Clinical Immunology (2013) 148, 99–109

available at www.sciencedirect.com

Clinical Immunology

www.elsevier.com/locate/yclim

Profiling celiac disease antibody repertoire☆

Sara D'Angelo [a,g], Flavio Mignone [b], Cecilia Deantonio [a], Roberto Di Niro [c], Roberta Bordoni [d], Roberto Marzari [e], Gianluca De Bellis [d], Tarcisio Not [f], Fortunato Ferrara [g], Andrew Bradbury [g], Claudio Santoro [a,*], Daniele Sblattero [a,*]

[a] *Department of Health Sciences and Interdisciplinary Research Center on Autoimmune Diseases (IRCAD), University of Eastern Piedmont "Amedeo Avogadro", Novara, Italy*
[b] *Dipartimento di Scienze e Innovazione Tecnologica, University of Eastern Piedmont "Amedeo Avogadro", Alessandria, Italy*
[c] *Department of Laboratory Medicine, Yale University School of Medicine, New Haven, CT 06510, USA*
[d] *Institute of Biomedical Technologies, National Research Council (ITB CNR), Milan, Italy*
[e] *Department of Life Sciences, University of Trieste, Trieste, Italy*
[f] *Department of Medical Sciences University of Trieste, Institute of Child Health IRCCS "Burlo Garofolo", Trieste, Italy*
[g] *Los Alamos National Laboratory, Los Alamos, NM, USA*

Received 22 January 2013; accepted with revision 10 April 2013

KEYWORDS
Celiac disease;
Autoantibody;
Autoantigen;
Protein microarray;
ORF-display libraries;
Next generation sequencing

Abstract The aim of this study was to dissect the autoantibody response in celiac disease (CD) that remains largely unknown, with the goal of identifying the disease-specific autoantigenic protein pattern or the so called epitome. Sera from CD patients were used to select immunoreactive antigens from a cDNA phage-display library. Candidate genes were identified, the corresponding proteins produced and their immunoreactivity validated with sera from CD patients and controls. Thirteen CD-specific antigens were identified and further validated by protein microarray. The specificity for 6 of these antigens was confirmed by ELISA. Furthermore we showed that this antibody response was not abolished on a gluten free diet and was not shared with other autoimmune diseases. These antigens appear to be CD specific and independent of gluten induction. The utility of this panel extends beyond its diagnostic value and it may drive the attention to new targets for unbiased screens in autoimmunity research.

1. Introduction

The circulating antibody repertoire represents an important source of diagnostic information, serving as a biomarker of the immune state of the body [1]. The identification of disease signatures for known and unknown etiology can be carried out in an unbiased way by performing serum antibody-based

☆ *Grant support*: This work was supported by: Fondazione Cariplo Bando Ricerca Biomedica 2009; EC Marie Curie Research Training Network [contract no. MRTN-CT-20010-289964]; Compagnia San Paolo; MIUR FIRB NG-Lab RBLA03ER38; IRCCS Burlo Garofolo RF 35/07; Regione Piemonte Ricerca Sanitaria and Piattaforma Immonc.
* Corresponding authors at: Università del Piemonte Orientale, Dip. Scienze della Salute, Via Solaroli 17, 28100 Novara, Italy.
E-mail addresses: csantoro@med.unipmn.it (C. Santoro), sblatter@med.unipmn.it (D. Sblattero).

1521-6616/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.clim.2013.04.009

profiling. The application of this unbiased discovery-driven approach has led to the identification of a large number of novel auto-immune antigens that have been identified using the sera of patients with cancer [2,3] and auto-immune diseases [4–6]. The complex network of the antibody–antigen immune response has recently been investigated by several innovative approaches including bacterial [7], lambda [8,9], T7 [10], and filamentous phage display systems [11]. In particular, phage display provided simple and fast procedures for the identification of novel antigens, allowing the screening of libraries created from appropriate sources directly with immune sera. Furthermore, the implementation of next generation sequencing technologies (NGS) at the screening level resulted in a more comprehensive and faster analysis of the interactions under study [12,13]. By combining the antibody repertoire with cDNA display technologies, thousands of antigen–antibody interactions are simultaneously screened and give, altogether, an extensive analysis of the "immunome" [14]. We developed a robust approach to identify novel autoantigens and to describe the profile of the corresponding autoantibody reactivity. Among autoimmune pathologies, celiac disease (CD) is of particular interest due to its high prevalence (1%) in the general population [15]. CD is an enteropathy developing as a consequence of dietary exposure to wheat gluten and related proteins in barley and rye [16], characterized by the production of antibodies to the autoantigen tissue transglutaminase (TG2) [17]. This depends upon dietary exposure to gluten [18,19] and is found only in genetically predisposed individuals. Although TG2 is the main autoantigen of the disease, with increasing diagnostic importance [20], several other autoantigen targets may exist. The disease association with other autoimmune diseases is well known [21], as well as the fact that these autoimmune disorders developed in unrecognized and/or untreated celiac subjects due to prolonged exposure to gluten [22]. The need for an early and correct diagnosis to prevent morbidity and mortality associated with untreated CD is a growing need. Hence, the identification of other Ab specificities is gaining importance. To date, reactivity to only a limited number of CD specific autoantigens has been clearly defined like actin [23], calreticulin [24], enolase alpha [25], and few others (reviewed in [26,27]). Furthermore only few systematic efforts have been attempted to describe the whole autoantigen repertoire (*i.e.* celiac disease epitome) [7,28]. Here, we used CD sera to select autoantigenic proteins from a large, cDNA library [12] applying an innovative discovery platform where we integrate genomic (NGS) and proteomic (protein microarray) technologies to perform high-throughput large-scale screening for potential autoantigens. We identified a novel panel of 13 CD-associated autoantigens, with a high antibody specificity in patients, that may represent a useful descriptive tool for the understanding of the immunological and pathogenic processes involved in the onset of CD and autoimmunity in general.

2. Materials and methods

2.1. Sera samples

Sera were obtained from IRCCS Burlo Garofolo, Italy; informed consent was obtained from all donors. Celiac patients underwent a complete small intestine mucosal biopsy and test for antitissue transglutaminase antibodies. Healthy controls tested negative for anti-TG2 and were asymptomatic. Sample details are summarized in Table 1. Briefly, the sera used for antigen selections were obtained from untreated CD patients, all confirmed by histological analysis of the intestinal mucosa biopsy. The microarray screening was performed with 46 sera from CD patients and 45 sera from healthy donors. The ELISA screening was performed with 3 different sets of patients' sera: the first comprised 46 CD patients, the second (termed pre-GFD–gluten free diet) consisted of 31 CD patients, and the third (termed post-GFD) consisted of the same set of 31 pre-GFD patients, whose sera were collected after 3 to 9 months of GFD; 44 healthy sera were used as controls. The ELISA was also performed with sera derived from patients affected by other autoimmune diseases: 28 type 1 diabetes (T1D) and 30 multiple sclerosis (MS) patients. All sera were tested for both IgG and IgA TG2 reactivity (Eurospital Eu-tTg kit).

Table 1 Sera used for autoantigen discovery platform. Set of sera for phage selection: pools H and J were used for library selections, pool C for the negative selection. Discovery set of sera for microarray: CD patient and healthy control sera used to screen antigen microarrays. Validation set of sera for ELISA: CD and healthy control sera for antigen validation after microarray screening (F, female; M, male).

	No.	Sex	Average age
Selection set for phage selection			
CD pool H	5	M 3 F 2	11.8 (8–21)
CD pool J	5	M 2 F 3	13.8 (5–28)
Healthy pool C	5	M 3 F 2	14.3 (5–28)
Discovery set for microarray test			
CD	46	M 17 F 29	11.2 (1–37)
Healthy	45	M 25 F 20	23.7 (17–50)
Validation set for ELISA test			
CD	46	M 10 F 36	7.5 (2–27)
Pre-GFD	31	M 12 F 19	5.8 (1–14)
Post-GFD	31	M 12 F 19	6.5 (2–15)
Healthy	44	M 20 F 24	9.2 (7–10)
Autoimmune set for ELISA test			
T1D	28	M 18 F 10	8.6 (3–14)
MS	30	M 16 F 14	43.4 (32–62)
Healthy	30	M 18 F 12	13.3 (5–28)

2.2. Construction and selection of cDNA phage display library

The human cDNA ORF library was prepared as described in Di Niro et al. [12]. The biopanning procedure was performed with sera from pools H and J (CD sera), while pool C (the healthy control sera) was used for the negative (subtractive) selection. Phage selections were performed as described in Di Niro et al. [29].

2.3. 454 sequencing and reads ranking of selected library fragments

cDNA fragments from selected phagemids were used to prepare 454 libraries according to Di Niro et al. [12]. The reads were first mapped onto the human genome reference sequence using GMAP software. The identified genes were then ranked according to the total number of reads (TNR) obtained from fragments recovered after the first, second and third rounds of selection on both pools H and J (TNR1H–2H–3H and TNR1J–2J–3J). After ranking genes present in at least 2 of the 3 selection rounds for each pool were identified and within these genes with a number of reads greater than a cut off, empirically set to 0.25% of corresponding TNR for each pool and round (0.25%TNR), were finally selected. The genes satisfying these criteria were considered putative antigens and further analyzed.

2.4. Putative antigen cloning and protein expression

cDNA fragments from the third round of selection were subcloned from the phagemid DNA into a compatible pGEX 4T-1 (GE-Healthcare) expression vector to yield GST-fusion products. The so obtained plasmids were used as a template for the following inverse PCR-based cloning strategy: two back-to-back outward phosphorylated primers corresponding to the center of each identified overlapping sequence contig, were used to carry out inverse PCR and ligation [30]. Positive colonies were grown and protein purifications were performed as reported in [29].

2.5. Antigenic protein microarray screening and identification

Quadruplicates of 42 different proteins, including negative and positive controls, were spotted onto nitrocellulose slides. Slides were saturated with blocking solution (3% milk in PBS 0.1% Tween 20–PBST) for 1 h at RT. Sera were diluted 1:200 in binding buffer (2% nonfat dry milk in PBST) and incubated for 90 min. Slides were washed twice for 15 min with PBST and once for 15 min with PBS. Secondary antibody, Cy5 conjugated anti-human IgG, was used at 1:100 dilution for 1 h at RT in binding buffer. Final washing was performed twice with PBST for 15 min, once with PBS for 15 min and once with water for 5 min. Slides were dried and scanned. Arrays were normalized with a 2-step protocol. Anti-GST background response was evaluated using the signal generated for each serum against a reference GST protein and a set of serial dilutions of purified IgG printed on each array was used to generate a calibration curve of arbitrary IgG units. A cut-off value for positivity for each antigen was defined as the mean value obtained with control sera plus 2 standard deviations (SD). Positive proteins were identified by comparing the frequency of reactive sera using Fisher's exact test: the proteins with a significant difference ($P < 0.05$) in reactivity between patients and controls were identified as putative candidates for further analysis.

2.6. ELISA on immunoreactive antigens

ELISA was performed as follows: recombinant GST-proteins were diluted in PBS to 10 μg/ml and 100 μl were coated in ELISA wells (Nunc), overnight (O/N) at 4 °C. Wells were washed with PBS and 200 μl of blocking solution (PBS plus 0.2% Tween20), were added to each well, for 1 h at RT. All sera were used as primary antibodies, diluted 1:100 in blocking solution and incubated for 2 h at 30 °C. Extensive washes were performed with PBST and PBS. Secondary antibody was a goat anti-human-IgG HRP conjugated (Jackson) diluted 1:3000 in blocking solution, for 1 h at 30 °C. After extensive washing, immunocomplexes were revealed with TMB and the plate read at 450 nm. A cut-off value for positivity for each antigen was defined as the mean OD_{450} value obtained with control sera plus 2 standard deviations (SD).

2.7. Statistical analysis

The ELISA data assessing the presence of antibodies directed towards the identified peptides were compared between patients and controls using Fisher's exact test for 2×2 contingency tables and T–Student test for the mean comparison.

The thirteen specific antigens were tested for their association with CD by means of univariate binary logistic regression analyses and, if statistically significant after controlling for age and gender, included in a multivariable model. For each antigen that resulted independently associated with CD from regression analyses a ROC curve was plotted and a cut-off value for classification was identified according to maximum sensitivity/specificity criteria. According to these thresholds, diagnostic performances were then explored for all identified antigens and for their possible combinations. In all of the analyses statistical significance was set for P-values < 0.05.

3. Results

3.1. Project strategy summary

To date, there is no systematic analysis of the autoantigens recognized by CD-specific autoantibodies. To address this issue we designed a comprehensive discovery and validation strategy, outlined in Fig. 1:

1–2) cDNA library construction from human tissues. The cDNA is filtered to retain only clones encoding open reading frame (ORF) fragments which are then displayed on filamentous phage. 3–7) Selection of immunoreactive antigens. Phages displaying proteins are first challenged with control sera to reduce background and then with sera from CD patients through cycles of selection and

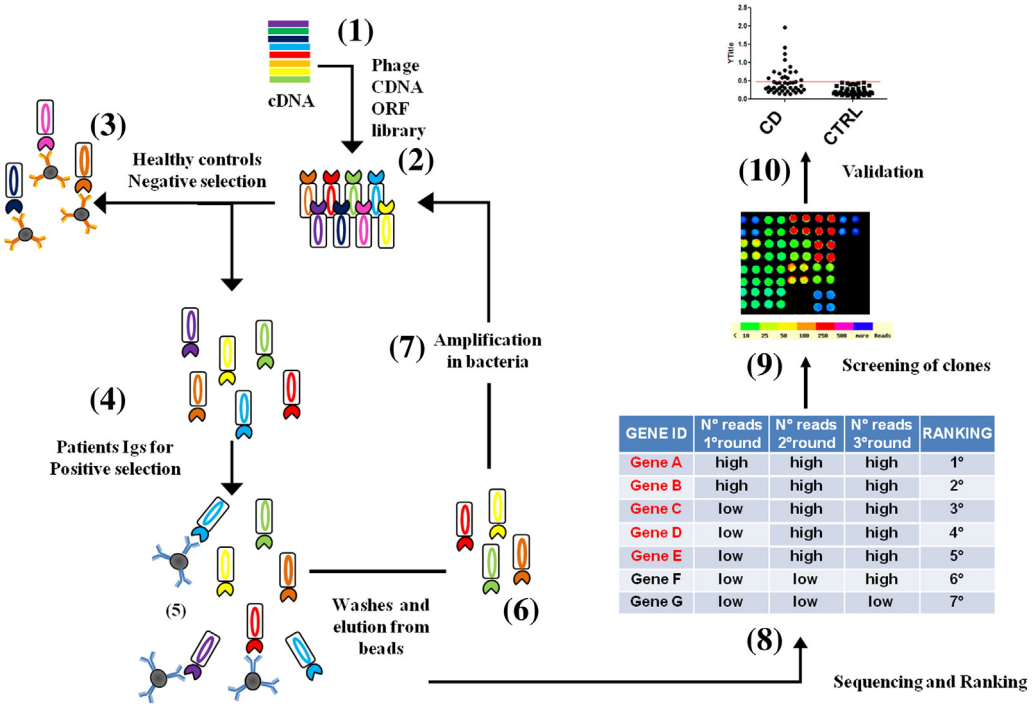

Figure 1 Project summary. cDNA encoding ORFs (1) derived from the tissue of interest were cloned into a phage display system (2) to display the corresponding protein fragments. A negative selection (3) with healthy control sera was performed before each selection round to remove polyreactive clones. Unbound phages (4) were used for positive selection on patient serum antibodies immobilized on magnetic beads (5), extensive washes were followed by phage amplification in bacteria (6–7). The selection outputs were deep-sequenced (8) allowing the identification of enriched gene sequences. After sequence analysis, putative antigens were produced and screened by protein microarray (9). The most immunoreactive clones were validated with an independent set of sera and assessed for their specificity (10).

amplification. 8) Identification of putative antigens. After each selection cycle, the gene repertoire present in the library is assessed by massive sequencing and single genes are ranked according to the number of reads mapping to each. The genes with the highest frequencies are listed as putative antigens. 9–10) Validation of antigenic fragments. Candidate ORFs are cloned and expressed as GST-fusion products, and finally validated by independent sets of CD and control sera by both protein microarray and ELISA assays.

3.2. Selections of autoantigenic proteins

The cDNA library used in this study was constructed and extensively characterized as described in Di Niro et al. [12]. Sera from 10 CD patients were used as a "selection set". All patients were diagnosed by the presence of both the intestinal lesion and serum antibody positivity to TG2. The sera were randomly grouped into two different pools (5 sera each) named H and J (Table 1). Each pool was independently used for three consecutive rounds of selection, in order to enrich the library for immunoreactive clones, with increasing washing and binding stringency. Before each round, a "subtractive step" (point 3, Fig. 1) was introduced by incubating phages with pool C, composed of sera from healthy blood donors (testing negative for the CD serological markers). This step was carried out in order to eliminate polyreactive clones.

3.3. Antigenic clone identification by 454 pyrosequencing and ranking of selected inserts

After each round of selection, cDNA fragments were recovered and used for 454 pyrosequencing. The rationale for this strategy was based on our previous finding [12] that selection cycles can progressively and specifically enrich for specific clones that can be rapidly identified and ranked by sequencing. Genes with greater numbers of reads are those that have been specifically enriched and therefore correspond to possible antibody targets. A total of 275,569 reads were obtained from the 6 sequencing runs (Table 1S), 79% of the reads were positively mapped on the human genome sequence and the corresponding mapped genes were identified. For each selection cycle the genes identified were then ranked according to their coverage (*i.e.* total number of reads mapping on the gene). Finally, ranking lists obtained from the different rounds of selection were compared and candidate genes were selected. After matching and comparing the rankings generated from each selection cycle, a total of 60 unique genes were identified (Table 2S).

3.4. Biomarker panel validation by protein microarray

To test the immunoreactivity of the 60 predicted autoantigens identified *via* 454 ranking, we cloned and expressed the corresponding fragments as C-terminal GST

fusion proteins. Using an inverse-PCR based strategy [12], we recovered clones corresponding to 42 of the 60 identified genes. All 42 clones were sequenced and analyzed: 37 were defined as "genic" (translated in the correct genic frame); and 5 clones were non-genic ORFs, potentially selected as mimotopes similar to conformational epitopes of real antigens recognized by the serum antibodies. GST fusion proteins were successfully produced and purified by affinity chromatography. We performed the initial validation of the identified antigenic proteins by protein arrays. Purified proteins were printed, along with controls and calibrator IgGs, on nitrocellulose slides. Each array was then incubated individually with sera from the "discovery set", composed of 46 sera from CD patients and with 45 sera from healthy controls (see Table 1) and reactivity was revealed with anti-human IgG antibody. Fig. 2 shows representative high resolution scanning images of microarrays: panel A demonstrates that proteins were arrayed at similar levels, panels B and C show the differences in immunoreactivity between two samples from either CD patients (B) or controls (C). Data obtained from the arrays were normalized for signal intensity by using the IgG calibration curve and a cut-off value for positivity was calculated for all the proteins tested using the signals obtained with the healthy donors' sera. The positivity of sera on the selected antigens is shown in Fig. 2D, where each dot represents one of the 42 antigen fragments, plotted against the number of sera recognizing the specific Ag (Y axis) in both CD and healthy control sets. When analyzing the reactivity with the healthy controls, two antigens were recognized by 5 sera, one was recognized by 4 sera and all the remaining proteins showed limited or negative reactivity. In contrast, the majority of proteins were recognized by the CD test set, with 20 antigens recognized by at least 7 CD sera samples. According to the specific reactivity measured for each protein, 13 antigens (Table 2) were identified by Fisher's exact test as those with the most significant difference in immunoreactivity between patients and controls, and were therefore identified as the most informative (Table 3S). Among these, 9 were derived from the pool H screening (HERPUD1, MCM5, MML2, PRPF31, SPTAN1, AASS, CDC37, MATR3, PRKCSH) and 4 from pool J (RAI1, NGFRAP1, NASP, IPPK). Within the identified

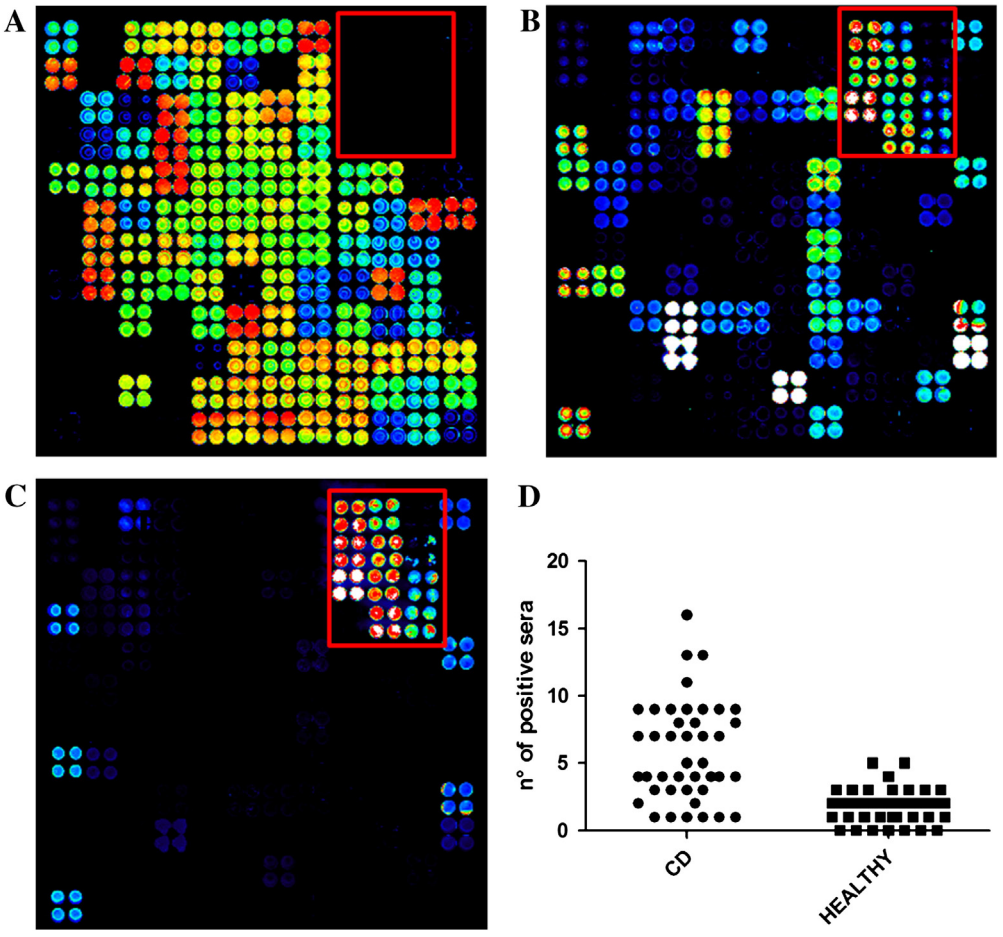

Figure 2 Protein microarray screening. A) Control of protein array quality by anti-GST mAb. B–C). Representative microarrays incubated with CD patient serum (B) and healthy donor serum (C). On the top right corner in each array, the IgG calibrator (red box). Each protein is spotted in 4 replicates. D) Distribution of the reactivity of sera on CD focused protein microarrays. Each spot on the chart represents one of the antigens printed on the microarray. For each of the antigens identified by massive sequencing approach, the number of positive sera in each subset (46 CD and 45 controls) is shown.

Table 2 ELISA immunoreactivity of CD patient and control sera to top-performing antigens.
The 13 best performing antigens identified by microarray were validated with 2 different sets of sera: 46 sera from celiac disease patients, 44 from healthy donors. Name, length and number of positive sera is reported for each antigen (sensitivity and specificity are showing in parentheses).

HUGO	Definition	AA length	CD (n. 46)	HEALTHY (n. 44)	P-value
MLL2	Myeloid/lymphoid or mixed-lineage leukemia 2	29	21 (45.7)	0 (100)	<.0001
NGFRAP1	Nerve growth factor receptor (TNFRSF16) associated protein 1	68	20 (43.5)	3 (93.2)	<.0001
AASS	Aminoadipate-semialdehyde synthase	65	15 (32.6)	2 (95.5)	.0008
RAI1	Retinoic acid induced 1	126	9 (19.6)	1 (97.7)	.0153
MCM5	Minichromosome maintenance complex component 5	50	9 (19.6)	1 (97.7)	.0153
NASP	Nuclear autoantigenic sperm protein (histone-binding)	88	11 (23.9)	3 (93.2)	.0397
CDC37	Cell division cycle 37 homolog (*S. cerevisiae*)	79	11 (23.9)	4 (90.9)	.0888
PRKCSH	Protein kinase C substrate 80K-H	97	8 (17.4)	2 (95.5)	.0905
HERPUD 1	Homocysteine-inducible, endoplasmic reticulum stress-inducible, ubiquitin-like domain member	86	7 (15.2)	2 (95.5)	.1582
SPTAN1	Spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)	55	7 (15.2)	2 (95.5)	.1582
PRPF31	PRP31 pre-mRNA processing factor 31 homolog (*S. cerevisiae*)	50	6 (13.0)	2 (95.5)	.2673
IPPK	Inositol 1,3,4,5,6-pentakisphosphate 2-kinase	27	4 (4.8)	1 (97.7)	.3614
MATR3	Matrin 3	67	3 (6.5)	4 (90.9)	1

fragments, 11 were genic ORFs, while two (IPPK and AASS) encoded mimotopes (nongenic ORFs).

3.5. ELISA validation of selected antigens

We validate the 13 most immunoreactive antigens with a new panel of sera, the "validation set", from CD patients (46 samples) and healthy controls (44 samples) (see Table 1), and by ELISA as a different analytical procedure. Of the 13 antigens tested, 6 showed a statistically significant difference in immunoreactivity between CD sera and controls as reported in Fig. 3 and summarized in Table 2. As observed by protein arrays, the specificity of these proteins was very high, ranging from 90.9% (NGFRAP1 and NASP) to 100% for MLL2. Sensitivity was more variable, with the highest value (45.7%) obtained for MLL2, followed by NGFRAP1 with 43.5%. Least effective were RAI1 and MCM5, which were recognized by nearly 20% of CD sera (Table 2). Thirteen different CD sera, out of the 46, tested positive for at least 4 of the 6 antigens.

In multivariate analyses we found that 4 antigens (suppl. Table 4A) were independently associated to the outcome. By optimizing the cut off value of the receiver operating characteristic (ROC) curve as a function of varying thresholds we obtained a classifier (CDC37, MLL2 and NGFRAP) that enables the discrimination of CD patients from healthy controls with 80% overall accuracy, with 80.4% sensitivity and 80% specificity (suppl. Table 4B).

3.6. Gluten dependence of antibody response

In order to monitor the gluten dependence of the antibody response, we selected a novel set of sera comprising 31 sera taken at the time of diagnosis (pre-GFD) and after a period of gluten free diet (post-GFD). All pre-GFD sera, but one, contained anti-TG2 IgAs, while 16 sera were positive for IgGs (Fig. 4A). After an average of 6 months of GFD, all post-GFD sera, but one, showed a decline in the antibody titer with 15 being negative for IgAs and 28 for IgGs. These sera were then tested by ELISA against the 6 antigens we identified. Using the pre-GFD set, NASP had the highest sensitivity (48.5%), followed by MLL2 and AASS both with 35.5%. MCM5 was confirmed as the least effective (Fig. 4B). Using post-GFD sera, AASS showed the highest sensitivity with 12 positive sera, followed by NASP with 11 and MLL2 with 10. Interestingly, unlike the response to TG2, there was not a clear decrease in reactivity after the gluten free diet but the response to the different antigens was more complex with sera becoming positive and *vice versa*. For example for RAI1 6 pre-GFD sera became positive in post-GFD and 5 became negative in post-GFD. Overall there was no statistically significant difference between the number of positive sera in pre-GFD and in the post-GFD population.

3.7. Autoimmunity dependence of antibody response

Finally we determined whether the identified antigens were specific for the CD auto-immune response or shared with other autoimmune diseases. We selected a final set of sera, "autoimmunity set", composed of 28 type 1 diabetic (T1D) sera and 30 multiple sclerosis (MS) sera. All sera were tested for the presence of anti-TG2 antibodies and were found to be negative for both IgG and IgA (Fig. 5A). When tested on the 6 antigens, we found that there was no specific reactivity for either set of sera: both T1D and MS sera showed the same low level of immunoreactivity found in the

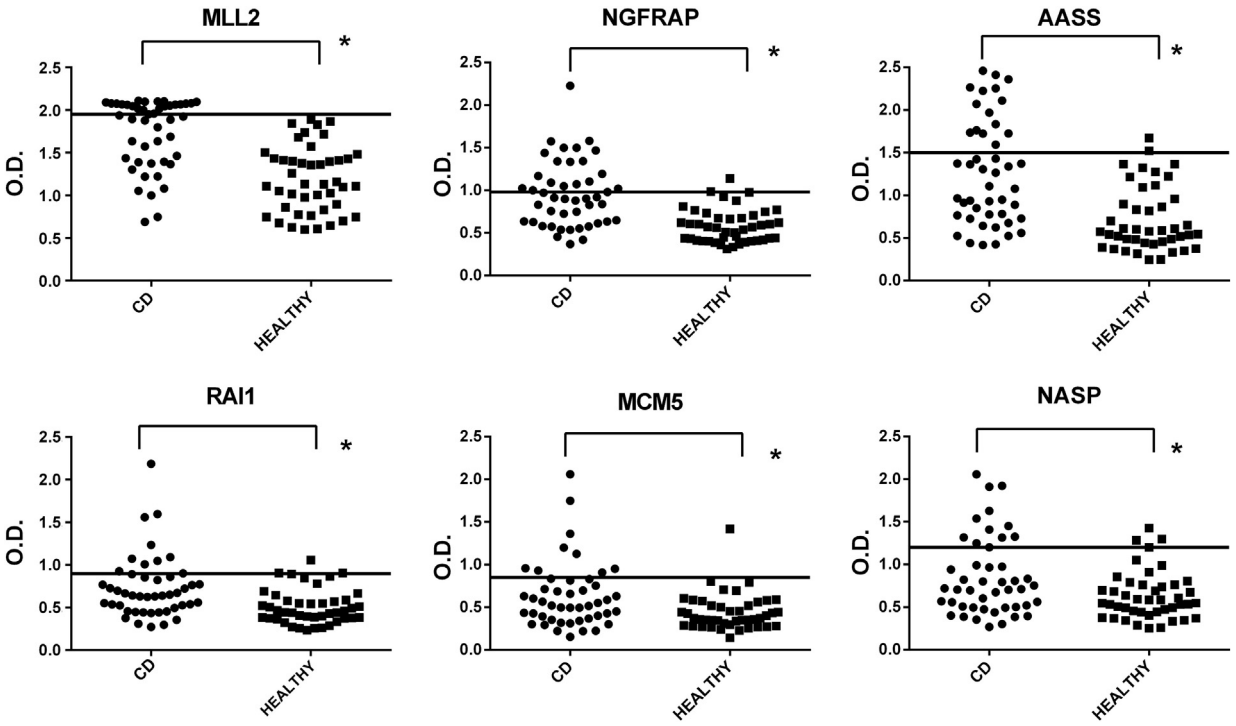

Figure 3 Specificity of autoantibody response. Immunoreactivity of 46 CD patients and 44 control sera towards 6 of the top-performing antigenic fragments identified is reported. 95% confidence cut-off for positivity is shown as a line. *Statistically significant ($P < .05$).

healthy control set of sera (Fig. 5B), therefore confirming the specificity of the antigens for CD.

4. Discussion

Few attempts have been made to obtain a comprehensive analysis of the CD autoimmune response. Zanoni et al. [28] used CD patient sera to screen a peptide library identifying a set of peptides with partial homology to desmoglein-1, rotavirus VP-7 protein, TG2 and TLR4, suggesting a possible involvement of rotavirus as triggering antigen. Spatola et al. [7] identify a further set of six peptides that, although providing a good sensitivity, could not be linked to any specific protein. Despite the wide diversity of peptide libraries, these often yield controversial results: one of the main drawbacks is that antigen mimotopes are predominantly recovered, requiring additional analysis to identify the original antigen that elicited the immune response. For this reason, cDNA libraries appear to be more suitable for antigen identification. For this

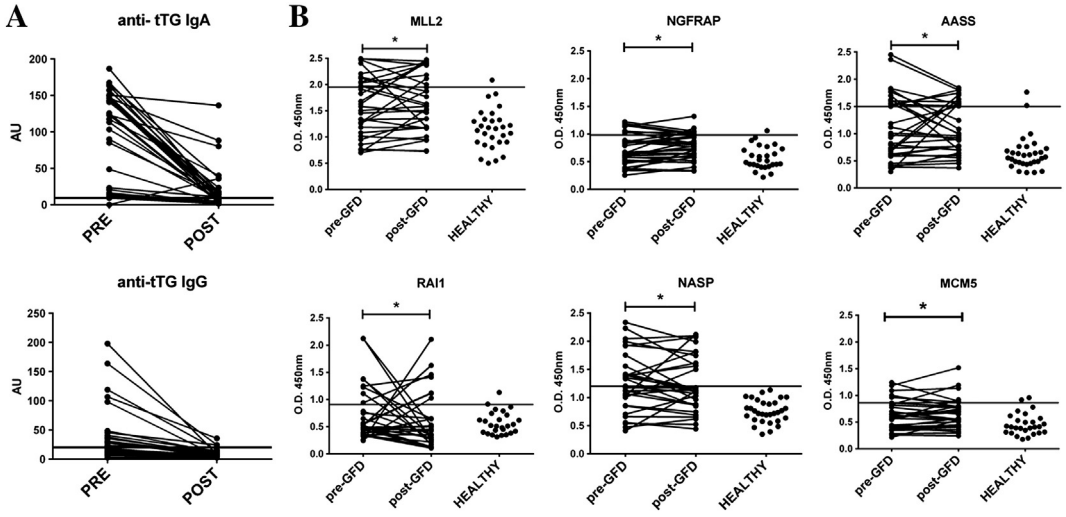

Figure 4 Gluten dependence of the antibody response. A) Antibody titers for both IgA and IgG response are shown for 31 CD patients at the time of diagnosis (pre) and after GFD (post). B) Immunoreactivity of 31 pre and post GFD sera and 27 controls is shown for the 6 top-performing antigenic fragments. 95% confidence cut-off value for positivity is shown as a line. *Statistically significant ($P < .05$).

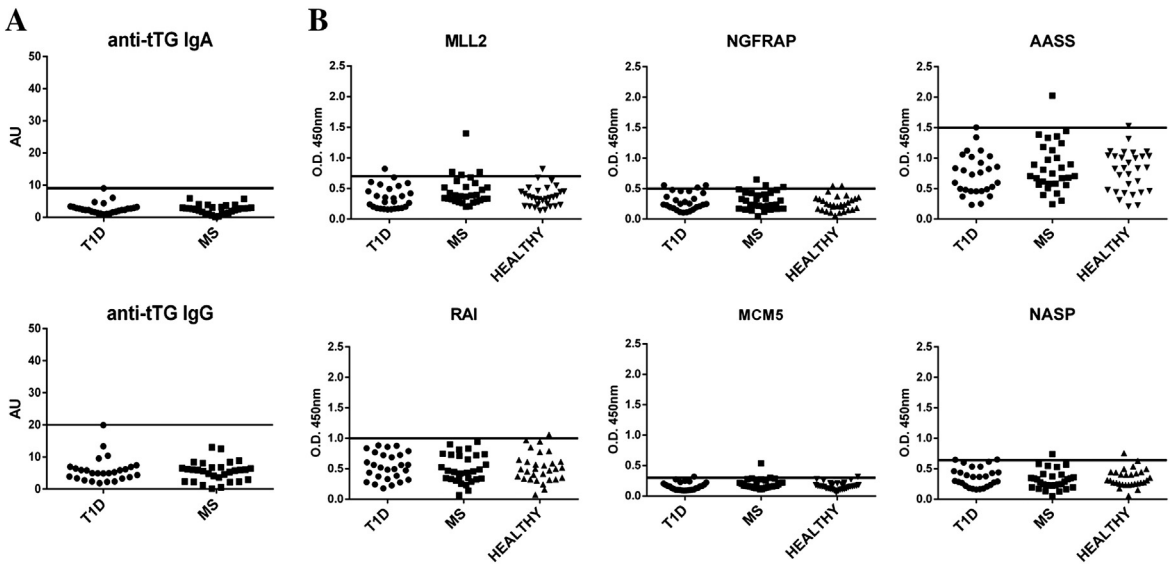

Figure 5 Autoimmunity dependence of antibody response. A) Antibody titer for both IgA and IgG response in 30 T1D (type 1 diabetes) and 30 MS (multiple sclerosis) patients sera. B) Immunoreactivity of 30 T1D and MS patient and 30 control sera towards 6 of top-performing antigenic fragments identified is shown. 95% confidence cut-off for positivity is shown as a line. *Statistically significant (P-val < 0.5).

study we used an innovative approach to identify antigenic proteins based on: i) construction of a cDNA library filtered and derived from real genes [12,31]; ii) selection with patient sera and iii) target gene identification using next generation sequencing technology and protein microarray. We have previously demonstrated that this approach can dramatically reduce the screening effort by associating the frequency of selected phages expressing sequence mapping to a defined gene to positive clone identification [12]. Together with the possibility of multiplex analysis provided by protein microarray, this combined approach led to the identification of 42 putative antigens, 13 of which were successfully confirmed to be specific for CD patient sera using protein microarray technology. When tested in ELISA, the reactivity of 6 out of 13 proteins was significantly higher with CD sera compared to healthy controls. The identified antigens were tested in different assays, with independent set of sera, thus excluding any bias coming from specific patients and increasing the confidence in the consistency of the results observed in different serological assays. Although present in the ORF library, none of the selected clones corresponded to the major CD antigen TG2. The TG2 epitopic region recognized by CD antibodies is strictly conformational and discontinuous [32] and any denaturation abolishes the serum reactivity [33]. Thus, as it is well known that conformational and/or discontinuous domains or epitopes are unlikely displayed on phage, the lack of selected ORF fragments corresponding to the CD major autoantigen (and possibly other autoantigens besides TG2) was not a surprising result. In addition, none of our identified antigens were shown to be TG2 mimotopes, as pre-absorption of sera on TG2 did not change their ability to recognize the same antigens (data not shown). The whole procedure was performed using CD serum IgGs. Although IgA anti-TG2 are routinely detected it is known that IgG antibodies to TG2 have a very high sensitivity as well [34]. Not only this, but in some cases IgG TG2 positivity improve sensitivity of the screening procedure when IgA are negative [35]. Furthermore IgA deficiency is found at high frequency in CD patients. Finally we have to consider that a systemic antibody response in CD as the one we are describing here could have a different mechanism of activation not involving primarily IgA as it is in the very specific case of TG2. Interestingly, the majority of the identified clones encode nuclear antigens, in keeping with a recent report [36] and most are involved in interaction networks with one another as well as with nucleic acids and carbohydrates. Furthermore, the products of the NASP, CDC37, MCM5, PRPF31, MATR3, MLL2 and RAI1 genes interact, directly or indirectly, with RNA or DNA. Several antigenic proteins identified have already been described as associated at various degrees with some diseases [37–39] and in some types of cancer [40]. However, none of the proteins encoded by these genes have been previously described as targets of autoimmune responses. Only 3 clones were found to encode protein fragments previously associated with autoimmune conditions other than CD. The homocysteine-induced ER protein (Herp), the product of the HERPUD 1 gene, is involved in protein degradation in the endoplasmic reticulum and has been proposed to mimic DNA structural determinants [41]. Accordingly, as Herp may be immunogenic *in vivo*, it represents a candidate triggering autoantigen for anti-DNA responses. Alpha fodrin, encoded by the SPTAN1 gene, is a widely expressed protein with a broad range of activities that was originally described as a relevant autoantigen with potential involvement in the initial stages of Sjogren's syndrome [42]. Both IgA and IgG antibodies to alpha fodrin have been described [43], but their usefulness in diagnostic practice is limited due to their presence in other conditions, including systemic lupus erythematosus (SLE), rheumatoid

arthritis and systemic sclerosis [44]. Our finding that such antibodies are also present in the serum of 7 CD patients not diagnosed with Sjogren's syndrome indicates a further lack of specificity, suggesting that a general failure in tolerance mechanisms common to several autoimmune conditions rather than to a specific disease-associated mechanism is the basis for the production of these autoantibodies. The nuclear autoantigenic sperm protein (encoded by the NASP gene) is a human histone chaperone [45] that owes its name to the ability to elicit autoantibodies following immunization in rabbits. There is some evidence that targeting of this protein by autoantibodies may result in reproductive failure in animal models [46]. Intriguingly this could be of some relevance since the role of untreated CD on male fertility is still a controversial issue [47,48]. Somewhat surprisingly, the antibody titers to the 6 most informative antigens were not decreasing following a GFD, whereas, as expected, the anti-TG2 response over the period of the diet is significantly reduced. Consistent with previous reports that diet therapy takes almost one year [15] to abolish the anti-TG2 response it is possible that the response to other autoantigens requires even longer gluten free diets for reductions to occur. This hypothesis is partially supported by the fact that a number of patients became negative for some of the antigens after the 6 months of GFD. On the other hand, the lack of a response suggests that the regulation of antibody response to TG2 and these newly described antigens is fundamentally different. Perhaps TG2 serves as the initial trigger for the autoimmune response, but once activated, responses to the other antigens is unaffected by removal of the root cause: gluten.

It is also important to note that the responses to these autoantigens are specific for CD and are not shared with at least the two other autoimmune conditions (MS and T1D) we examined. This indicates for the first time that although CD has a genetic background common to these and other autoimmune diseases [49], a specific and unique response over and above that directed towards TG2, also exists. None of these proteins was identified in the recent genome wide association studies that have revealed several new candidate genes for celiac disease [49,50]. None of the antigens alone provides sufficient sensitivity and specificity as diagnostic tool but using a small classifier of 3 independent proteins specificity and sensitivity values raised up to 80%. Therefore we cannot exclude the possibility that this classifier, after being validated in other datasets, could be helpful in the diagnosis especially in the case of sero-negative patients and in patients that have already started a gluten-free diet before the final analysis. From a general clinical point of view, the availability of this integrated discovery platform will be useful to define the autoimmune profiles of untreated CD patients with the possibility of identifying autoantibodies associated with clinical features that may be predictive of the development of later specific autoimmune disorders. The identification of such "pre-disease auto-antibodies" may provide patients with additional motivation to comply with GFD in order to avoid such additional autoimmune disorders. Since evidence suggests that GFD may be able to abolish the production of antibodies before the development of clinically apparent disease [51], this should be considered an important area for further study, particularly in the planning of autoimmune disorder prevention trials. Similar diagnostic approaches may even be applicable to the recently described cryptic genetic gluten intolerance [52].

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Acknowledgments

We thank Fulvio Adorni for statistical advice.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.clim.2013.04.009.

References

[1] K.S. Anderson, J. LaBaer, The sentinel within: exploiting the immune system for cancer biomarkers, J. Proteome Res. 4 (2005) 1123–1133.
[2] V.A. Somers, R.J. Brandwijk, B. Joosten, P.T. Moerkerk, J.W. Arends, P. Menheere, W.O. Pieterse, A. Claessen, R.J. Scheper, H.R. Hoogenboom, S.E. Hufton, A panel of candidate tumor antigens in colorectal cancer revealed by the serological selection of a phage displayed cDNA expression library, J. Immunol. 169 (2002) 2772–2780.
[3] Z. Kalnina, K. Silina, I. Meistere, P. Zayakin, A. Rivosh, A. Abols, M. Leja, O. Minenkova, D. Schadendorf, A. Line, Evaluation of T7 and lambda phage display systems for survey of autoantibody profiles in cancer patients, J. Immunol. Methods 334 (2008) 37–50.
[4] V. Somers, C. Govarts, K. Somers, R. Hupperts, R. Medaer, P. Stinissen, Autoantibody profiling in multiple sclerosis reveals novel antigenic candidates, J. Immunol. 180 (2008) 3957–3963.
[5] Y. Zhang, J.L. Davis, W. Li, Identification of tribbles homolog 2 as an autoantigen in autoimmune uveitis by phage display, Mol. Immunol. 42 (2005) 1275–1281.
[6] E.H. Kemp, L.M. Herd, E.A. Waterman, A.G. Wilson, A.P. Weetman, P.P. Watson, Immunoscreening of phage-displayed cDNA-encoded polypeptides identifies B cell targets in autoimmune disease, Biochem. Biophys. Res. Commun. 298 (2002) 169–177.
[7] B.N. Spatola, J.A. Murray, M. Kagnoff, K. Kaukinen, P.S. Daugherty, Antibody repertoire profiling using bacterial display identifies reactivity signatures of celiac disease, Anal. Chem. 85 (2) (Jan. 15 2013) 1215–1222.
[8] R.A. Young, R.W. Davis, Efficient isolation of genes by using antibody probes, Proc. Natl. Acad. Sci. U. S. A. 80 (1983) 1194–1198.
[9] P. Seto, H. Hirayu, R.P. Magnusson, J. Gestautas, L. Portmann, L.J. DeGroot, B. Rapoport, Isolation of a complementary DNA clone for thyroid microsomal antigen. Homology with the gene for thyroid peroxidase, J. Clin. Invest. 80 (1987) 1205–1208.
[10] W. Li, ORF phage display to identify cellular proteins with different functions, Methods 58 (2012) 2–9.
[11] A.R. Bradbury, S. Sidhu, S. Dubel, J. McCafferty, Beyond natural antibodies: the power of *in vitro* display technologies, Nat. Biotechnol. 29 (2011) 245–254.
[12] R. Di Niro, A.M. Sulic, F. Mignone, S. D'Angelo, R. Bordoni, M. Iacono, R. Marzari, T. Gaiotto, M. Lavric, A.R. Bradbury, L. Biancone, D. Zevin-Sonkin, G. De Bellis, C. Santoro, D. Sblattero, Rapid interactome profiling by massive sequencing, Nucleic Acids Res. 38 (2010) e110.

[13] U. Ravn, F. Gueneau, L. Baerlocher, M. Osteras, M. Desmurs, P. Malinge, G. Magistrelli, L. Farinelli, M.H. Kosco-Vilbois, N. Fischer, By-passing *in vitro* screening—next generation sequencing technologies applied to antibody display and *in silico* candidate selection, Nucleic Acids Res. 38 (2010) e193.
[14] T. Pederson, The immunome, Mol. Immunol. 36 (1999) 1127–1128.
[15] A. Tommasini, T. Not, V. Kiren, V. Baldas, D. Santon, C. Trevisiol, I. Berti, E. Neri, T. Gerarduzzi, I. Bruno, A. Lenhardt, E. Zamuner, A. Spano, S. Crovella, S. Martellossi, G. Torre, D. Sblattero, R. Marzari, A. Bradbury, G. Tamburlini, A. Ventura, Mass screening for coeliac disease using antihuman transglutaminase antibody assay, Arch. Dis. Child. 89 (2004) 512–515.
[16] P.H. Green, C. Cellier, Celiac disease, N. Engl. J. Med. 357 (2007) 1731–1743.
[17] W. Dieterich, T. Ehnis, M. Bauer, P. Donner, U. Volta, E.O. Riecken, D. Schuppan, Identification of tissue transglutaminase as the autoantigen of celiac disease, Nat. Med. 3 (1997) 797–801.
[18] W. Dieterich, E. Laag, H. Schopper, U. Volta, A. Ferguson, H. Gillett, E.O. Riecken, D. Schuppan, Autoantibodies to tissue transglutaminase as predictors of celiac disease, Gastroenterology 115 (1998) 1317–1321.
[19] S. Sulkanen, T. Halttunen, K. Laurila, K.L. Kolho, I.R. Korponay-Szabo, A. Sarnesto, E. Savilahti, P. Collin, M. Maki, Tissue transglutaminase autoantibody enzyme-linked immunosorbent assay in detecting celiac disease, Gastroenterology 115 (1998) 1322–1328.
[20] A. Rostom, C. Dube, A. Cranney, N. Saloojee, R. Sy, C. Garritty, M. Sampson, L. Zhang, F. Yazdi, V. Mamaladze, I. Pan, J. MacNeil, D. Mack, D. Patel, D. Moher, The diagnostic accuracy of serologic tests for celiac disease: a systematic review, Gastroenterology 128 (2005) S38–S46.
[21] A. Di Sabatino, G.R. Corazza, Coeliac disease, Lancet 373 (2009) 1480–1493.
[22] A. Ventura, G. Magazzu, L. Greco, Duration of exposure to gluten and risk for autoimmune disorders in patients with celiac disease. SIGEP Study Group for Autoimmune Disorders in Celiac Disease, Gastroenterology 117 (1999) 297–303.
[23] M.G. Clemente, M.P. Musu, F. Frau, G. Brusco, G. Sole, G.R. Corazza, S. De Virgiliis, Immune reaction against the cytoskeleton in coeliac disease, Gut 47 (2000) 520–526.
[24] L. Tuckova, K. Karska, J.R. Walters, M. Michalak, P. Rossmann, S. Krupickova, E.F. Verdu, R. Saalman, L.A. Hanson, H. Tlaskalova-Hogenova, Anti-gliadin antibodies in patients with celiac disease cross-react with enterocytes and human calreticulin, Clin. Immunol. Immunopathol. 85 (1997) 289–296.
[25] J. Stulik, L. Hernychova, S. Porkertova, O. Pozler, L. Tuckova, D. Sanchez, J. Bures, Identification of new celiac disease autoantigens using proteomic analysis, Proteomics 3 (2003) 951–956.
[26] A. Alaedini, P.H. Green, Autoantibodies in celiac disease, Autoimmunity 41 (2008) 19–26.
[27] R. Shaoul, A. Lerner, Associated autoantibodies in celiac disease, Autoimmun. Rev. 6 (2007) 559–565.
[28] G. Zanoni, R. Navone, C. Lunardi, G. Tridente, C. Bason, S. Sivori, R. Beri, M. Dolcino, E. Valletta, R. Corrocher, A. Puccetti, In celiac disease, a subset of autoantibodies against transglutaminase binds toll-like receptor 4 and induces activation of monocytes, PLoS Med. 3 (2006) e358.
[29] R. Di Niro, S. D'Angelo, P. Secco, R. Marzari, C. Santoro, D. Sblattero, Profiling the autoantibody repertoire by screening phage-displayed human cDNA libraries, Methods Mol. Biol. 570 (2009) 353–369.
[30] R.A. Hoskins, M. Stapleton, R.A. George, C. Yu, K.H. Wan, J.W. Carlson, S.E. Celniker, Rapid and efficient cDNA library screening by self-ligation of inverse PCR products (SLIP), Nucleic Acids Res. 33 (2005) e185.
[31] S. D'Angelo, N. Velappan, F. Mignone, C. Santoro, D. Sblattero, C. Kiss, A.R. Bradbury, Filtering "genic" open reading frames from genomic DNA samples for advanced annotation, BMC Genomics 12 (Suppl. 1) (2011) S5.
[32] Z. Simon-Vecsei, R. Kiraly, P. Bagossi, B. Toth, I. Dahlbom, S. Caja, E. Csosz, K. Lindfors, D. Sblattero, E. Nemes, M. Maki, L. Fesus, I.R. Korponay-Szabo, A single conformational transglutaminase 2 epitope contributed by three domains is critical for celiac antibody binding and effects, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 431–436.
[33] D. Sblattero, F. Florian, E. Azzoni, T. Zyla, M. Park, V. Baldas, T. Not, A. Ventura, A. Bradbury, R. Marzari, The analysis of the fine specificity of celiac disease antibodies using tissue transglutaminase fragments, Eur. J. Biochem. 269 (2002) 5175–5181.
[34] D. Sblattero, I. Berti, C. Trevisiol, R. Marzari, A. Tommasini, A. Bradbury, A. Fasano, A. Ventura, T. Not, Human recombinant tissue transglutaminase ELISA: an innovative diagnostic assay for celiac disease, Am. J. Gastroenterol. 95 (2000) 1253–1257.
[35] C. Trevisiol, A. Ventura, V. Baldas, A. Tommasini, D. Santon, S. Martelossi, G. Torre, I. Berti, A. Spano, S. Crovella, A. Amoroso, D. Sblattero, R. Marzari, A. Bradbury, T. Not, A reliable screening procedure for coeliac disease in clinical practice, Scand. J. Gastroenterol. 37 (2002) 679–684.
[36] A. Picarelli, L. Sabbatella, M. Di Tola, M. Silano, A. Nicolussi, S. D'Inzeo, A. Coppa, Nuclear fluorescence serum reactivity on monkey oesophagus: a new antibody for the follow-up of coeliac disease? Clin. Exp. Immunol. 161 (2010) 417–425.
[37] A. Li, S. Davila, L. Furu, Q. Qian, X. Tian, P.S. Kamath, B.F. King, V.E. Torres, S. Somlo, Mutations in PRKCSH cause isolated autosomal dominant polycystic liver disease, Am. J. Hum. Genet. 72 (2003) 691–703.
[38] W. Bi, G.M. Saifi, C.J. Shaw, K. Walz, P. Fonseca, M. Wilson, L. Potocki, J.R. Lupski, Mutations of RAI1, a PHD-containing protein, in nondeletion patients with Smith–Magenis syndrome, Hum. Genet. 115 (2004) 515–524.
[39] E.N. Vithana, L. Abu-Safieh, M.J. Allen, A. Carey, M. Papaioannou, C. Chakarova, M. Al-Maghtheh, N.D. Ebenezer, C. Willis, A.T. Moore, A.C. Bird, D.M. Hunt, S.S. Bhattacharya, A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11), Mol. Cell 8 (2001) 375–381.
[40] T.J. Dudderidge, J.D. Kelly, A. Wollenschlaeger, O. Okoturo, T. Prevost, W. Robson, H.Y. Leung, G.H. Williams, K. Stoeber, Diagnosis of prostate cancer by detection of minichromosome maintenance 5 protein in urine sediments, Br. J. Cancer 103 (2010) 701–707.
[41] Y. Hirabayashi, Y. Oka, T. Ikeda, H. Fujii, T. Ishii, T. Sasaki, H. Harigae, The endoplasmic reticulum stress-inducible protein, Herp, is a potential triggering antigen for anti-DNA response, J. Immunol. 184 (2010) 3276–3283.
[42] N. Haneji, T. Nakamura, K. Takio, K. Yanagi, H. Higashiyama, I. Saito, S. Noji, H. Sugino, Y. Hayashi, Identification of alpha-fodrin as a candidate autoantigen in primary Sjogren's syndrome, Science 276 (1997) 604–607.
[43] T. Witte, T. Matthias, F.C. Arnett, H.H. Peter, K. Hartung, C. Sachse, R. Wigand, A. Braner, J.R. Kalden, H.J. Lakomek, R.E. Schmidt, IgA and IgG autoantibodies against alpha-fodrin as markers for Sjogren's syndrome. Systemic lupus erythematosus, J. Rheumatol. 27 (2000) 2617–2620.
[44] A. Ruffatti, P. Ostuni, P. Grypiotis, C. Botsios, M. Tonello, C. Grava, M. Favaro, S. Todesco, Sensitivity and specificity for primary Sjogren's syndrome of IgA and IgG anti-alpha-fodrin antibodies detected by ELISA, J. Rheumatol. 31 (2004) 504–507.
[45] R.T. Richardson, O.M. Alekseev, G. Grossman, E.E. Widgren, R. Thresher, E.J. Wagner, K.D. Sullivan, W.F. Marzluff, M.G. O'Rand, Nuclear autoantigenic sperm protein (NASP), a linker histone chaperone that is required for cell proliferation, J. Biol. Chem. 281 (2006) 21526–21534.

[46] M. Wang, J.L. Shi, G.Y. Cheng, Y.Q. Hu, C. Xu, The antibody against a nuclear autoantigenic sperm protein can result in reproductive failure, Asian J. Androl. 11 (2009) 183–192.
[47] J.F. Ludvigsson, J. Ludvigsson, Coeliac disease in the father affects the newborn, Gut 49 (2001) 169–175.
[48] D. Zugna, L. Richiardi, O. Akre, O. Stephansson, J.F. Ludvigsson, Celiac disease is not a risk factor for infertility in men, Fertil. Steril. 95 (2011) 1709–1713, (e1701–1703).
[49] C. Cotsapas, B.F. Voight, E. Rossin, K. Lage, B.M. Neale, C. Wallace, G.R. Abecasis, J.C. Barrett, T. Behrens, J. Cho, P.L. De Jager, J.T. Elder, R.R. Graham, P. Gregersen, L. Klareskog, K.A. Siminovitch, D.A. van Heel, C. Wijmenga, J. Worthington, J.A. Todd, D.A. Hafler, S.S. Rich, M.J. Daly, Pervasive sharing of genetic effects in autoimmune disease, PLoS Genet. 7 (2011) e1002254.
[50] P.C. Dubois, G. Trynka, L. Franke, K.A. Hunt, J. Romanos, A. Curtotti, A. Zhernakova, G.A. Heap, R. Adany, A. Aromaa, M.T. Bardella, L.H. van den Berg, N.A. Bockett, E.G. de la Concha, B. Dema, R.S. Fehrmann, M. Fernandez-Arquero, S. Fiatal, E. Grandone, P.M. Green, H.J. Groen, R. Gwilliam, R.H. Houwen, S.E. Hunt, K. Kaukinen, D. Kelleher, I. Korponay-Szabo, K. Kurppa, P. MacMathuna, M. Maki, M.C. Mazzilli, O.T. McCann, M.L. Mearin, C.A. Mein, M.M. Mirza, V. Mistry, B. Mora, K.I. Morley, C.J. Mulder, J.A. Murray, C. Nunez, E. Oosterom, R.A. Ophoff, I. Polanco, L. Peltonen, M. Platteel, A. Rybak, V. Salomaa, J.J. Schweizer, M.P. Sperandeo, G.J. Tack, G. Turner, J.H. Veldink, W.H. Verbeek, R.K. Weersma, V.M. Wolters, E. Urcelay, B. Cukrowska, L. Greco, S.L. Neuhausen, R. McManus, D. Barisani, P. Deloukas, J.C. Barrett, P. Saavalainen, C. Wijmenga, D.A. van Heel, Multiple common variants for celiac disease influencing immune gene expression, Nat. Genet. 42 (2010) 295–302.
[51] A. Ventura, E. Neri, C. Ughi, A. Leopaldi, A. Citta, T. Not, Gluten-dependent diabetes-related and thyroid-related autoantibodies in patients with celiac disease, J. Pediatr. 137 (2000) 263–265.
[52] T. Not, F. Ziberna, S. Vatta, S. Quaglia, S. Martelossi, V. Villanacci, R. Marzari, F. Florian, M. Vecchiet, A.M. Sulic, F. Ferrara, A. Bradbury, D. Sblattero, A. Ventura, Cryptic genetic gluten intolerance revealed by intestinal antitransglutaminase antibodies and response to gluten-free diet, Gut 60 (2011) 1487–1493.