

# Identifying Epstein-Barr virus EBNA-1 sequence variation using 454 FLX technology in Multiple Sclerosis patient samples

Tschochner M.<sup>1</sup>, Leary S.<sup>1</sup>, Strautins K.<sup>1</sup>, Clark H.<sup>1</sup>, Chopra A.<sup>1</sup>, Choo L.<sup>1</sup>, Carroll W.<sup>2</sup>, Kermode A.<sup>2</sup>, Nolan D.<sup>1,3</sup>



<sup>1</sup> Institute for Immunology & Infectious Diseases, Murdoch University, Perth, Western Australia

<sup>3</sup> Sir Charles Gairdner Hospital, Nedlands, Perth, Western Australia

<sup>2</sup> Department of Clinical Immunology & Immunogenetics, Royal Perth Hospital, Perth, Western Australia;



## Introduction

Both genetic and environmental factors contribute to Multiple Sclerosis (MS) risk. Susceptibility to MS is conferred by immune system associated genes, with human leukocyte antigen (HLA) DRB1\*1501 identified as a major MS risk allele. Infection with Epstein-Barr Virus (EBV) has also been shown to increase MS risk, and in particular anti-EBV nuclear antigen-1 (EBNA-1) antibodies are strongly associated with the disease. Additionally, it has been shown that EBNA1-specific CD4+ T-cells can cross-recognize MS-associated myelin antigens, and are selectively expanded in MS patients.

We therefore sought to assess MS patient-specific EBNA-1 sequence variation, including the potential influence of minority sequence variants, in shaping immune responses to known and predicted HLA-DRB1\*1501-restricted epitopes.

## Methods

MS samples were utilised from the Perth Demyelinating Disease Database (PDDD) (Western Australia). DNA was isolated from buffy coats and the C-terminal end of the EBNA-1 gene was amplified using a novel semi-nested PCR approach with a fully automated setup utilising Biomek FX robots. First round PCR was performed using Roche High Fidelity Taq in a 25µl reaction with primers EBV109111F-EBV109951R, resulting in a 840bp fragment. A semi-nested PCR was followed using the primer combination EBV109111F-EBV109869R resulting in a final 749 base pair product (Fig.1). Successful PCR samples were purified using AMPure and Sanger sequenced on an automated 96 capillary ABI 373 DNA sequencer, followed by analysis using the ASSIGN V4.0.1.36 software (Conexio Genomics) (Fig.2). Threshold for mixture detection in sanger sequencing has been established to be ~20%.

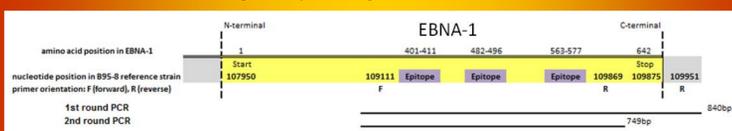


Fig.1: Semi-nested, C-terminal EBNA-1 PCR. Numbers reflect the position of the primers in B95-8.

To study sequence minorities in quasispecies present at  $\geq 1\%$ , ultra-deep sequencing was performed using the FLX-454 Roche technology for a subset of 23 MS patients and a B95-8 control. The same first round PCR products obtained were used for high throughput sequencing and Sanger sequencing approaches. For the second round semi-nested PCR, primers were 5' end tagged with 10 nucleotides of unique sequence (one each primer pair and sample) to enable pooling of samples in one library. Samples were purified using AMPure and concentrations measured with Nanodrop. Copy number per sample was calculated and  $1 \times 10^{11}$  copies per sample were pooled. After equimolar pooling, the combined samples were purified with minielute spin column (Qiagen). Final concentration for the pooled fragments was 150.4 ng/ul. One library of all pooled samples was prepared and 454 sequenced (Fig.2).

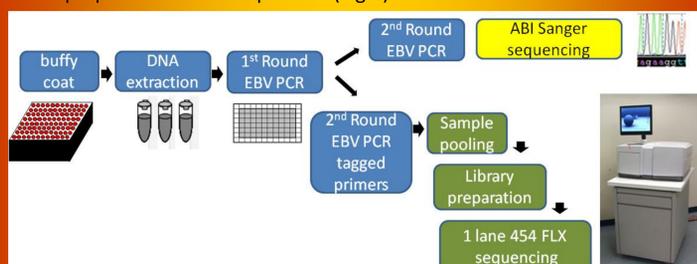


Fig. 2: workflow sanger and 454 FLX sequencing

For FLX data analysis all homopolymers not present in the consensus sequence have been excluded. Minorities present at  $< 1\%$  were not taken into consideration and insertion deletions were also excluded.

## Results

EBV could successfully be amplified in 72 samples without previous culture to enrich for EBV episomes. EBV Sanger sequences clustered in four main groups. Interestingly, most of the samples did not cluster with the EBV reference strain B95-8 (Fig.3). Genotyping of position 487 of EBNA-1 identified the strain 487-threonine (n=52) and 487-alanine (n=14) as most prevalent, with alanine also present in the B95-8 reference strain (Table 1).

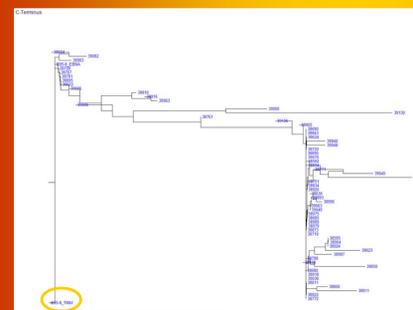


Fig. 3: Phylogenetic tree of C-terminal EBNA-1 MS sequences. Circled is the EBV reference strain B95-8

codon	amino acid	n
ACA	Threonine	1
ACT	Threonine	51
CTC	Leucine	4
GAA	Glutamate	1
GCT	Alanine	14
GTT	Valine	1

Table 1: Genotyping position 487 of the EBNA-1 gene in 72 samples.

14 samples tested on FLX had high coverage with average reads between 80-1632 for the C-terminal EBNA-1 fragment. 10 samples had low amount of coverage and or reads (average reads  $< 15$ ). Four of these samples (38585, 38695, 38939, 38593) were successful for PCR amplification but could not be sequenced with conventional Sanger sequencing technology. Unmatched contigs were blasted and were found to represent chromosomal DNA.

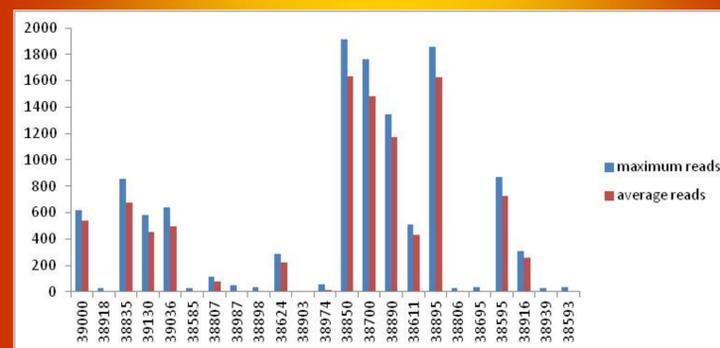


Fig. 4: Maximum and average reads for 24 MS samples tested on 454 FLX.

Amino acid variability was low and Sanger sequencing results revealed no mixed positions. Utilizing 454 FLX deep sequencing, minor species could be detected at  $>10\%$  frequency in 2.9% (21 positions of all 714 positions investigated), increasing to 25.5% of all positions investigated with minor species detected at  $\geq 1\%$  (Table 2).

At a 1% cut off, 105 positions contained mixed amino acids for one sample only whereas in 30 different positions there were at least 2 samples with an amino acid mixture in the same position (Table 2).

	samples with mutation in the same position														mixed positions	% mixed positions
	1	2	3	4	5	6	7	8	9	10	11	12	13	14		
1% positions with mutation	105	30	15	11	6	4	0	2	1	4	0	3	1	0	182	25.5%
2% positions with mutation	35	17	3	3	4	2	0	1	1	0	2	1	0	0	69	9.7%
5% positions with mutation	21	7	0	1	1	1	1	0	0	0	0	0	0	0	32	4.5%
10% positions with mutation	19	1	1	0	0	0	0	0	0	0	0	0	0	0	21	2.9%

Table 2: Mixed amino acid positions for 14 samples with high coverage tested on 454 FLX. Minorities at 1%, 2%, 5% and 10% cut off. Percentage of mixed positions was calculated in relation to the 714 investigated nucleotides of EBNA-1.

Epitope coding sequence variability was directly compared using 14 samples sequenced with Sanger sequencing and 454 FLX technology for three previously described epitopes: a putative B cell epitope GRRPFFHPVGE (aa 401-411), and two HLA-DRB1\*1501 epitopes (AEGRLALLARSHVER (aa 482-496), MVFLQTHIFAEVLKD (aa 563-477)).

As shown in Figure 5, both methods were concordant in detecting sequence variation potentially relevant to epitope recognition and presentation, with only a minority of sequences sharing identity with the B95-8 reference strain. 454 FLX technology identified mixtures of  $>1\%$ , although generally at a low percentage with no mixtures  $> 10\%$  within the epitopes.

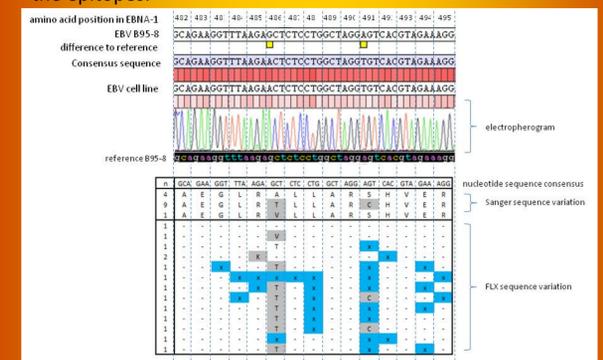


Fig. 5: Comparison of HLA-DRB1\*1501 class-II epitope variation between Sanger and 454 FLX sequencing (epitope position in EBNA-1: 482-496) for 14 samples. Blue boxes = mixtures in 454 FLX sequencing. Grey boxes = variation to reference sequence B95-8. n = number of samples with this epitope.

## Discussion

In this study, both sanger and 454 FLX approaches were utilised to identify EBNA-1 sequence variation potentially relevant to immune response in MS patients.

Sanger sequencing revealed no mixed positions but identified different viral clusters mainly divergent from the reference strain B95-8.

Using the sensitive 454 FLX technology, several minority species could be detected. Additionally, four samples could be sequenced using 454 FLX that could not be sequenced with sanger sequencing technology. However, these samples preferentially amplified chromosomal DNA with low EBV reads, suggesting low EBV viral load in these samples.

Most quasispecies were present at less than 1%, consistent with known limited dsDNA viral sequence variation.

Analysing sequence variation in previously described EBNA-1 epitopes demonstrated concordant results using these two methods, with limited evidence of viral quasispecies.

## Conclusions

This study confirms the feasibility of direct amplification and sequencing of EBNA-1 from patient samples in identifying patient-specific sequence variation, with evidence of high level of concordance between sanger and FLX 454 sequencing techniques.

The frequent detection of sequences that are divergent from B95-8 reference strain suggests that studies investigating epitope-specific immune responses can be usefully guided by autologous patient sequence information.

Low-level sequence variation was identified by sensitive FLX technology but was generally rare. It remains to be determined if these minority species affect EBV-specific immune responses or disease outcomes.

## Acknowledgments

We would like to thank all patients and study participants, all nurses and doctors involved as well as staff at IID for their assistance and input.

This study is funded by the McCusker Charitable Foundation