Multiple Sclerosis Patient Sequence Variation In Epstein-Barr Virus Nuclear Antigen 1 Using 454 FLX technology

Tschochner M.[1], Leary S.[1], Strautins K.[1], Clark H.[1], Cooper D.[1], Chopra A.[1], Choo L.[1], Carroll W.[2,3], Kermode A.[1,2,3], Nolan D.[1,4]

[1]*Institute for Immunology & Infectious Diseases, Murdoch University, Perth, Western Australia* [2]*Department of Neurology, Sir Charles Gairdner Hospital, Queen Elizabeth II Medical Centre, Western Australia ,* [3]*Centre for Neuromuscular and Neurological Disorders, Australian Neuromuscular Research Institute, UWA, Australian Neuromuscular Research Institute, Sir Charles Gairdner Hospital, University of Western Australia, Perth, Western Australia*, [4]*Department of Clinical Immunology, Royal Perth Hospital, Perth, Western Australia*

Background:

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.

Objective:

To identify MS patient-specific EBNA-1 sequence variation and the potential influence of minority variants in shaping epitope-specific immune responses.

Materials and 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 (*Fig 1*). First round PCR was performed with the primers EBV109111F-EBV109951R and Expand High Fidelity Taq (Roche) in a 25µl reaction, 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 3730XL DNA sequencer, followed by analysis with the ASSIGN V4.0.1.36 software (Conexio Genomics) (*Fig 2*). Threshold for mixture detection in sanger sequencing has been established to be ~20%.

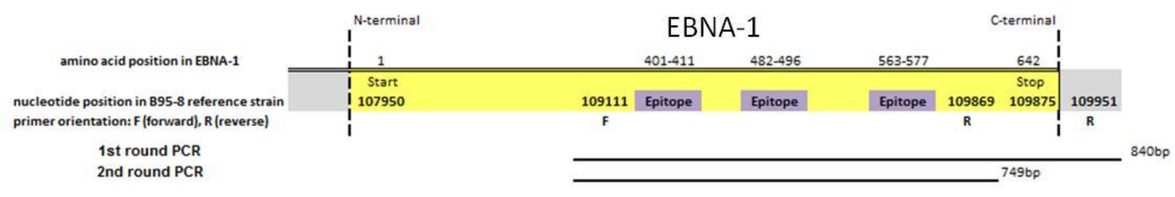

Fig 1. Semi-nested, C-terminal EBNA-1 PCR. Numbers reflect the position of the primers in the B95-8 reference sequence.

To study sequence minorities in quasispecies present at ≥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 for the sanger sequencing approach were utilized for the high throughput sequencing approach. Second round primers contained an additional tag on the 5' end composing of 11 nucleotides of unique sequence (one each primer pair and sample) to enable pooling of samples in one FLX library. Samples were purified using AMPure and concentrations measured with Nanodrop. Copy number per sample was calculated and $1x10^{11}$ copies per sample were pooled. After equimolar pooling, the combined samples were purified and concentrated to 150.4 ng/µl using a minielute spin column (Qiagen). One library of all pooled samples was prepared and sequenced with 454 FLX in one lane out of eight (*Fig 2*) .

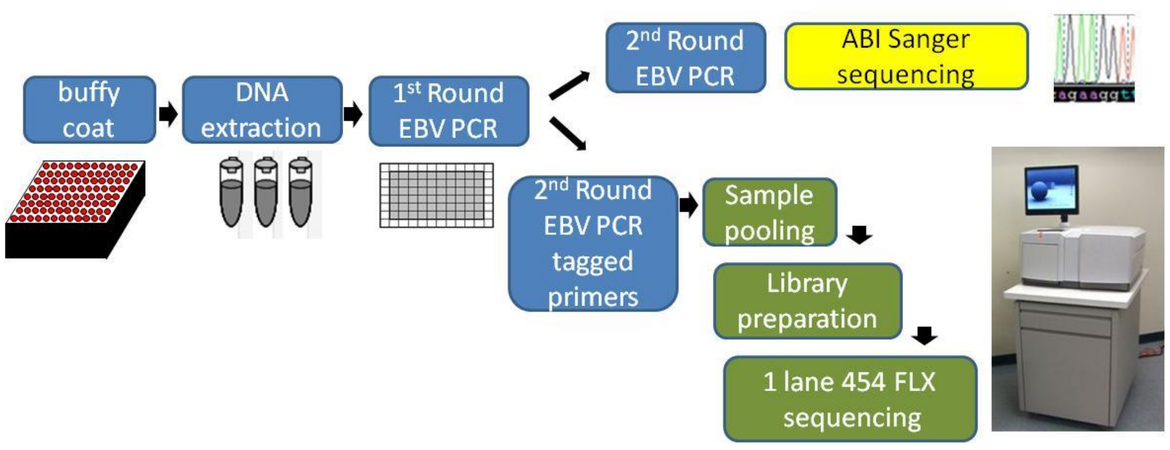

Fig 2. Workflow comparison sanger and 454 FLX sequencing

For FLX data analysis all homopolymers not present in the majority consensus sequence of the patient sample have been excluded. All 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 well with the EBV reference strain B59-8 (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).

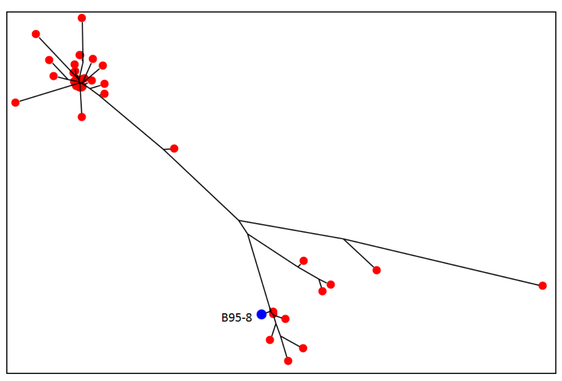

Fig. 3: Phylogenetic tree of C-terminal EBNA-1 patient sequences and the B95-8 reference sequence.

genotyping position EBNA-1: 487

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 amino acid position 487 of the EBNA-1 gene in 72 samples.

14 samples tested with 454 FLX had high coverage of the C-terminus with average reads between 80-1632. 10 samples had low coverage and or reads (average reads <15) (*Fig 4*). Four samples (38585, 38695, 38939, 38593) which could be amplified but not Sanger sequenced were successfully sequenced using FLX. Detailed analysis showed high amount of contigs representing chromosomal DNA but all showed also low frequency EBV reads of unique sequence.

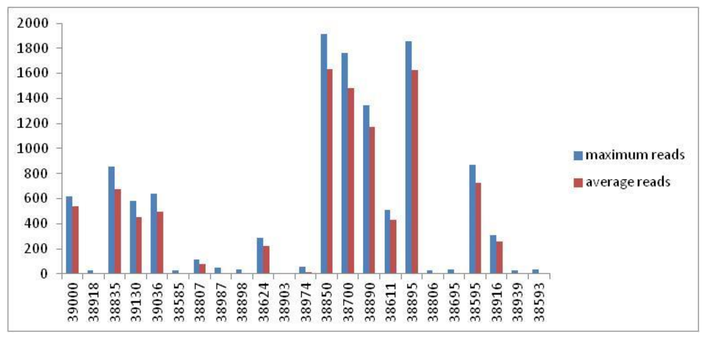

Fig 4. Maximum and average reads for 24 MS samples tested on 454 FLX in one lane out of eight.

Overall, amino acid variability was low and Sanger sequencing results revealed few mixed positions. Utilizing 454 FLX deep sequencing, quasispecies could be detected, although generally at a low percentage. In the high coverage samples minorities at ≥10% frequency were detected in 3.5% (26 positions of 749 positions investigated), increasing to 7.3% for minor species detected at ≥1% (Table 2). At a 1% cut off, 55 of 749 positions contained mixed amino acids across all 14 samples. 41 of these occurred in a single sample only but 12 different positions had mutations which were shared by two samples and 2 positions were shared by three samples (Table 2).

Samples with mutation in same position	1	2	3	total mutations
minorities 1%	41	12	2	55
minorities 2%	21	13	1	35
minorities 5%	23	6	1	30
minorities 10%	20	6	0	26

Table 2. Mixed amino acid positions for 14 samples with high coverage tested on 454 FLX. Minorities present at 1%, 2%, 5% and 10% cut off.

Epitope coding sequence variability was directly compared using 14 samples sequenced with Sanger and 454 FLX sequencing technology for three previously described epitopes: a putative B cell epitope GRRPFFHPVGE (aa 401-411), and two HLA-DRB1*1501 epitopes (AEGLRALLARSHVER (aa 482-496) and MVFLQTHIFAEVLKD (aa 563-477). In our cohort, the most frequent variants were identified as PPPGRRPFFHPVGEAD, AEGLRTLLARCHVER (*Fig. 5*) and IV/AFLQTHIFAEGLK. Binding scores (NetMHCIIpan) of patient derived epitope sequences revealed that FLX technology identified additional low frequency variants with HLA-DRB1*1501 restriction.

Both sequencing 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.

codons B95-8	gca	gaa	ggt	tta	aga	gct	ctc	ctg	gct	agg	agt	cac	gta	gaa	agg	
aa B95-8	A	E	G	L	R	A	L	L	A	R	S	H	V	E	R	samples
FLX	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Sanger	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
FLX	-	-	-	-	-	GTT	-	-	-	-	-	-	-	-	-	
Sanger	-	-	-	-	-	GTT	-	-	-	-	-	-	-	-	-	1
FLX	-	-	-	-	AAA	-	-	-	-	-	-	-	-	-	-	
Sanger	-	-	-	-	AAA	-	-	-	-	-	-	-	-	-	-	1
FLX	-	-	-	-	-	RCT	-	-	-	-	WGT	-	-	-	-	
Sanger	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
FLX	-	-	GGY	-	-	ACT	-	-	-	-	TGT	-	-	-	-	
Sanger	-	-	-	-	-	ACT	-	-	-	-	TGT	-	-	-	-	1
FLX	-	-	-	-	-	RCT	-	-	-	-	WGT	-	-	-	-	
Sanger	-	-	-	-	-	ACT	-	-	-	-	TGT	-	-	n.a.	-	1
FLX	-	-	-	-	-	ACT	-	-	-	-	TGT	-	-	-	-	
Sanger	-	-	-	-	-	ACT	-	-	-	-	TGT	-	-	-	-	7

Fig. 5. Epitope sequence variability in the AEGLRALLARSHVER epitope detected by FLX.
grey: difference to reference strain B95-8
blue: mixtures detected with FLX

Discussion and Conclusions:

This study confirms the feasibility of direct EBNA-1 amplification from patient samples. Patient derived sequences were divergent from the B95-8 reference strain and clustered in four close groups. FLX sequencing technology could detect additional low-level sequence variants but they were generally rare, consistent with known limited dsDNA viral sequence variation. Of note, four samples could be sequenced using 454 FLX which 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.

Analysing sequence variation in previously described EBNA-1 epitopes demonstrated concordant results between Sanger and FLX 454 sequencing techniques, with limited evidence of viral quasispecies. 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.

In conclusion, 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.

PACTRIMS
November 2013

Acknowledgments:
We would like to thank all patients and study participants, all nurses and doctors involved as well as staff at IIID for their assistance and input.
This study is funded by the McCusker Charitable Foundation

Presenting author contact:
m.tschochner@iiid.com.au