Appendix e-1

EBNA2 amplicon generation and deep-sequencing

The EBNA2 amplicons were generated using a multistep PCR and the Illumina Nextera strategy. In the 1st PCR, 30 ng of DNA were amplified using the EBNA2 type specific primers E2C and E2SEQ4.[e1] Then, 2,5 µl of the 1st PCR were used as templates in a 2nd nested PCR step adopting specific overhangs primer pairs. The primer overhangs used named Next_E4C_For and Next_E2A2mod_Rev (Next_E4C_For-5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG[CGTCGGGCATGGACCTCTA]3' and Next_E2A2mod_Rev 5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG[TGGTTCTGGACTATCTGGATCAT]-3') contain (from 5' to 3'ends) the Nextera's protocol adaptor sequences followed by the specific EBNA2 priming sequences (underlined nucleotides). The products of the 2nd nested PCR (~530 bp) were visualized on 1.2% agarose gel and purified using the AMPure XP Beads at a concentration of 0.8X vol/vol (Agencourt Bioscience Corporation, Beverly, Massachusetts). Fourty ng of the purified amplicons were used as templates in a final PCR step using Nextera index primers (index1 and 2), a primer mix containing the Illumina P5 and P7 primers (0.2 µM of each primer), and the high fidelity Phusion DNA Polymerase system (error rate 1/150,000). The dual indexed amplicons obtained (~620 bp) were purified using magnetic beads AMPure XP (0.8X, vol/vol), checked for quality control on Bioanalyzer 2100 (Agilent Technologies, Santa Clara, California) and quantified by fluorimetry using the Quant-iTTM PicoGreen® dsDNA Assay Kit (Invitrogen, Carlsbad, California). All the amplification steps were performed using the Phusion® High-Fidelity DNA polymerase system (Thermo Fisher Scientific, Inc., New England Biolabs) and 0.5 µM of each primer. All PCR reactions were performed in the presence

of a negative (Molecular Biology Grade Water, RNase/DNase-free water) and a positive (genomic DNA from *B95-8*) control. Finally, equimolar ratios of the dual indexed purified amplicons were pooled and subjected to 2x250 pb paired-end sequencing on the MiSeq platform.

EBNA2 variant detection and analysis from amplicon deep-sequencing data

Raw data (in FASTQ format) obtained by MiSeq deep sequencing from each sample were initially checked by means of FASTQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The automatic trimming of low-quality ends from reads was performed using Trim Galore, setting the minimal Phred score value to 20 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Cleaned FASTQ files were aligned onto the complete human herpesvirus 4 genome strain B95-8 (downloaded from GenBank with accession NC_007605) using GSNAP software[e2] and resulting SAM files were converted in the binary BAM files by means of SAMtools.[e3]

Individual amplicon positions were explored using REDItools[e4] in order to obtain textual summary tables including the base distribution per site, the mean quality, the nucleotide change (if any) and its relative frequency. REDItools were launched fixing the minimal Phred quality score of a base to 25 and requiring only concordant read pairs.

Viral allele assignment was performed using an ad hoc python script (available upon request). In brief, the script reads in a table containing per each potential viral allele all representative positions allocating them in appropriate arrays. Then, browsing each individual position of an input BAM file, it looks for candidate sites stored in the arrays and calculates the frequency distribution of all four bases. The relative probability to observe a given viral allele is then calculated multiplying the nucleotide frequency per each position of the allele. In order to assess if the detected variants were either indicative of possible coinfections or simply random

recurring sequencing errors we considered all reads carrying two variants and performed a chi-square test to evaluate the null hypothesis of independent segregation.

The most likely viral allele for Sanger reads was detected using the same script described above. Before the viral assignment, Sanger reads were aligned onto the human herpesvirus 4 genome strain B95-8 using GMAP[e3] and saving the output in the standard SAM format.

EBNA2 sequencing in post-mortem brain tissue

The study was performed using post-mortem brain tissue from a 44-year-old female with secondary progressive MS (disease duration 19 years, age at wheelchair 34 years, EDSS 7) who died of pneumonia. Brain tissue was provided by the UK Multiple Sclerosis Tissue Bank at Imperial College London (http://www.ukmstissuebank.imperial.ac.uk). Based on the available clinical documentation no treatment is reported in the 6 months before death. One snap-frozen tissue block (4 cm^3) was selected for laser capture microdissection (LCM) on the basis of the inflammatory degree, as assessed by immunohistochemistry.[e5] Ten serial brain sections were cut and mounted on membrane slides for LCM (MMI AG, Glattbrugg, Switzerland), air-dried, fixed, dehydrated and stained to visualize the areas of interest, as previously described.[e5] Using a laser microdissector SL Cut (MMI AG) equipped with a Nikon Eclipse TE2000-S microscope, large B cell-enriched meningeal immune infiltrates were selectively cut from all the serial sections and pooled in a single cap (total microdissected area = 1.470.000 μm^2). The isolated tissue fragments were immediately stored at- 80°C. DNA was extracted using QIAmp DNA micro kit (Qiagen). Sequence analysis was performed using the Illumina MiSeq platform, as described in the Methods section.

e-References

e1 Tierney RJ, Edwards RH, Sitki-Green D, et al. Multiple Epstein-Barr virus strains in patients with infectious mononucleosis: comparison of ex vivo samples with in vitro isolates by use of heteroduplex tracking assays. J Infect Dis. 2006; 193:287-297.

e2 Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 2010;26:873-881.

e3 Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078-2079.

e4 Picardi E, Pesole G. REDItools: high-throughput RNA editing detection made easy. Bioinformatics. 2013;29:1813-1814

e5 Serafini B, Severa M, Columba-Cabezas S, C, et al. Epstein-Barr virus latent infection and BAFF expression in B cells in the multiple sclerosis brain: implications for viral persistence and intrathecal B-cell activation. J Neuropathol Exp Neurol. 2010;69:677-693.