Check for updates

DOI: 10.1111/jdv.15817 JEADV

ORIGINAL ARTICLE

Inter-rater reliability of the BIOCHIP indirect immunofluorescence dermatology mosaic in bullous pemphigoid and pemphigus patients

A. Yang,[1,2] R.R. Xuan,[1,†] W. Melbourne,[2] T. Hashimoto,[3] S. Uzun,[4] M. Daneshpazhooh,[5] J. Yamagami,[6] G. Di Zenzo,[7] J.M. Mascaro Jr,[8] H. Mahmoudi,[5] A. Patsatsi,[9] K. Drenovska,[10] S. Vassileva,[10] D.F. Murrell[1,2,*]

[1]University of New South Wales, Kogarah, NSW, Australia
[2]Department of Dermatology, St George Hospital, Sydney, NSW, Australia
[3]Department of Dermatology, Osaka City University Graduate School of Medicine, Osaka, Japan
[4]Department of Dermatology, Akdeniz University School of Medicine, Antalya, Turkey
[5]Autoimmune Bullous Diseases Research Center, Tehran University of Medical Sciences, Tehran, Iran
[6]Department of Dermatology, Keio University School of Medicine, Tokyo, Japan
[7]Molecular and Cell Biology laboratory, IDI-IRCCS, Rome, Italy
[8]Hospital Clinic and Barcelona University Medical School, Barcelona, Spain
[9]2nd Dermatology Department, Aristotle University School of Medicine, Thessaloniki, Greece
[10]Department of Dermatology and Venereology, Sofia University of Medicine, Sofia, Bulgari,
*Correspondence: D.F. Murrell. E-mail: d.murrell@unsw.edu.au

Abstract

Background The BIOCHIP (Dermatology Mosaic 7; EUROIMMUN, Lubeck, Germany) is a novel multiplex indirect immunofluorescence (IIF) technique used in the serological diagnosis of bullous pemphigoid (BP) and pemphigus.

Objective To validate the accuracy and inter-rater reliability (IRR) of the BIOCHIP in the diagnosis of BP, pemphigus foliaceus (PF) and pemphigus vulgaris (PV).

Methods Sera from patients with BP (n = 38), PF (n = 8), PV (n = 23), control patients (n = 64) and healthy control volunteers (n = 39) were tested. Sera were collected and analysed during the course of the disease at 1–5 different time points. The BIOCHIP was performed for all patients, digital images were captured of each incubated field, and the images were shared with 10 dermatologists experienced in reading IF from around the world to report. There were 312 BIOCHIP slides consisting of 1872 photos in total. All patients were de-identified. Fleiss Kappa was used to estimate the IRR.

Results Fleiss Kappa was computed for each category (Oesophagus, Oesophagus immunofluorescence pattern, Salt-Split Skin (SSS), SSS immunofluorescence location, BP180, BP230, Dsg 1 and Ds3). The inter-rater agreement between the 10 raters varied between fair and moderate for all categories. Those that demonstrated fair concordance included monkey oesophagus (k = 0.257, P < 0.0001), oesophagus pattern (k = 0.357, P < 0.0001), Dsg1 (k = 0.390, P < 0.0001) and BP230 (k = 0.281, P < 0.0001). Moderate agreement was demonstrated for SSS (k = 0.416, P < 0.0001), SSS immunofluorescence location (k = 0.505, P < 0.0001), Dsg3 (k = 0.437, P < 0.0001) and BP180 (k = 0.559, P < 0.0001).

Conclusion The BIOCHIP mosaic-based immunofluorescence test is a simple, time and effort saving test that can aid in the diagnosis and screening of BP, PV and PF. However, the level of agreement was relatively low. The authors found the most common causes to be variable levels of training, indicating the presence of a learning curve in the interpretation of the results and ambiguous staining patterns leading to incongruent results.

Received: 24 February 2019; Accepted: 3 July 2019

Conflict of interest

None declared.

Funding source

None.

[†]Co-first author.

© 2019 European Academy of Dermatology and Venereology

Introduction

Bullous pemphigoid (BP), pemphigus vulgaris (PV) and pemphigus foliaceus (PF) are autoimmune bullous diseases (AIBD) that are potentially life-threatening and have significant negative impacts on the psychological health of patients. They are characterized by circulating and tissue-bound autoantibodies against BP180/BP230, Desmoglein (dsg) 3 or Dsg1, respectively. The timely and accurate identification of autoantibodies is important for patient management to reduce the severity of disease, morbidity and mortality.[1]

In recent years, the BIOCHIP (Dermatology Mosaic 7; EUROIMMUN) mosaic-based indirect immunofluorescence (IIF) technique has been developed as a new method for the diagnosis of AIBD. The design of the BIOCHIP enables differential diagnosis amongst the different AIBD subtypes from a single serum sample. This is possible through the presence of multiple antigenic structures in separate windows on a single incubation field.[2] These fields are interpreted under a fluorescent microscope in a similar way to traditional IIF investigations. The simultaneous processing of the most common AIBD autoantibodies using a single investigation such as the BIOCHIP can be useful to aid a timely and precise diagnosis.

We demonstrated the validity of the BIOCHIP, and we found that the BIOCHIP technique afforded better specificity than sensitivity in the diagnosis of BP, PV or PF.[3] The authors found the main limitation with the BIOCHIP was the accurate interpretation of the BIOCHIP slides. Despite the use of a control, the intensity of the fluorescence was variable and at times ambiguous. Some windows would display positive staining but only at the edge of a few cells, not sufficiently strong for a positive diagnosis yet not completely negative. Due to the rater-dependent nature of this investigation, in an unblinded clinical situation, this could lead to interpretation bias. It is important to measure inter-rater reliability (IRR) of the BIOCHIP to determine its utility in clinical practice. This is the first international study to analyse the inter-rater agreement of the BIOCHIP substrates using evaluations from several experienced dermatologists.

Bullous pemphigoid

Bullous pemphigoidis the most common autoimmune subepidermal bullous disease in the Western world. It is characterized by circulating immunoglobulin (Ig) G autoantibodies targeting the 180-kD BP antigen (BP180, BPAG2) and/or the 230-kD BP antigen (BP230 or BPAG1).[4,5] These two antigens are components of the hemidesmosomes, which are adhesion complexes promoting epithelial–stromal adhesion in stratified and other complex epithelia. Most patients with BP have circulating IgG autoantibodies that bind to the immunodominant non-collagenous NC16A extracellular domain of BP180.[6–8]

Pemphigus

Pemphigus is a life-threatening AIBD that presents clinically with blisters of the skin and mucous membranes.[9] The disease manifests in patients aged 20–40 years and PV most prevalent in people of Mediterranean or Jewish ancestry; however, individuals with certain human leucocyte antigen allotypes are also predisposed to the disease.[2,9] Pemphigus is mediated by autoantibodies targeting the 130kDa antigen (Dsg3) in patients with mucosal PV and 165kDa antigen (Dsg1) in patients with PF and mucocutaneous PV.[10–12] These two antigens are components of the desmosomal glycoproteins (cadherins) which connect neighbouring keratinocytes. This is in the suprabasal spinous and the subcorneal granular layer, respectively, leading to acantholysis and intraepithelial blisters.[13–15]

Methods/Subjects/Raters

Ethics

The study was approved by Bellberry Ethics. All subjects were over the age of 18 and informed consent was obtained.

Method

Twenty millilitres of blood was collected and centrifuged at 2.5 rates per minute for 10 min before being aliquoted into dilutions of 1 : 10 with saline buffer solution. All sera were stored at −20°C and anonymized before testing. BIOCHIP (Dermatology Mosaic 7; EUROIMMUN) was processed according to the manufacturer's instructions. The kits were provided free of charge by the manufacturers, and all sera were processed at St George Hospital, Sydney, Australia by a single immunopathologist.

Photographs of each substrate in each window on the BIOCHIP slide were taken using a mounted IMAGING MicroPublisher 5.0 RTV camera (OLYMPUS, Tokyo, Japan). The same settings were used for all the images captured. Individual photos of each substrate from 312 BIOCHIP slides produced a total of 1872 photos. The photos were magnified and displayed in a blinded document. Each photo was rated by each rater. Raters selected an option from a dropdown box alongside each image. The 'monkey oesophagus (MO)' window was rated as 'Positive', 'Negative' or 'Indeterminate'. The options for the type of immunofluorescence pattern were 'Desmosomal', 'Linear' or 'Indeterminate'. The 'Salt-Split Skin (SSS)' window was rated as 'Positive', 'Negative' or 'Indeterminate' and the location – 'epidermal/above', 'dermal/below', 'both' or 'indeterminate'. BP180, BP230, Dsg1 and Dsg3 were rated as 'Positive', 'Negative' or 'Indeterminate'. If raters selected indeterminate, they were asked for the reason.

Subjects

In total, 312 blood samples were collected and processed with the BIOCHIP (Dermatology Mosaic 7; EUROIMMUN).

© 2019 European Academy of Dermatology and Venereology

Cases Sera from patients with BP (n = 38), PF (n = 8), and PV (n = 23) were collected prospectively. The diagnosis was confirmed based on the routine multistep diagnostic algorithm: clinical presentation, appropriate findings of histopathology, and/or positive direct or IIF results, and/or serum autoantibodies against BP180, BP230, Dsg1 or Dsg3, respectively. A total of 209 blood samples were collected from the 69 cases. Samples were collected at 1–5 different time points over the course of the disease.

Disease controls Sera from patients (n = 64) with a variety of diseases different from BP, PV or PF were used as disease controls. Those in the diseased control group had a variety of dermatological conditions including linear IgA dermatosis, psoriasis, Sjogren's syndrome, epidermolysis bullosa acquisita (EBA), epidermolysis bullosa simplex, mucous membrane pemphigoid, junctional epidermolysis bullosa, porphyria, prurigo nodularis, urticarial pruritus and Galli–Galli disease.

Healthy volunteers Healthy volunteers (n = 39) were recruited if they were not treated for a dermatological disease.

Raters

Ten raters were included in this study. Each photo was rated by each rater. An invitation to participate in the study was sent to 21 dermatologists and one immunopathologist with BIOCHIP IIF experience. Those that responded (n = 14) were provided examples of control positive and negative BIOCHIP slides (Figs 1–3) and were required to demonstrate a priori level of knowledge by passing a test to correctly interpret standard slides before they were provided access to the subject images. Those from the same institution were given instructions to read the slides separately to avoid bias. Thirteen returned the standardizing test and were then granted access to the images. Ten raters completed evaluating all 312 BIOCHIP slides. Upon completion, the participants were asked to complete a questionnaire (Appendix I).

Statistics

All data were analysed using SPSS (IBM SPSS Statistics for Macintosh, Version 25.0. Armonk, NY, USA). Inter-rater agreement was tested using Fleiss Kappa statistical test that was appropriate for fully crossed designs with three or more coders.[16] The rating system used was a nominal rating system to limit the strain on IRR estimates. Fleiss Kappa was computed for category. Values of k of 0.21–0.40 are rated as a fair concordance, 0.41–0.60 as a moderate concordance, 0.61–0.80 as substantial concordance and 0.81–1.00 as almost perfect concordance. A P-value of <0.05 was considered statistically significant.[17]

Results

All 312 biochip slides consisting of 1872 photos were rated by 10 raters. Table 1 summarizes the kappa statistic for the inter-rater agreement on each individual substrate of the BIOCHIP. The

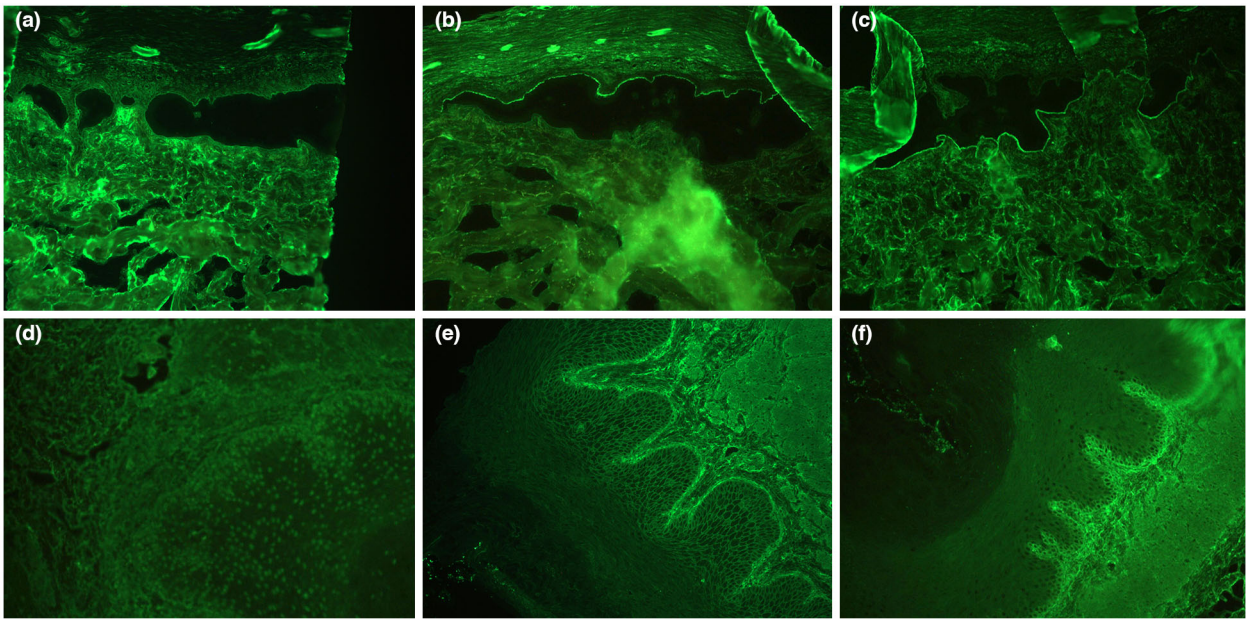

Figure 1 Normal staining patterns on salt-split skin (SSS) and monkey oesophagus substrate (MO). (a) Negative reaction on SSS. (b) Positive reaction above the blister. (c) Positive reaction below the blister. (d) Negative reaction on MO. (e) Positive reaction on MO showing desmosomal reactivity. (f) Positive reaction on MO showing linear pattern.

© 2019 European Academy of Dermatology and Venereology

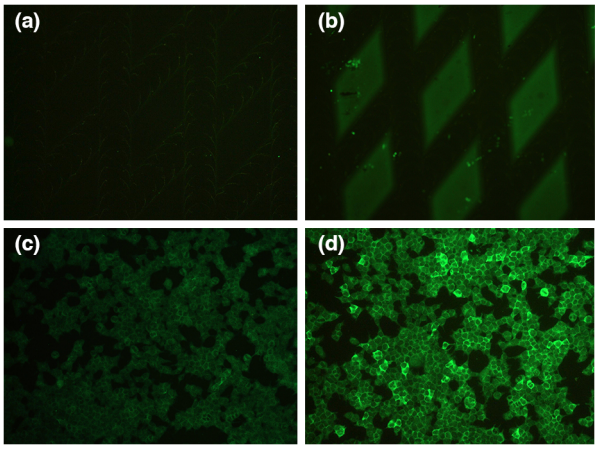

Figure 2 Normal staining patterns on BP180 and BP230 substrates. (a) Negative reaction for BP180. (b) Positive reaction for BP180. (c) Negative reaction for BP230. (d) Positive reaction for BP230.

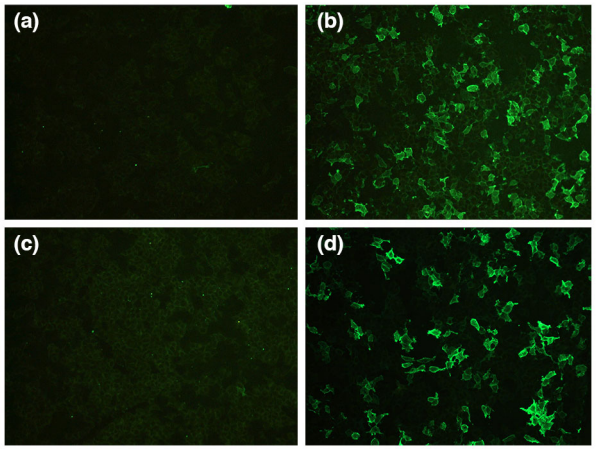

Figure 3 Normal staining patterns on Dsg 1 and 3 substrates. (a) Negative reaction on Dsg1. (b) Positive reaction for Dsg1. (c) Negative reaction for Dsg3. (d) Positive reaction for Dsg3.

inter-rater agreement between the 10 raters varied between fair and moderate for all substrates. Those that demonstrated fair concordance included, MO ($k = 0.257$, $P < 0.0001$), MO immunofluorescence pattern ($k = 0.357$, $P < 0.0001$), Dsg1 ($k = 0.39$, $P < 0.0001$) and BP230 ($k = 0.281$, $P < 0.0001$). Moderate agreement was demonstrated for SSS ($k = 0.416$, $P < 0.0001$), SSS immunofluorescence location ($k = 0.505$, $P < 0.0001$), BP180 ($k = 0.559$, $P < 0.0001$); and Dsg3 ($k = 0.437$, $P < 0.0001$).

Six raters responded to the postassessment questionnaire. The images took on average 7.25 h to rate, 13.9 s per window. Many of the raters found that there was weak desmosomal staining on the MO which made it difficult to differentiate from background staining. Fifty percent of the participants responded that the fluorescence on the Dsg1 and Dsg3 substrates was superficial or displayed peripheral staining which they deemed indeterminate.

Where the diagnosis inferred from the BIOCHIP differed from that of the conventional multistep diagnostic procedure, the BIOCHIP slides were reviewed. These windows differed to the typical patterns observed on the positive and negative controls provided by EUROIMMUN. Figure 4 depicts some of these unconventional staining patterns observed on MO, SSS, BP180, BP230 and Dsg1 substrates. It was common in these slides for the raters to have provided incongruent evaluations.

Discussion

Validity and reliability are the two most important factors to consider when judging the value of a new diagnostic test. Validity measures the extent to which the test measures what it is supposed to measure. Reliability refers to whether the test produces consistent results when repeated measurements are made. The BIOCHIP has demonstrated high validity in the diagnosis of BP and pemphigus diseases; however, it falls short of producing reliable results between users.

On a standard BIOCHIP (Dermatology Mosaic 7; EUROIMMUN) slide, there are 10 incubation fields each containing sections of six different substrates including (i) frozen tissue section of MO; (ii) 1 mol/L NaCl-split skin/SSS; (iii) human embryonic kidney (HEK293) cells transfected with Dsg1 protein ectodomain, (iv) HEK293 cells transfected with Dsg3 protein ectodomain; (v) microdrops of BP180 free antigen; and (vi) HEK293 cells transfected with C-terminal globular domain of the BP230 domain. These substrates are coated on thin glass slides that have been mechanically cut into millimetre-sized fragments and glued onto microscope slides. After incubation, these are viewed under a fluorescent microscope, and the windows are interpreted in a similar manner to conventional IIF.

Several studies have demonstrated its relatively high diagnostic accuracy.[18–23] However, due to the rater-dependent nature of immunofluorescent slides, there is a potential for an interpretation bias and poor reliability. Furthermore, despite the use of a control as a comparison, the intensity and distribution of the fluorescence were variable and at times ambiguous and difficult to interpret.

The assessment of IRR provides a way of quantifying the degree of agreement between raters who make independent ratings about the features of the BIOCHIP. IRR analysis aims to determine how much of the variance in the observed result is due to variance in the true result after the variance due to measurement error between raters has been removed.[24]

The study was a fully crossed design where all the subjects were rated by all the raters – this allowed for systematic bias between raters to be assessed and controlled for in the IRR

© 2019 European Academy of Dermatology and Venereology

Table 1 Summary of Fleiss Kappa statistic

	Fleiss Kappa	CI upper 95%	CI lower 95%	Strength of agreement
Oesophagus	0.257*	0.243	0.272	Fair
Oesophagus pattern	0.357*	0.343	0.370	Fair
SSS	0.416*	0.402	0.431	Moderate
Location	0.505*	0.494	0.515	Moderate
Dsg1	0.39*	0.376	0.404	Fair
Dsg3	0.437*	0.423	0.452	Moderate
BP180	0.559*	0.544	0.574	Moderate
BP230	0.281*	0.267	0.295	Fair

The inter-rater agreement between the 10 raters varied between fair and moderate for all substrates.
CI, confidence interval; SSS, salt-split skin.
*$P < 0.0001$.

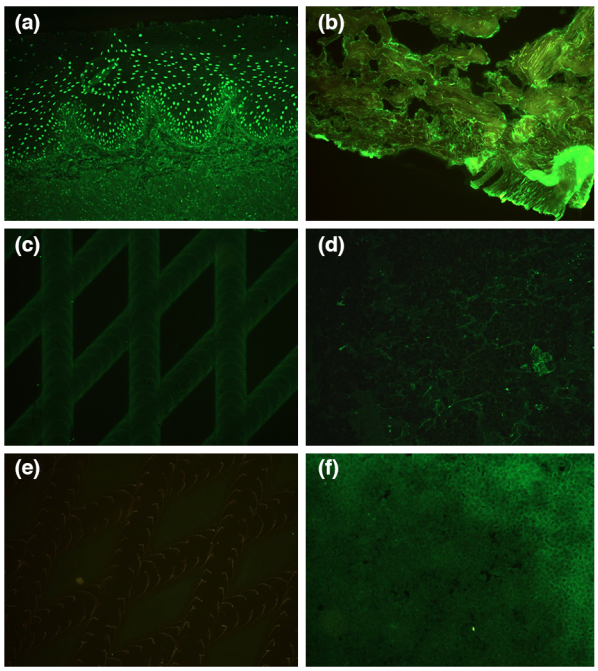

Figure 4 Examples of ambiguous staining. (a) Nuclear staining on epithelium on monkey oesophagus (MO) of a control patient. (b) Ambiguous staining on salt-split skin. (c) Inverse staining of BP180. (d) Ambiguous staining on Dsg 1. (e) Faint staining on BP180. (f) Uneven staining of cells on BP230.

estimate.[25] This is instead of only a subset of subjects being reviewed by multiple raters and generalization applied to the full sample.

An IRR analysis was performed to assess the degree that raters consistently interpreted each individual substrate of the BIOCHIP. Fleiss Kappa (κ) was computed for each coder pair then averaged to provide a single index of IRR.[26] The resulting kappa indicated fair-to-moderate agreement, ($\kappa = 0.257$–0.559).[27] There are no other published IRR estimates to compare these findings in the literature. The IRR analysis suggested that raters had better agreement in SSS ($k = 0.416$, $P < 0.0001$), SSS immunofluorescence location ($k = 0.505$, $P < 0.0001$) and BP180 ($k = 0.559$, $P < 0.0001$). Only fair agreement is observed in interpretation of the MO ($k = 0.257$, $P < 0.0001$), staining pattern on MO ($k = 0.357$, $P < 0.0001$), Dsg1 ($k = 0.39$, $P < 0.0001$) and BP230 ($k = 0.281$, $P < 0.0001$).

The low IRR indicated that the observed ratings contained a large amount of measurement error. The authors postulate that the low IRR could be due to variability in experience amongst raters and occasional ambiguous staining patterns.

The raters that were selected are known within the international community to be users of the BIOCHIP. To ensure a priori level of knowledge, they were required to correctly interpret control images on a qualifying test prior to having authorization to interpret the real subjects in the study. Despite this, some raters were more experienced in the use of the BIOCHIP compared with others. Those with more experience with the BIOCHIP demonstrated more congruent results compared to those with less experience. This highlights the importance of ensuring appropriate training and that users should be aware of the learning curve present in the use of this investigation.

Ambiguous staining patterns accounted for many of the 'indeterminate' results. Of the MO substrates deemed indeterminate, raters reported difficulty interpreting the staining pattern. Several reasons included 'weak desmosomal staining', 'weak linear staining', 'focal linear fluorescence' and 'missing epidermis'. Additionally, two participants reported the presence of nuclear staining in the epithelium (Fig. 4a) on several MO substrates. This indicated the presence of antinuclear antibodies (ANA), and these findings further correlated with the presence of antibodies against the dermal side of the basement membrane on their respective SSS substrates. These results are suggestive of bullous systemic lupus erythematosus (BSLE) or EBA as recently published studies have found positive ANA in 25% of patients.[28] EUROIMMUN have developed alternative BIOCHIP slides with

© 2019 European Academy of Dermatology and Venereology

Collagen VII to screen for BSLE or EBA that has demonstrated promising results.[29]

The majority commented that the electronic format presenting the BIOCHIP photos in a more convenient layout that allowed them to see their evaluations of all six windows simultaneously. They found this a more efficient and convenient format than the conventional use of a microscope. Two participants stated that the electronic format interfered with their evaluations due to not being able to magnify the photos. This was an inevitable disadvantage of presenting the images electronically.

Conclusion

The design of the BIOCHIP enables differential diagnosis amongst the different AIBD subtypes from a single serum sample. This has great potential to be a sensitive and specific test to aid in the timely and accurate diagnosis of AIBD. An interrater analysis was performed to assess the degree that raters consistently interpreted each individual substrate. The level of agreement was relatively low, and the authors found the most common causes to be variable levels of training, indicating the presence of a learning curve in the interpretation of the results, and ambiguous staining patterns leading to incongruent results.

References

1 Ghodsi SZ, Chams-Davatchi C, Daneshpazhooh M *et al*.. Quality of life and psychological status of patients with pemphigus vulgaris using Dermatology Life Quality Index and General Health Questionnaires. *J Dermatol* 2012; **39**: 141–144.

2 Zhao CY, Murrell DF. Autoimmune blistering diseases in females: a review. *Int J Womens Dermatol* 2015; **1**: 4–12.

3 Yang A, Xuan R, Wei M *et al*. Validation of the BIOCHIP test for the diagnosis of bullous pemphigoid, pemphigus vulgaris and pemphigus foliaceous. *J Eur Acad Dermatol Venereol* 2019. In press. https://doi.org/10.1111/jdv.15770

4 Stanley JR, Tanaka T, Mueller S *et al*. Isolation of complementary DNA for bullous pemphigoid antigen by use of patients' autoantibodies. *J Clin Invest* 1988; **82**: 1864–1870.

5 Giudice GJ, Emery DJ, Diaz LA. Cloning and primary structural analysis of the bullous pemphigoid autoantigen BP180. *J Invest Dermatol* 1992; **99**: 243–250.

6 Kobayashi M, Amagai M, Kuroda-Kinoshita K *et al*. BP180 ELISA using bacterial recombinant NC16a protein as a diagnostic and monitoring tool for bullous pemphigoid. *J Dermatol Sci* 2002; **30**: 224–232.

7 Di Zenzo G, Grosso F, Terracina M *et al*. Characterization of the anti-BP180 autoantibody reactivity profile and epitope mapping in bullous pemphigoid patients. *J Invest Dermatol* 2004; **122**: 103–110.

8 Sakuma-Oyama Y, Powell AM, Oyama N *et al*. Evaluation of a BP180-NC16a enzyme-linked immunosorbent assay in the initial diagnosis of bullous pemphigoid. *Br J Dermatol* 2004; **151**: 126–131.

9 Bystryn JC, Rudolph JL. Pemphigus. *Lancet* 2005; **366**: 61–73.

10 Amagai M, Komai A, Hashimoto T *et al*. Usefulness of enzyme-linked immunosorbent assay using recombinant desmogleins 1 and 3 for serodiagnosis of pemphigus. *Br J Dermatol* 1999; **140**: 351–357.

11 Amagai M, Tsunoda K, Zillikens D *et al*. The clinical phenotype of pemphigus is defined by the anti-desmoglein autoantibody profile. *J Am Acad Dermatol* 1999; **40**(2 Pt 1): 167–170.

12 Kershenovich R, Hodak E, Mimouni D. Diagnosis and classification of pemphigus and bullous pemphigoid. *Autoimmun Rev* 2014; **13**: 477–481.

13 Mahoney MG, Wang ZH, Stanley JR. Pemphigus vulgaris and pemphigus foliaceus antibodies are pathogenic in plasminogen activator knockout mice. *J Invest Dermatol* 1999; **113**: 22–25.

14 James KA, Culton DA, Diaz LA. Diagnosis and clinical features of pemphigus foliaceus. *Dermatol Clin* 2011; **29**: 405–412, viii.

15 Hashimoto T, Amagai M, Garrod DR *et al*. Immunofluorescence and immunoblot studies on the reactivity of pemphigus vulgaris and pemphigus foliaceus sera with desmoglein 3 and desmoglein 1. *Epithelial Cell Biol* 1995; **4**: 63–69.

16 Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971; **76**: 378.

17 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **159–74**.

18 Russo I, Saponeri A, Peserico A *et al*. The use of biochip immunofluorescence microscopy for the diagnosis of Pemphigus vulgaris. *Acta Histochem* 2014; **116**: 713–716.

19 Tampoia M, Zucano A, Villalta D *et al*. Anti-skin specific autoantibodies detected by a new immunofluorescence multiplex biochip method in patients with autoimmune bullous diseases. *Dermatology* 2012; **225**: 37–44.

20 Chiang YZ, Zhao C, Melbourne W *et al*. Biochip immunofluorescence microscopy as a new diagnostic tool for autoimmune blistering skin diseases. *Br J Dermatol* 2015; **173**: 7–8.

21 van Beek N, Rentzsch K, Probst C *et al*. Serological diagnosis of autoimmune bullous skin diseases: prospective comparison of the BIOCHIP mosaic-based indirect immunofluorescence technique with the conventional multi-step single test strategy. *Orphanet J Rare Dis* 2012; **7**: 49.

22 Van Beek N, Rentzsch K, Probst C *et al*. Prospective comparison of a specially composed biochip mosaic with the conventional multistep procedure in the serological diagnosis of autoimmune bullous diseases. *Exp Dermatol* 2012; **21**: e34.

23 Ozkesici B, Mutlu D, Donmez L *et al*. The Value of the BIOCHIP mosaic-based indirect immunofluorescence technique in the diagnosis of pemphigus and bullous pemphigoid In turkish patients. *Acta Dermatovenerol Croat* 2017; **25**: 202–209.

24 Novick MR. The axioms and principal results of classical test theory. *J Math Psychol* 1966; **3**: 1–18.

25 Putka DJ, Le H, McCloy RA *et al*. Ill-structured measurement designs in organizational research: implications for estimating interrater reliability. *J Appl Psychol* 2008; **93**: 959–981.

26 Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. *Psychol Bull* 1971; **76**: 365–377.

27 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–174.

28 Delgado L, Aoki V, Santi C *et al*. Clinical and immunopathological evaluation of epidermolysis bullosa acquisita. *Clin Exp Dermatol* 2011; **36**: 12–18.

29 Marzano AV, Cozzani E, Biasin M *et al*. The use of Biochip immunofluorescence microscopy for the serological diagnosis of epidermolysis bullosa acquisita. *Arch Dermatol Res* 2016; **308**: 273–276.

© 2019 European Academy of Dermatology and Venereology

Appendix 1

BIOCHIP Questionnaire

Regarding the BIOCHIP itself

1. Approximately how long did it take to finish reading the BIOCHIP slides?

2. What were the strengths of reading the BIOCHIP compared to conventional indirect immunofluorescence slides?

3. What were the challenges you encountered reading the BIOCHIP?

4. If there were windows that you deemed as indeterminate, what were the reasons?
   - Oesophagus:
   - Salt Split Skin:
   - Desmoglein 1:
   - Desmoglein 3:
   - BP180:
   - BP230:

5. Would you recommend the BIOCHIP for clinical use? If so, why? If not, why?

6. Any other comments

Regarding the electronic format of how the BIOCHIP slides were presented

1. What are some obvious differences (strengths or difficulties) you encountered reading the BIOCHIP electronically versus in person?

2. What were some challenges you encountered with the electronic format?

3. Did these challenges affect your evaluation of the BIOCHIP results?

4. Do you have any suggestions for improvement?

5. Any other comments

© 2019 European Academy of Dermatology and Venereology