Detection of Merkel Cell Polyomavirus Using Whole Exome Sequencing Data

Sandra Garcia-Mulero
IDIBELL
Ferran Moratalla-Navarro
Institut d'Investigacio Biomedica de Bellvitge
Soraya Curiel-Olmo
IDIVAL
Victor Moreno
Institut Catala d' Oncologia
José Pedro Vaqué
Universidad de Cantabria
Rebeca Sanz-Pamplona (rebecasanz@iconcologia.net)
IDIBELL https://orcid.org/0000-0002-2187-3527
Josep María Piulats
Institut Catala d' Oncologia

Technical advance

Keywords: Sequencing, Merkel cell carcinoma, viral insertion, pipeline

Posted Date: September 21st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-60576/v1

Abstract

Background

Merkel cell carcinoma (MCC) is a highly malignant neuroendocrine tumor of the skin in which Merkel cell polyomavirus (MCV) DNA virus insertion can be detected in 75–89% of cases. Etiologic and phenotypic differences exist between MCC tumors with and without the inserted virus, thus it is important to distinguish between MCV+ and MCV- cases.

Methods

Here we report a freely available bioinformatics methodology to identify MCV+ MCC tumors using whole exome sequencing (WES) data and to infer the virus insertion site into the tumor genome.

Results

Our method has been validated in a set of MCC samples previously characterized in the laboratory as MCV+ or MCV-. MCV insertions have also been reported in non-smokers lung tumors but no MCV+ was found in two tested datasets using our pipeline.

Conclusions

With our pipeline it is possible to use WES to interrogate for MCV insertions in cancer samples.

Background

Merkel cell carcinoma (MCC) is a highly malignant neuroendocrine tumor of the skin [1]. MCC typically affects immunosuppressed individuals thus suggesting an infectious origin. Indeed, in 2008, a 5.4 kbp polyomavirus DNA was found to be integrated in some MCC genomes in a clonal pattern [2]. Although MCC is rare, its incidence has tripled over the past two decades in the United States. This could be due for both the increasing prevalence of risk factors such as ultraviolet (UV) exposure or systemic immune suppression [3] or because of the improvement of diagnostic methods. Merkel cell polyomavirus (MCV) DNA virus can be detected in 75–89% of MCC. It is noteworthy that etiologic and phenotypic differences exist between MCC tumors with and without the inserted virus [4]. In this regard, recent work by independent laboratories has shown important genetic differences between MCV+ and MCV- MCC tumors, the latter harboring higher mutational burdens with ultraviolet (UV) signatures [5–8]. Moreover, relevant clinical features are associated with this phenotype being the MCV- subtype the one showing a worst prognosis [9]. Also, from a therapeutic point of view, immunotherapy could play an important role in this tumor [10,11]. Thus, it is important to distinguish between MCV+ MCC and MCV- MCC. Normally, MCV insertions in MCC genomes were detected by PCR or using an anti-MCP antibody [9]. Here we show a 2-alignment-steps methodology to identify MCV+ tumors using sequencing data.

Methods

Patients and samples

Whole exome sequencing (WES) data from a total of 15 MCC patients (9 freshly frozen -FF- and 6 formalin fixed paraffin embedded -FFPE-) and their normal paired samples were used to set up the pipeline. For clinical characteristics and genomic analysis, refer to González-Vela et al. study [6]. Also, MCV insertion was assessed in two lung adenocarcinomas datasets downloaded from the European Nucleotide Archive (ENA) public repository: series SRP090460 including 25 RNA-seq samples [12] and SRP022932 comprising 30 WES samples [13].

2-alignment-steps pipeline

The following pipeline was set up using a positive control sample (WES data from a MCV+ MCC tumor analyzed with standard laboratory techniques): after quality control assessment using FastQC [14], WES raw data was pre-processed with TrimGalore [https://github.com/FelixKrueger/TrimGalore] for adapters and bad quality reads removal. Then, reads were aligned against the human reference genome with Bowtie2 [15] using strict parameters to achieve a very sensitive local alignment. Next, unmapped reads (reads that did not map to the human genome) were retrieved using SamTools [16]. Finally, these non-human reads were aligned to the NCBI Merkel cell polyomavirus reference genome (EU375803.1; isolate MCC350, complete genome). This second alignment was performed with BWA aligner [17]. If reads were found matching with MCV genome, the tumor was classified as MCV+ (Figure 1). The script used to make this analysis is freely available at Github repository (https://github.com/odap-ubs/merkel_virus).

Previously, we assessed that no similarities between the MCV genome and any sequence in the Homo sapiens genome exist. Wgsim software [https://github.com/lh3/wgsim] was used to generate simulated 75-bp-length sequences from the MCV reference genome (10000, 100000, 1000000 reads, and 1000000 reads including polymorphisms). Any read mapped in the Human Genome.

Per sample coverage was calculated as twice the read length multiplied by total number of reads and divided by total genome length.

The pipeline was also applied on RNA-seq data. Reads were trimmed using the same parameters than those in WES data. Alignments were performed with STAR v2.6.0 (given the simplicity of the MCC genome without intergenic regions or introns, alignment could be done with aligners not accounting for spliced sites). Unmapped and discordant reads were extracted from bam files to realign it against the MCV genome.

Virus insertion site

Reads mapped to the MCV genome whose mates were mapped to the human genome were used to get information of the proximal insertion site (Supplementary Figure 3). Also, a visual inspection of the bam files using the Integrative Genome Viewer (IGV) [18] tool was done in order to find the exact site of insertion.

Results And Discussion

Identification of MCC MCV+ tumors using WES

We interrogated sequencing data for a total of 15 MCC primary tumors and their normal paired samples, previously characterized in the laboratory for MCV insertions (seven MCV+, and eight MCV-) [6]. Our method was able to detect all seven MCC positive for MCV (ranged from 2 to 44 mapped reads). Supplementary Figure 1 shows the MCV+ MCC aligned reads in the IGV tool. Interestingly, there was a tendency towards detecting more aligned reads in the first 2,500 bp of the MCV. As expected, none of the normal paired samples were found to present MCV insertions.

Sensitivity and specificity were 100% and 62.5%, respectively. However, due to the extremely low number of reads in false positive samples, better values of specificity would be obtained if a sample was catalogued as positive only when at least two MCV reads were detected. Strikingly, three out of eight previously classified as MCV- MCC were found to have at least one read mapping into MCV genome whereas the remaining five had zero mapped reads (Table 1). These three false positive samples were evaluated in detail. Two out of the three samples were probably artefacts since only one 20 bp length read with two mismatches was mapped in MCV genome. However, the third sample had two 75-bp-length reads mapped in MCV genome thus we cannot exclude the possibility that this is a true positive sample. It was also interesting to note that there was a correlation between the sequencing coverage and the number of mapped reads. This may indicate that when depth of sequencing increases also increases the probability of finding an inserted viral genome (Table 1).

Table 1. Alignment results.

SAMPLE	TISSUE	MCV	Nº MAPPED READS	COVERAGE
48120427	FF	Positive	44	3.519
48120428	FF	Positive	34	2.704
48121209	FF	Positive	16	2.543
48120987	FF	Positive	15	2.286
48140221	FF	Positive	7	2.090
48090369	FF	Positive	4	2.467
48130247	FFPE	Positive	2	1.794
48130206	FFPE	Negative	2	1.304
48120431	FF	Negative	1	3.213
48141029	FFPE	Negative	1	1.784
48141028	FFPE	Negative	0	1.869
48130208	FFPE	Negative	0	1.364
48130207	FFPE	Negative	0	1.691
48121576	FF	Negative	0	2.520
48120426	FF	Negative	0	3.192

It is worth to mention that a bioinformatic approach to detect MCV insertions was previously described by Knepper et al [19]. However, they performed de-novo assembly and then aligned into a variety of virus genomes rather than directly aligned off-targets reads into MCV genome.

Since interrogated samples were a mixture of fresh frozen (FF) and formalin-fixed paraffin-embedded (FFPE) preserved tumors, we wondered if our method performed well in both types of samples. Figure 2 suggests that our pipeline is more robust when NGS was performed in FF samples. This is not surprising since sequencing performs better in FF tissues. However, only one out of seven MCV+ MCC were FFPE. Moreover, the MCV+ FFPE sample was the one with the lower sequencing depth so a coverage bias rather than a tissue preservation effect cannot be excluded. Thus, more samples need to be assessed in order to settle this topic.

MCV site of insertion

Finally, paired-end WES data was used to try to infer the virus insertion site into the tumor genome. After applying this strategy in all positive samples, only one of them (48121209) showed soft-clipped reads whose mates were mapped on chr19:48,445,990. Interestingly, this region was into GRWD1 gene intronic region (Supplementary Figure 2A). This is a P53 regulator whose loss of function has been previously

associated with tumorigenesis [20]. The fact that this is a highly covered region might indicate that our method failed to detect insertion positions in poor covered regions. Thus, whole-genome sequencing should be a better technique to identify not only MCV+ tumors but also the virus insertion site. Supplementary Figure 2B showed how soft-clipped reads also mapped against the MCV genome.

MCV insertion in non-small cell lung patients

Apart from MCC, MCV insertions have been found in non-small cell lung cancer in non-smokers [21]. However, this is a controversial result since other study has reported no MCV insertions in this type of cancer [22]. Therefore, the 2-alignment-steps methodology was applied on non-smokers lung adenocarcinoma data sets; 25 RNA-seq samples [12] and 30 WES samples [13]. As a result, no MCV+ tumor was found. To note, we performed RNA-seq alignments with STAR and Bowtie2 with no differences. Our result appears to support a study reporting that MCV DNA fragments could be detected in the lower respiratory tract when a high-sensitive PCR assay was used thus generating false positives results [23].

Conclusions

In conclusion, here we present an easy, freely-available method to detect MCV insertions as a bystander product of sequencing. When tested in previously characterized MCV+ and MCV- MCC tumors, high concordance was achieved. In two non-small cell lung cancer datasets no MCV insertions have been found. Since human gene exons DNA are enriched in WES, MCV will be most likely detected in MCC if its genome is inserted in exome regions or when high sequencing depth is obtained and MCV reads are detected as off-target reads. We cannot rule out the possibility that samples were contaminated with viral DNA, but the lack of positive results in paired normal samples sequenced with similar coverage was reassuring. Also, the identification of an insertion site in one of the samples. This methodology can also be applied using RNA-seq as input. Thus, with enough depth of sequencing, it is possible to apply the pipeline described here to take advantage of WES experiments and assess the presence of MCV insertions in any dataset of interest.

Abbreviations

MCC: Merkel cell carcinoma, MCV: Merkel cell polyomavirus, WES: Whole exome sequencing, FF: Freshly frozen, FFPE: Formalin fixed paraffin embedded, PCR: Polymerase Chain Reaction, NGS: Next Generation Sequencing.

Declarations

Ethics approval and consent to participate:

Samples were collected in agreement with the Declaration of Helsinki protocols after obtaining written informed consent from all the patients and doctors involved, as required by the Comité Ético de

Investigación Clínica, Cantabria (CEIC). All processes were conducted in adherence with the recommendations of the CEIC.

Consent for publication:

We, (Josep María Piulats and Rebeca Sanz-Pamplona) hereby declare that we participated in the development of the manuscript titled Detection of Merkel cell polyomavirus using whole exome sequencing data. We have read the final version and give our consent for the article to be published in BMC Medical Genomics.

Availability of data and materials:

The data used and analysed during the current study is available from González-Vela et al. [6] but restrictions apply to the availability of this data, which was used under license for the current study, and so is not publicly available. Data is however available from the corresponding author on reasonable request and with permission of Dr. José Pedro Vaqué. The code used to do the analysis is available in the Github repository (https://github.com/odap-ubs/merkel_virus).

Competing interests:

Victor Moreno is consultant to Bioiberica S.A.U. and Grupo Ferrer S.A., received research funds from Universal DX and is co-investigator in grants with Aniling. Josep M Piulats is consultant for Roche-Genentech, Bristol Myers Squibb, Merck Sharp & Dohme, Merck-Serono, Janssen, Astellas, VCN-Biotech, and BeiGene; Josep M Piulats has received research grants from Bristol Myers Squibb, Merck Shart & Dohme, Merck Serono, Janssen, and Astra Zeneca.

Funding:

Agency for Management of University and Research Grants (AGAUR) of the Catalan Government grant 2017SGR723; Instituto de Salud Carlos III, co-funded by FEDER funds –a way to build Europe– grants PI14-00613, PI17-00092.

Author Contributions:

RSP and JMP designed the study, conceived the experiments and wrote the article. SGM and FMN carried out the sequencing analyses. SCO and JPV performed wet lab experiments. VM provides biostatistics expertise. VM and JPV helped to draft the manuscript. All authors critically reviewed and had final approval of the article.

Acknowledgments:

We thank Josipa Bilic for critical reading of the manuscript and English language assistance. We thank Spanish Association Against Cancer (AECC) Scientific Foundation.

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;15:763–76.
2. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science. 2008;319:1096–100.
3. Fitzgerald TL, Dennis S, Kachare SD, Vohra NA, Wong JH, Zervos EE. Dramatic Increase in the Incidence and Mortality from Merkel Cell Carcinoma in the United States. Am Surg. 2015;81:802–6.
4. Shuda M, Arora R, Kwun HJ, Feng H, Sarid R, Fernández-Figueras M-T, 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;125:1243–9.
5. Harms PW, Vats P, Verhaegen ME, Robinson DR, Wu Y-M, Dhanasekaran SM, et al. The Distinctive Mutational Spectra of Polyomavirus-Negative Merkel Cell Carcinoma. Cancer Res. 2015;75:3720–7.
6. González-Vela MDC, Curiel-Olmo S, Derdak S, Beltran S, Santibañez M, Martínez N, et al. Shared Oncogenic Pathways Implicated in Both Virus-Positive and UV-Induced Merkel Cell Carcinomas. J Invest Dermatol. 2017;137:197–206.
7. Wong SQ, Waldeck K, Vergara IA, Schröder J, Madore J, Wilmott JS, et al. UV-Associated Mutations Underlie the Etiology of MCV-Negative Merkel Cell Carcinomas. Cancer Res. 2015;75:5228–34.
8. 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;7:3403–15.
9. 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;137:819–27.
10. Colunga A, Pulliam T, Nghiem P. Merkel Cell Carcinoma in the Age of Immunotherapy: Facts and Hopes. Clin Cancer Res. 2018;24:2035–43.
11. Nghiem P, Bhatia S, Lipson EJ, Sharfman WH, Kudchadkar RR, Brohl AS, et al. Durable Tumor Regression and Overall Survival in Patients With Advanced Merkel Cell Carcinoma Receiving Pembrolizumab as First-Line Therapy. J Clin Oncol. 2019;37:693–702.
12. Sun Z, Wang L, Eckloff BW, Deng B, Wang Y, Wampfler JA, et al. Conserved recurrent gene mutations correlate with pathway deregulation and clinical outcomes of lung adenocarcinoma in never-smokers. BMC Med Genomics. 2014;7:32.
13. Ahn JW, Kim HS, Yoon J-K, Jang H, Han SM, Eun S, et al. Identification of somatic mutations in EGFR/KRAS/ALK-negative lung adenocarcinoma in never-smokers. Genome Med. 2014;6:18.
14. Andrews S. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. 2010;
15. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

16. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.
17. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
18. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6.
19. Knepper TC, Montesion M, Russell JS, Sokol ES, Frampton GM, Miller VA, et al. The Genomic Landscape of Merkel Cell Carcinoma and Clinicogenomic Biomarkers of Response to Immune Checkpoint Inhibitor Therapy. Clin Cancer Res. 2019;25:5961–71.
20. Kayama K, Watanabe S, Takafuji T, Tsuji T, Hironaka K, Matsumoto M, et al. GRWD1 negatively regulates p53 via the RPL11-MDM2 pathway and promotes tumorigenesis. EMBO Rep. 2017;18:123–37.
21. Hashida Y, Imajoh M, Daibata M. Integrated and mutated forms of Merkel cell polyomavirus in non-small cell lung cancer. Br J Cancer. 2013;108:2624.
22. 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. The American Journal of Surgical Pathology. 2009;33:1378.
23. Babakir-Mina M, Ciccozzi M, Lo Presti A, Greco F, Perno CF, Ciotti M. Identification of Merkel cell polyomavirus in the lower respiratory tract of Italian patients. J Med Virol. 2010;82:505–9.

Figures

Figure 1

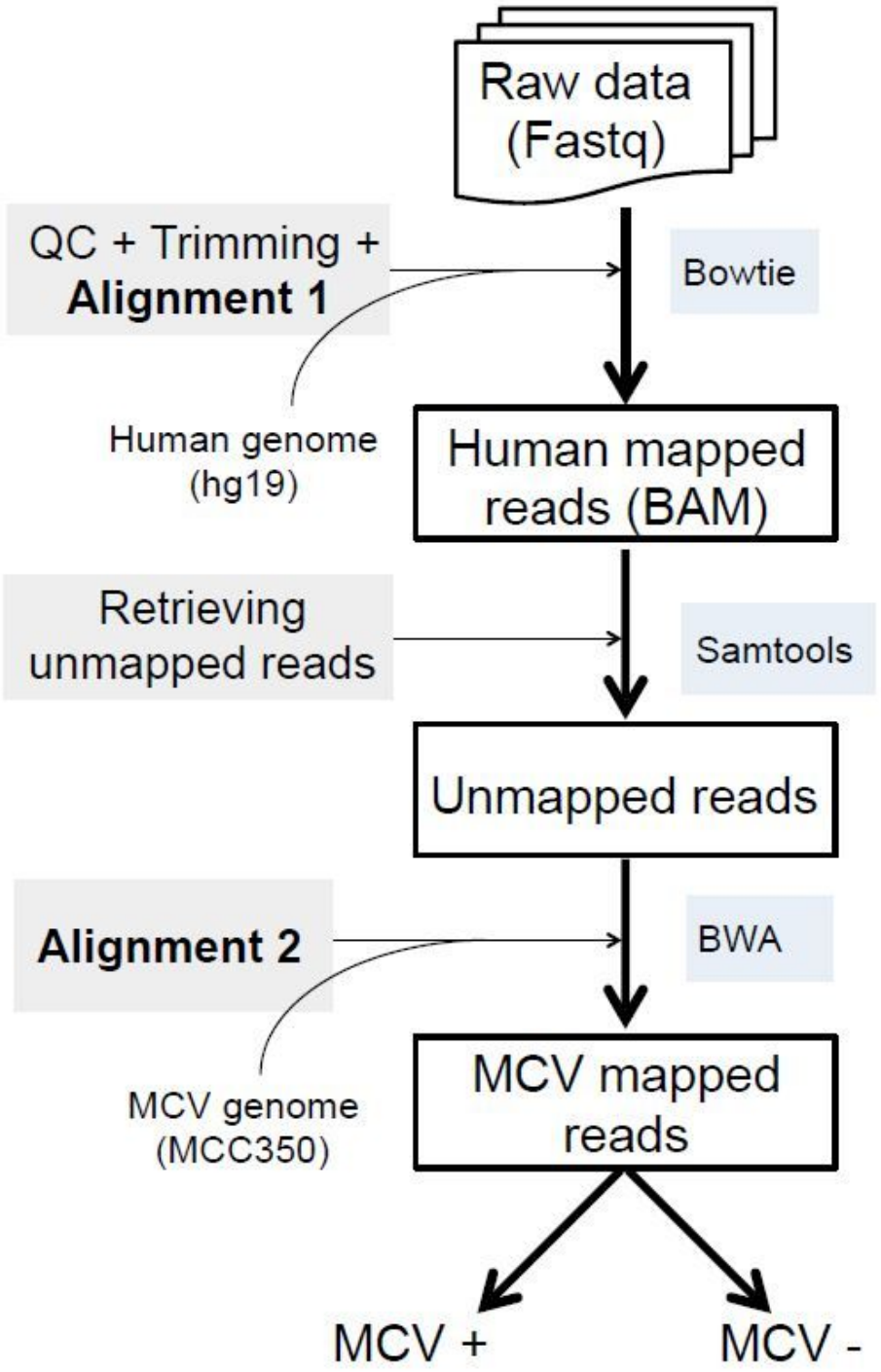

Figure 1

Pipeline for Merkel virus insertion detection using WES data. First, raw data were trimmed if necessary with TrimGalore (v0.4.0) and aligned against the human reference genome build hg19/GRCh37, previously indexed, using Bowtie 2.0 (v2.2.5) —alignment 1—. From this first alignment, unmapped reads are selected with Samtools (v1.3.1). Function Samtools view, with -f and -F options, is used to select and filter the desired reads. The online utility “Decoding SAM flags" from the Broad Institute

(https://broadinstitute.github.io/picard/explain-flags.html) was used to decide the filtering criteria. The selection criteria filters: i) unmapped reads whose mate are mapped (-f 4 -F 264), ii) mapped reads whose mate is unmapped (-f 8 -F 260), iii) both unpaired reads (-f 12 -F 256). The three outputs are merged and next, those retrieved unmapped reads are aligned against MCV genome (5,381 bp length, downloaded from the NCBI —GenBank accession number EU375803—) with BWA (v0.7.15) —alignment 2—. Finally, the number of reads aligned to virus is calculated, and an output is generated with the following information: Sample ID, MCV+/- status, and read counts in case of MCV+.

Figure 2

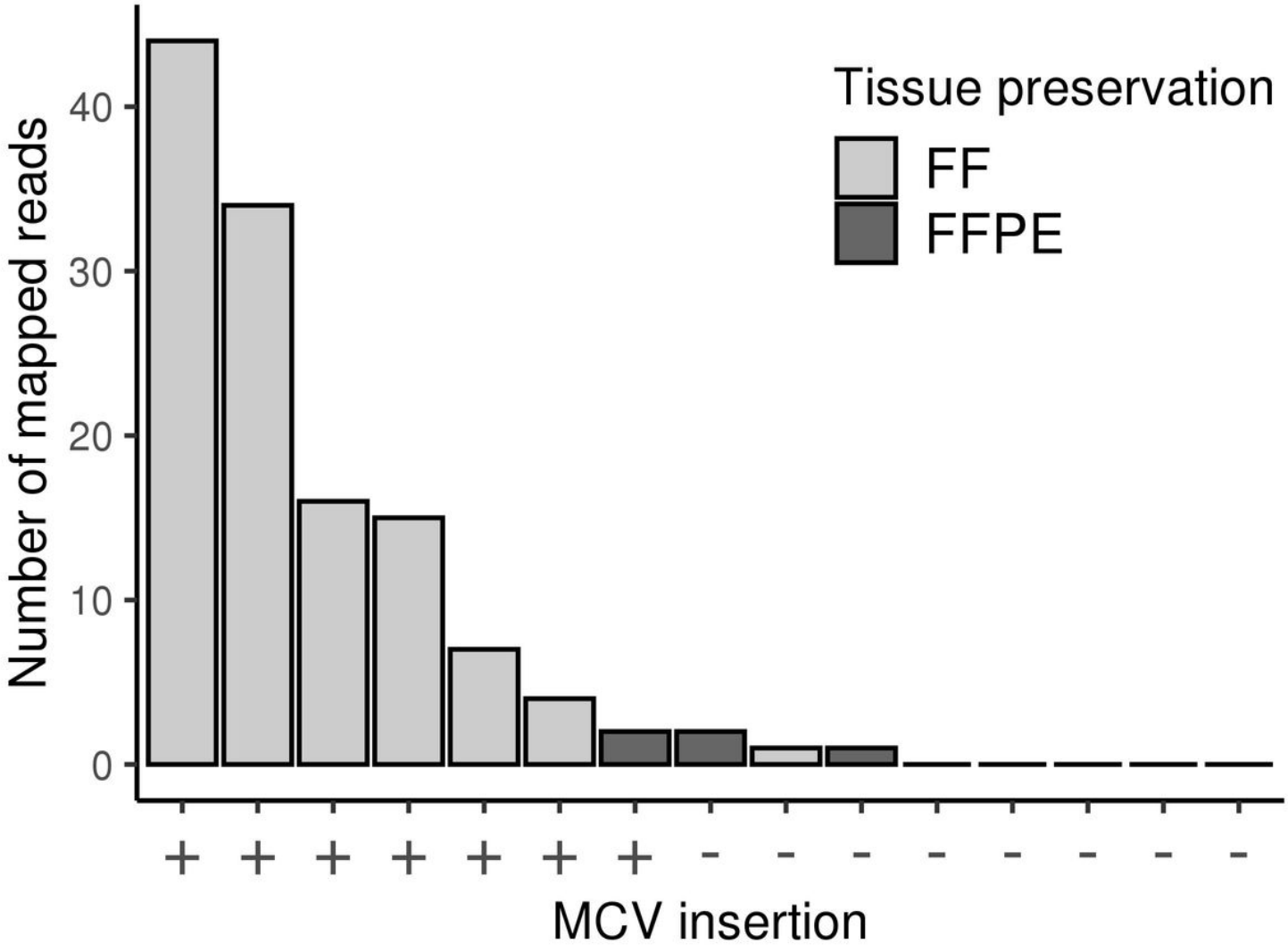

Figure 2

Barplot representing the number of mapped reads by sample. Fresh frozen (FF) samples are colored in light grey and formalin fixed paraffin embedded (FFPE) samples are colored in dark grey. In the x axis, MCV positive (+) and MCV negative (-) MCC tumors are indicated.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.tif
- supplement2.pdf
- supplement3.pdf