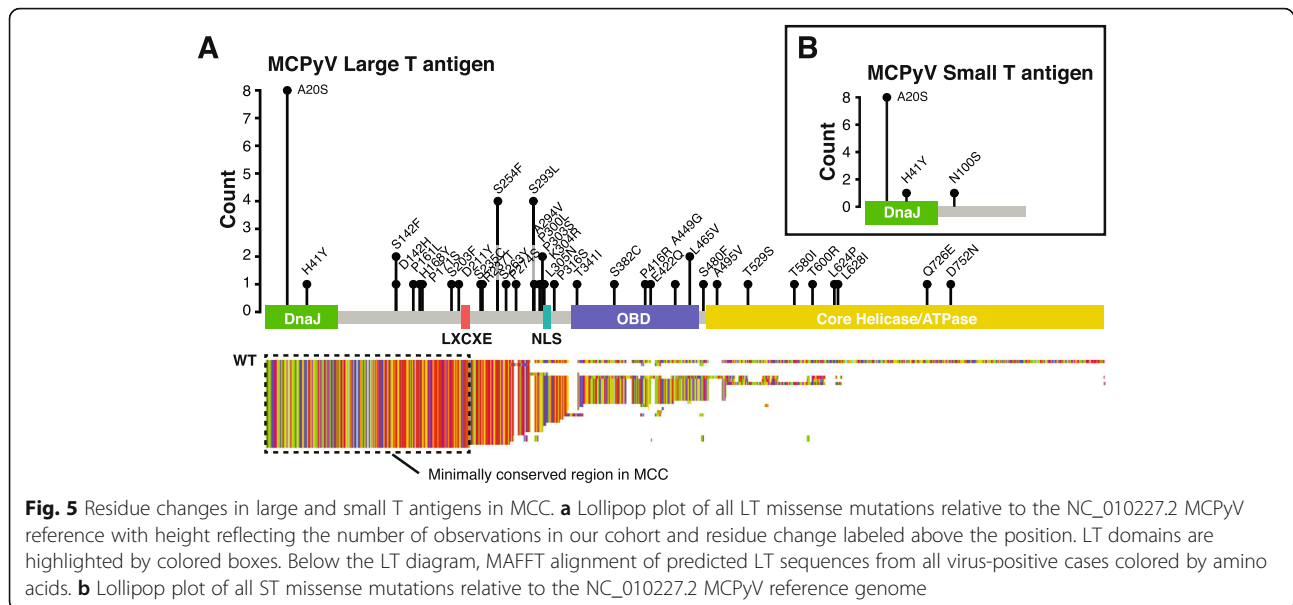
**Fig. 4** MCPyV coverage and mutations from virus-positive cases. Read coverage for MCPyV in gray and each plot represents a single patient with their ID in the upper left corner. Scales for the coverage plots are set from 0 to the maximum read coverage per patient. Point and insertion-deletion mutations 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

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). When annotating these regions, we observed that they frequently contain enhancer regions that may contribute to oncogenesis as seen in HPV-associated tumors [47]. Uniquely, patient MCC026 had integrations in chromosomes 9, 16, and 18, all of which had integration sites separated by between 107.5 and 129.9 kbp appearing to be distinct events.

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 (Additional file 3: Fig. S6).

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 (Fig. 6c, Additional file 3: Fig. S6). However, other integrations generated more complex assembly graphs with a multiple contigs linked together in a “pigtailed” conformation (Fig. 6d, Additional file 3: Fig. S6). Based on coverage and conformation, this graph likely represents an integration event containing partially duplicated viral genome concatemers fused to different segments of the human genome. For samples with distant integration sites, 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 (Fig. 6e, Additional file 3: Fig. S6).

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



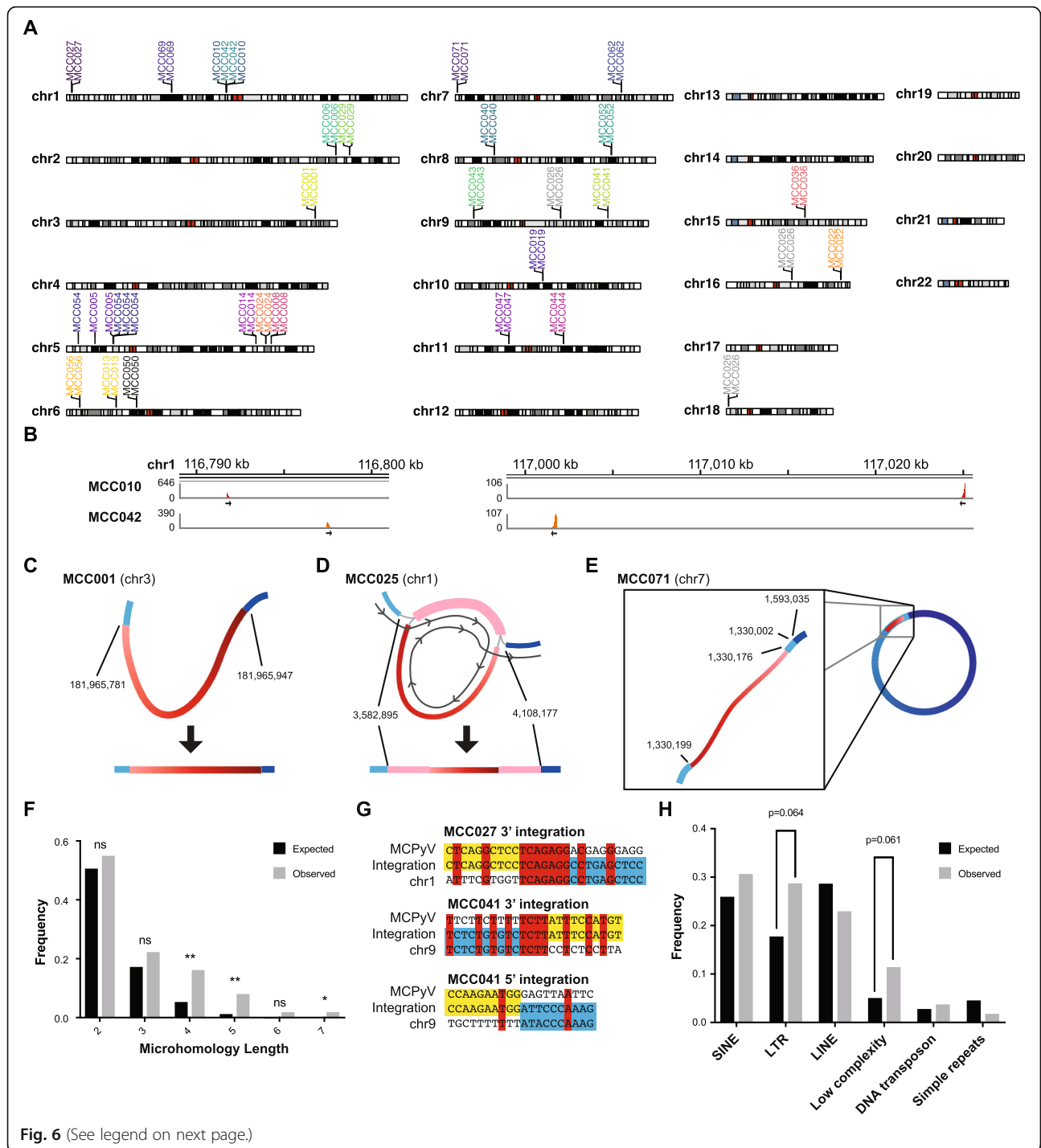
(Fig. 6f). There was no significant increase in overall homology between MCPyV and human DNA at integration sites versus randomly selected sites. Patient MCC027 had the integration site with the longest stretch of homology, and MCC041 had both the integration site with the greatest overall homology on its 3' end and lowest homology with no microhomology greater than 1 bp on its 5' end (Fig. 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 1.5 kb of a repeat element and there was a trend towards integrations near LTR retrotransposons and low-complexity regions (Fig. 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 ViroPanel dataset that strongly correlated with virus integration, we compared the data from the OncoPanel and POPv3/ViroPanel datasets to determine the viral status of all 71 tumors 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 ViroPanel data, there was a rough correlation between the number of off-target reads and the number of MCPyV reads in the ViroPanel dataset. There were 8 samples with MCPyV reads in the OncoPanel dataset that were not also analyzed by ViroPanel. 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 either 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 and 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 [48]. 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 virus-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 ViroPanel 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 primer set positive also ranked at the TMB borderline



**Fig. 6** (See legend on next page.)

(See figure on previous page.)

**Fig. 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 in 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 Additional file 3: Fig. S6. **c** Representative single linear assembly graph for integrated MCPyV from case MCC001 on chromosome 3. **d** 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 gray 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 2 and 7 bp. Expected values are in black and observed are in gray. 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 gray.  $P$  values from Fisher's exact test are represented above the bars

(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 (Additional file 1: Table S4) [25]. Criteria evaluated included procedure, 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 which 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 (Additional file 2: Fig. S5). According to the biopsy type and first relapse 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 biopsies annotated as a recurrence, the first recurrence occurred before the biopsy was obtained. Among the 19 recurrence biopsies, 15 were distant metastatic biopsies, one local recurrence (MCC027), one unspecified recurrence (MCC063), one second recurrence (MCC057), and one local recurrence with no prior chemo/XRT (MCC026). Regardless of the biopsy type, all patients and sequencing data were grouped into either no relapse ( $N = 30$ ) or 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 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 (Additional file 1: 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 cases were identified as virus-negative MCC with relatively high TMB compared to the 28% virus-negative MCC in immunocompetent patients (Fig. 7a, Table 5). Virus-negative MCC was present in three patients with solid organ transplantation; three with autoimmune diseases including myasthenia gravis, rheumatoid arthritis, and granulomatosis with polyangiitis; one with monoclonal gammopathy of undetermined significance (MGUS); and another with Waldenstrom's macroglobulinemia. 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 germline mutations in NF1 and GATA2 [49]. 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 months,  $p < 0.01$ ) (Fig. 7b, Table 5). Immunosuppressed patients also exhibited significantly shorter recurrence-free survival, 7.5 months (95% CI 3.5–20.1 months) and 20.2 months (95% CI 12.9–50.2 months,  $p = 0.01$ ), respectively. We acknowledge that some cases have relatively short follow-up times that may impact survival analysis; however, the association of OS and immunosuppression only shifts slightly after keeping patients with follow-up times greater than 6 months ( $n = 63$ ) remaining statistically significant. Under this criterion, the median OS for immunocompetent and immunosuppressed cases were 48.5 months (95% CI 35.4 to 113.3 months) and 21.6 months (95% CI 6.9 to 30.7 months,  $p$  value < 0.01), respectively.

#### Discussion

We undertook this study to develop an assay to more accurately distinguish between virus-positive and virus-negative MCC by genetic features. We built upon an

**Table 2** MCPyV integration sites

ID	Off-target reads	ViroPanel 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	87,721	0.77	5.60786	5:20753360,33939328
MCC006	4	37,150	0.77	2.50159	2:196945370,196945371
MCC007	0	971	0.97	0.00317	
MCC008	0	69,634	1.00	1.73052	5:149618981,149709442
MCC009	0	3	0.05	0.00004	
MCC010	4	111,147	1.00	5.37166	1:116791739,117025123
MCC011	0	212	0.87	0.00039	
MCC012	0				
MCC013	78	113,184	1.00	5.32306	6:36192882,36282634
MCC014	0	21,307	0.70	1.45712	5:138420218,138511276
MCC015	0	177	0.77	0.00036	
MCC016	0				
MCC017	0				
MCC018	2				
MCC019	0	13,343	0.50	0.83000	10:63999700,64000021
MCC020	0	120	0.83	0.00029	
MCC021	0	8	0.07	0.00004	
MCC022	0	22,715	0.93	0.72455	16:83581326,83890305
MCC023	0	1	0.02	0.00002	
MCC024	8	521,049	1.00	24.81107	
MCC025	0				
MCC026	0	48,346	0.99	2.25473	9:76893837,77023700; 16:47914233,48036152; 18:1561377,1668866
MCC027	0	17,748	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	
MCC035	0	5	0.08	0.00002	
MCC036	0	21,479	0.47	1.21591	15:57507670,57507677
MCC037	194	21,095,751	1.00	511.66883	
MCC038	0				
MCC039	0				
MCC040	0	181,807	0.96	7.46100	8:28408988,28457320
MCC041	0	12,858	0.74	0.78054	9:111568335,111579165
MCC042	0	75,177	1.00	1.99230	1:116797448,116797523
MCC043	1	31,969	0.48	3.14034	9:13451094,13451103

**Table 2** MCPyV integration sites (Continued)

ID	Off-target reads	ViroPanel reads	MCPyV coverage	Normalized coverage	Integration site(s)
MCC044	0	35,415	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	34,047	0.48	2.10949	6:51146411,51146421
MCC051	7				
MCC052	8	74,199	1.00	2.06071	8:113896842,114256794
MCC053	0				
MCC054	0	103,352	1.00	2.95039	5:8556313,34193826 (34349919–34349456)
MCC055	0	265	0.96	0.00052	
MCC056	6	85,232	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	10,735	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	25,483	0.55	1.31799	1:76825442,76826185
MCC070	0	5	0.09	0.00001	
MCC071	0	19,543	0.47	1.61800	7:1330002,1593035

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, ViroPanel, 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 (> 6000). 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 in 12 different chromosomes with the most occurring on chromosome 5. In addition, two fully overlapping integration sites from two different tumors were observed on chromosome 1 separated by only 10–20 kb. 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 [50]. For both MCPyV and HPV, it has previously been proposed that integration initiates after DNA double strand breaks in the host genome and viral genomes, likely during viral genome replication as integrated viral concatemers are common [8, 51]. In this study, we identified that integration is then likely mediated through erroneous DNA repair at sites of microhomology between the host and viral genomes. This is

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

ID	TMB	TMB category	PCR # primer sets	IHC	ViroPanel	Virus-positive Evidence	Add'l VP Evidence	UV	VP or VN
MCC001	8.5	Intermediate	2	1	Positive	ViroPanel	PCR IHC	0	VP
MCC002	30.4	High	0	0	Negative	UV ViroPanel TM		1	VN
MCC003	0.0	Low						0	VP
MCC004	6.6	Intermediate						0	VP
MCC005	1.6	Low		1	Positive	ViroPanel	IHC	0	VP
MCC006	3.7	Low			Positive	ViroPanel		0	VP
MCC007	9.6	Intermediate	1	0	Negative	ViroPanel		0	VN
MCC008	2.7	Low	5	1	Positive	ViroPanel	PCR IHC	0	VP
MCC009	28.2	High		0	Negative	UV ViroPanel 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	ViroPanel	PCR IHC	0	VP
MCC014	3.7	Low		1	Positive	ViroPanel		0	VP
MCC015	27.7	High	0	0	Negative	UV ViroPanel 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	ViroPanel		0	VP
MCC020	27.7	High		0	Negative	UV TM		1	VN
MCC021	25.0	High			Negative	UV ViroPanel TM		1	VN
MCC022	3.7	Low			Positive	ViroPanel		0	VP
MCC023	19.2	Intermediate			Negative	UV ViroPanel TM		1	VN
MCC024	3.2	Low			Positive	ViroPanel		0	VP
MCC025	0.0	Low						0	VP
MCC026	4.3	Low			Positive	ViroPanel		0	VP
MCC027	2.7	Low	4	1	Positive	ViroPanel	PCR IHC	0	VP
MCC028	29.3	High			Negative	UV ViroPanel TM		1	VN
MCC029	5.3	Low		0				0	VP
MCC030	38.9	High	0	0	Negative	UV ViroPanel TM		1	VN
MCC031	1.6	Low						0	VP
MCC032	1.1	Low						0	VP
MCC033	5.3	Low		1	Positive	ViroPanel	IHC	0	VP
MCC034	28.8	High	0	0	Negative	UV ViroPanel TM		1	VN
MCC035	11.2	Intermediate			Negative	UV ViroPanel TM		1	VN
MCC036	3.2	Low	4	1	Positive	ViroPanel	PCR IHC	0	VP
MCC037	4.8	Low		1	Positive	ViroPanel	IHC	0	VP
MCC038	1.6	Low						0	VP
MCC039	22.4	High				UV TM		1	VN
MCC040	3.7	Low			Positive	ViroPanel		0	VP
MCC041	2.1	Low			Positive	ViroPanel		0	VP
MCC042	2.1	Low			Positive	ViroPanel		0	VP
MCC043	4.3	Low			Positive	ViroPanel		0	VP
MCC044	5.3	Low	2	1	Positive	ViroPanel	PCR IHC	0	VP

**Table 3** Comparison of sequencing, PCR, and IHC for determination of tumor viral status (*Continued*)

ID	TMB	TMB category	PCR # primer sets	IHC	ViroPanel	Virus-positive Evidence	Add'l VP Evidence	UV	VP or VN
MCC045	2.1	Low		1	Positive	ViroPanel	IHC	0	VP
MCC046	20.8	High			Negative	UV ViroPanel TM		1	VN
MCC047	2.1	Low		1	Positive	ViroPanel	IHC	0	VP
MCC048	26.6	High			Negative	UV ViroPanel TM		1	VN
MCC049	34.1	High			Negative	UV ViroPanel TM		1	VN
MCC050	4.8	Low			Positive	ViroPanel		0	VP
MCC051	3.2	Low		1	Positive	ViroPanel	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	ViroPanel	PCR IHC	0	VP
MCC055	30.9	High			Negative	UV ViroPanel TM		1	VN
MCC056	3.7	Low			Positive	ViroPanel		0	VP
MCC057	29.8	High			Negative	UV ViroPanel TM		1	VN
MCC058	1.1	Low						0	VP
MCC059	19.2	Intermediate			Negative	UV ViroPanel TM		1	VN
MCC060	1.1	Low						0	VP
MCC061	2.1	Low						0	VP
MCC062	4.8	Low	4	0	Positive	ViroPanel	PCR	0	VP
MCC063	27.2	High			Negative	UV ViroPanel 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
MCC068	2.1	Low						0	VP
MCC069	4.3	Low		1	Positive	ViroPanel	IHC	0	VP
MCC070	29.8	High			Negative	UV ViroPanel TM		1	VN
MCC071	3.2	Low						0	VP

IHC MCPyV staining, 0 = negative, 1 = positive; UV UV signature present, 0 = negative, 1 = positive

similar to mechanisms identified for microhomology-mediated end joining (MMEJ) for HPV genome integration in tumors, but it has yet to be determined if the same host factors are involved [52]. For both MCPyV and HPV mediated tumors, the MMEJ event frequently leads to the formation of a transiently circular DNA intermediate, which can be amplified through aberrant firing of the viral origin of replication [53]. The resulting large linear DNA then reintegrates into the chromosome and appears as amplified regions of the host genome in a tandem head-to-tail conformation interspersed with the viral genome [8].

Case MCC026 has three apparently separate integration events occurring on different chromosomes. The integration event on chromosome 16 only contains a small section of the viral genome from positions 2853–3521, which would only encode the helicase domain of LT and therefore is unlikely to contribute to tumor survival. Conversely,

the event on chromosome 18 has a full copy of the viral genome while the event on chromosome 9 contains the NCCR, ST, and a truncated LT, likely sufficient to contribute to oncogenesis. Based on the assembly graphs and coverage, one or both of these integration events have more than one copy of ST and LT (Fig. 4 and Additional file 3: Fig. S6). Distinct sequences derived during assembly and the distances between the intrachromosomal junctions (107–129kbp) indicate that these likely are separate events, but only long-read sequencing of this tumor can definitively determine that these are not part of a larger interchromosomal translocation.

The most common chromosomal copy number changes involved chromosomes 1 and 10. Amplification of 1p (cluster 4) involving MYCL was observed more commonly in virus-negative cases, but was identified in a few virus-positive cases. Interestingly, in virus-positive MCC, MCPyV ST binds MYCL and the EP400 chromatin modifying



**Table 4** Association between relapse and genomic sequencing ( $N = 71$ )

Characteristics	All ( $N = 71$ )	Relapse or not		Fisher's 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%)	

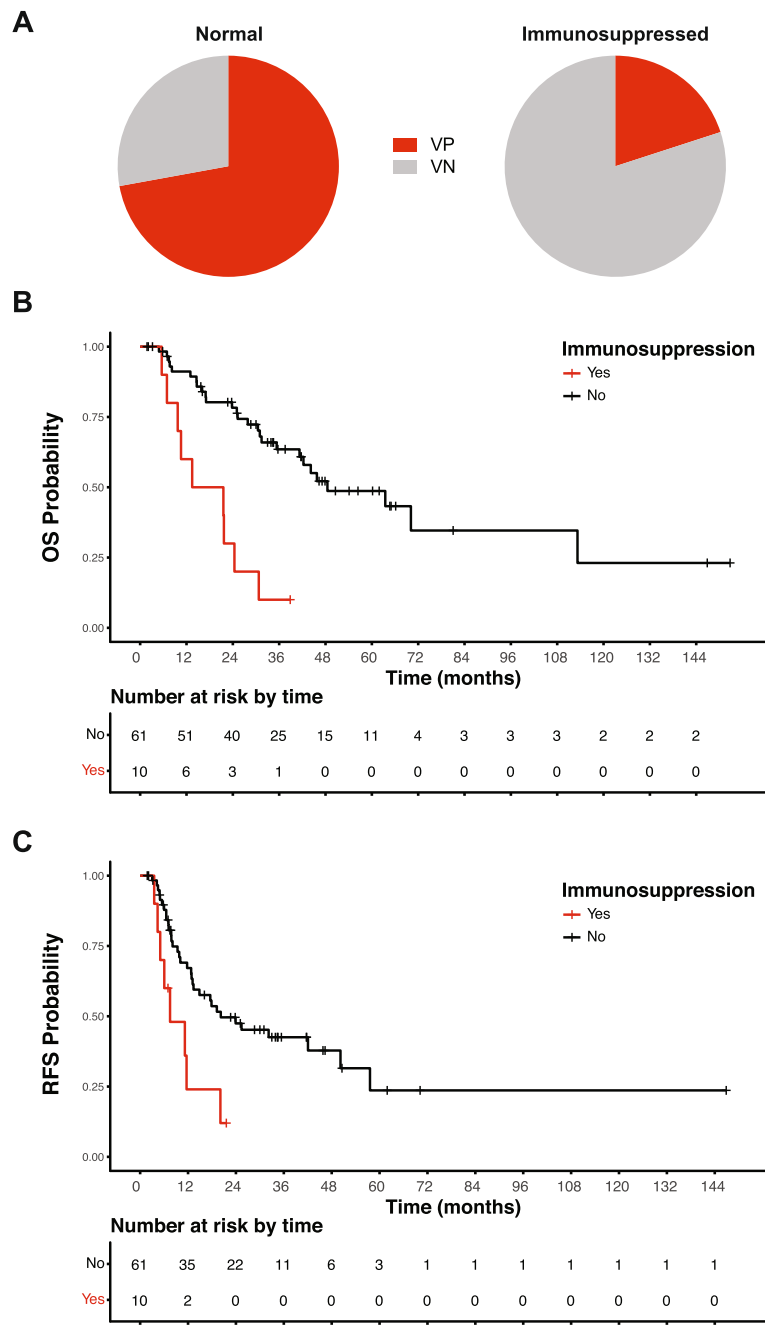
complex to activate transcription of several hundred target genes [54]. Amplification of MYCL is likely to be an oncogenic event that contributes to MCC aggressiveness. Amplification of Chr1q (cluster 13) was also observed in both virus-positive and virus-negative MCC. This region includes MDM4, whose protein product cooperates with MDM2 to promote the ubiquitination and subsequent degradation of p53 [43, 44]. There may be additional oncogenic genes in this cluster that contribute to MCC oncogenesis when p53 is mutated. Heterozygous loss of chromosome 10 (cluster 14) was observed in 26/71 MCC including both virus-positive and virus-negative tumors. Loss of chromosome 10 likely reduces PTEN levels contributing to activation of the PI3K signaling pathway.

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 [55]. In other cancers, such as basal cell carcinoma and ovarian cancer, this amplification is typically associated with worse outcome [55]. Although the chromosome 6 amplification in this study was significantly associated with better overall survival, it was also more frequent in metastasis. This amplification contains genes such as *VEGFA*, which promotes angiogenesis and has been observed to be expressed at higher levels in distant ovarian cancer metastases [56]. Interestingly, further analysis of genes that are more abundantly mutated in the sequenced metastatic lesions, revealed statistically significant ( $q = 2.47 \times 10^{-7}$ ) pathway enrichment for angiogenesis

as well as EGFR signaling and p53 dysregulation. *FANCE* is also carried on this chromosome and as a DNA damage response gene act to limit number of point mutations and copy number changes observed in virus-positive MCC, hindering tumor evolution. Together, this amplification co-occurring with MCPyV may represent a less fatal, but more metastatic subtype of MCC. Additionally, this result could be impacted by diagnosis, treatment, or study recruitment of 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 [57], it was not until 1997 when a direct link between immunosuppression and MCC was postulated [58]. 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. Notably, the search for a viral pathogen in MCC was initiated because of reports linking MCC with immunosuppression and with HIV-1/AIDS [2]. A subsequent report has shown similar rates of MCPyV-positivity in immunocompetent and suppressed patients, but relied on PCR and IHC for virus detection [18].

In the present report, three solid organ transplant recipients, three with chronic autoimmune diseases, and two with hematologic malignancies developed virus-negative



**Fig. 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, gray) and immunocompetent or immunosuppressed. **b** Kaplan-Meier plot of overall survival of immunocompetent (black) and immunosuppressed (red) MCC patients

MCC. It is well established that 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 [58–61]. Within the latter population, 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 [62].

Importantly, some therapeutics used in organ transplantation are known to further 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 [63]. In patients with rheumatoid arthritis, methotrexate and anti-TNF drugs were associated with an increased risk of nonmelanoma skin cancer

**Table 5** Association between patient 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, years				
< =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%)	

[64]. 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. Although this study is one of the largest molecular studies on MCC genetics to date, the small sample size and inherent confounding factors of studying outcome in a cancer that manifests in older populations are important limitations. This also

highlights the need for accurate determination of virus status and importance to continue to study this rare cancer to fully address the involvement of immune suppression on the etiology and outcome of this aggressive cancer.

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 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, 65]. 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 lifelong 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 [66]. 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 rather than direct effects on DNA itself [67].

## Conclusions

Here we present a comprehensive characterization of the Merkel cell carcinoma genetics using a clinically implemented sequencing platform. This platform was augmented using a hybrid capture baitset against Merkel cell polyomavirus. From our analyses, we identified CNV clusters unique to and common to virus-negative and virus-positive, which reflect the evolutionary mechanisms of the tumors. We also accurately reconstructed the viral integration events providing clear evidence for a circular host-fusion DNA intermediate initiated by recombination at 4+ bp microhomology enriched at low-complexity regions in the human genome. Lastly, we observed a surprising number of virus-negative tumors in immunosuppressed patients in our cohort potentially reflecting a previously misunderstood risk population.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s13073-020-00727-4>.

**Additional file 1: Table S1.** SNV data for all patients. Table S2. CNV data for all patients. Table S3. CNV cluster definitions. Table S4. Synoptic review of dermatopathology. Table S5. Association between relapse and genomic sequencing ( $N = 52$ ).

**Additional file 2: Fig. S1.** Oncoprint for all genes in this study. Oncoprint summarizing point mutations and CNVs for all genes and patients in this study. Sample are in order of descending TMB and genes are in order of highest point mutations to least. Fig. S2. Network graph for recurrent CNVs. Network graph with all significantly ( $q < 0.05$ ) co-occurrent CNVs showing how they cluster into reoccurring groups. Fig.

S3. Additional file 4.pdf: CNV frequency by cluster for all patients. Cluster number is shown in gray bars above the bar plots representing amplifications/gains (red) and deletions/losses (blue). Below the bar plots is a heat map of all CNVs (genes, x-axis) (amplifications/gains, red; deletions/losses, blue; no change, gray) across all samples (y-axis) annotated by cluster and chromosome. On the left side pRB, p53 shown in gray and black for 1 or 2 copy loss/mutant, respectively. Presence of UV mutations are shown in black. Fig. S4. RFS survival divided by CNV cluster 5. Kaplan-meier plot for MCC patients split by presence or absence of CNV cluster 5 showing no difference in regression free survival (RFS). Fig. S5: Plot of tumor sampling by patient for sequencing. Plot of all patients along the y axis and time since diagnosis on the x axis. Initial diagnosis, primary biopsy sequencing, recurrence biopsy sequencing, death, and last contact times are plotted as applicable.

**Additional file 3: Fig. S6.** Assembly graphs for all integration events. 28 assembly graphs annotated by MCPyV genome position (colors labeling each segment are under the header "as.factor(V9)"). The "V3" variable represents the coverage of each contig as determined by SPAdes and contigs are scaled to reflect this value.

## Abbreviations

CNV: Copy number variant; MCC: Merkel cell carcinoma; MCPyV: Merkel cell polyomavirus; SNV: Single nucleotide variant; TCGA: The Cancer Genome Atlas; TMB: Tumor mutation burden; UV: Ultraviolet; VN: Virus-negative; VP: Virus-positive

## Acknowledgements

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.

## Funding

This work was supported in part by US Public Health Service grants ([report.nih.gov](http://report.nih.gov)) R01CA63113, R01CA173023, P01CA203655, and R35CA232128 to JAD. Salary support for GJS comes from the NCI's cancer research training award.

## Availability of data and materials

OncoPanel data generated and analyzed for the current study is derived from patient samples containing identifiable sequencing data and is not publicly available by the IRB guidelines. However, the corresponding author will make every effort to share data by request. Non-human, ViroPanel sequences are available through the NCBI Sequence Read Archive (SRA) under Bioproject ID PRJNA606748. TCGA data used in this study are publicly available from cBioPortal (<https://www.cbioportal.org/>) [34, 35]. Additional TCGA sequencing data are available from the National Cancer Institute Genomic Data Commons data portal (<https://portal.gdc.cancer.gov/>). CNV and SNV calls are available in Additional file 1: Tables S1 and S2. All coordinates used are based on the hg19 human reference genome. Our software to determine viral integration sites and co-occurrent CNVs is available from [https://github.com/gstarrett/oncovirus\\_tools](https://github.com/gstarrett/oncovirus_tools) [46].

## Authors' contributions

GJS, LEM, and JAD contributed to the conceptualization. G.J.S., T.C., L.E.M., and J.A.D. contributed to the methodology. GJS and ART contributed to the software. GJS and JC contributed to the validation. GJS, TC, JN, RTB, and JAD contributed to the formal analysis. MT, CM, JC, ART, WP, MS, AP, FK, and JAD contributed to the investigation. MT, CP, GR, AG-H, LEM, and JAD contributed to the resources. GJS, TC, CP, ART, LEM, and JAD contributed to the data curation. GJS, TC, ART, and JAD contributed to the writing of the original draft. GJS, LEM, and JAD contributed to the writing and reviewing and editing. G.J.S. contributed to the visualization. ART, AG-H, LEM, and JAD contributed to the supervision. AG-H, LEM, and JAD contributed to the project administration. JAD contributed to the funding acquisition. The author(s) read and approved the final manuscript.

**Ethics approval and consent to participate**

This study was conducted according to the Declaration of Helsinki principles and approved by the Dana-Farber Cancer Institute institutional review board (DFCI Protocol No. 18–240). Written informed consent was received from participants prior to inclusion in the study.

**Consent for publication**

Not applicable.

**Competing interests**

JAD has received honoraria for participation in an advisory board with Merck & Co., Inc. JAD has received research funding from Constellation Pharmaceuticals, Inc. The remaining authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Laboratory of Cellular Oncology, CCR, NCI, NIH, Bethesda, MD, USA. <sup>2</sup>Department of Dermatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. <sup>3</sup>Merkel Cell Carcinoma Center of Excellence, Dana-Farber/Brigham Cancer Center, Boston, MA, USA. <sup>4</sup>Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA, USA. <sup>5</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, Mayer 440, 450 Brookline Avenue, Boston, MA 02215, USA. <sup>6</sup>Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. <sup>7</sup>Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, MA, USA. <sup>8</sup>Department of Informatics and Analytics, Dana-Farber Cancer Institute, Boston, MA, USA. <sup>9</sup>Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. <sup>10</sup>Present Address: Miami Cancer Institute, Miami, FL, USA.

Received: 23 September 2019 Accepted: 27 February 2020

Published online: 18 March 2020

**References**

- 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.
- Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science*. 2008;319(5866):1096–100.
- 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.
- 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.
- Schwalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. *Cell Host Microbe*. 2010;7(6):509–15.
- 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.
- 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.
- 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) <https://doi.org/10.1128/mBio.02079-16>.
- 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.
- 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.
- 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.
- Lipson EJ, Vincent JG, Loyo M, Kagohara LT, Lubner BS, Wang H, et al. PD-L1 expression in the Merkel cell carcinoma microenvironment: association with inflammation, Merkel cell polyomavirus and overall survival. *Cancer Immunol Res*. 2013;1(1):54–63.
- Leroux-Kozal V, Leveque N, Brodard V, Lesage C, Duzde 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.
- 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.
- 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. *Lancet Oncol*. 2016;17(10):1374–85.
- 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.
- Rodrig 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.
- 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.
- 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;23(18):5622–5630.
- 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.
- 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.
- 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.
- Garcia EP, Minkovskaya A, Jia Y, Ducar MD, Shivdasani P, Gong X, et al. Validation of OncoPanel: a targeted next-generation sequencing assay for the detection of somatic variants in Cancer. *Arch Pathol Lab Med*. 2017;141(6):751–8.
- Wagle N, Berger MF, Davis MJ, Blumenstiel B, Defelice M, Pochanard P, et al. High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer Discov*. 2012;2(1):82–93.
- 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.
- Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
- 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.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.
- 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.
- McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. *Bioinformatics*. 2010;26(16):2069–70.
- Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80–92.

32. 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.
33. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput Biol*. 2016;12(4):e1004873.
34. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401–4.
35. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013;6(269):p11.
36. Bloom BH. Space/time trade-offs in hash coding with allowable errors. *Comm ACM*. 1970;13(7):422–6.
37. 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.
38. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
39. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res*. 2019;47(D1):D419–D26.
40. 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.
41. Van Gele M, Leonard JH, Van Roy N, Cook AL, De Paepe A, 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.
42. Cohen PR, Kurzrock R. Merkel cell carcinoma with a suppressor of fused (SUFU) mutation: case report and potential therapeutic implications. *Dermatol Ther*. 2015;5(2):129–43.
43. 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.
44. 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.
45. 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.
46. Starrett GJ. Oncovirus tools. GitHub. <https://doi.org/10.5281/zenodo.3661416>. 2020.
47. 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.
48. 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.
49. Crall C, Morley KW, Rabinowitz G, Schmidt B, Broyles AD, Huang JT. Merkel cell carcinoma in a patient with GATA2 deficiency: a novel association with primary immunodeficiency. *Br J Dermatol*. 2015;.
50. 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;.
51. Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, et al. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res*. 2014;24(2):185–99.
52. Leeman JE, Li Y, Bell A, Hussain SS, Majumdar R, Rong-Mullins X, et al. Human papillomavirus 16 promotes microhomology-mediated end-joining. *Proc Natl Acad Sci U S A*. 2019;116(43):21573–9.
53. Nulton TJ, Olex AL, Dozmorov M, Morgan IM. Analysis of the cancer genome atlas sequencing data reveals novel properties of the human papillomavirus 16 genome in head and neck squamous cell carcinoma. *Oncotarget*. 2017;8(11):17684–99.
54. Cheng J, Park DE, Berrios C, White EA, Arora R, Yoon R, et al. Merkel cell polyomavirus recruits MYCL to the EP400 complex to promote oncogenesis. *PLoS Pathog*. 2017;13(10):e1006668.
55. Santos GC, Zielenska M, Prasad M, Squire JA. Chromosome 6p amplification and cancer progression. *J Clin Pathol*. 2007;60(1):1–7.
56. Sopo M, Anttila M, Hamalainen K, Kivela A, Yla-Herttuala S, Kosma VM, et al. Expression profiles of VEGF-A, VEGF-D and VEGFR1 are higher in distant metastases than in matched primary high grade epithelial ovarian cancer. *BMC Cancer*. 2019;19(1):584.
57. Pitale M, Sessions RB, Husain S. An analysis of prognostic factors in cutaneous neuroendocrine carcinoma. *Laryngoscope*. 1992;102(3):244–9.
58. 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.
59. Lanoy E, Engels EA. Skin cancers associated with autoimmune conditions among elderly adults. *Br J Cancer*. 2010;103(1):112–4.
60. 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) <https://doi.org/10.1093/jnci/dju382>.
61. Sahi H, Sihto H, Artama M, Koljonen V, Bohling T, 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.
62. 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 Dermatol*. 2017 153(3):296–303.
63. Wheless L, Jacks S, Mooneyham Potter KA, Leach BC, Cook J. Skin cancer in organ transplant recipients: more than the immune system. *J Am Acad Dermatol*. 2014;71(2):359–65.
64. Scott FJ, 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 Dermatol*. 2016; 152(2):164–72.
65. Busam KJ, Jungbluth AA, Rekhman 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.
66. 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.
67. Dowlathshahi 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.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

