Published in final edited form as:

Anal Chem. 2013 January 15; 85(2): 1215–1222. doi:10.1021/ac303201d.

Antibody repertoire profiling using bacterial display identifies reactivity signatures of celiac disease

Bradley N. Spatola[†], Joseph A. Murray[‡], Martin Kagnoff^{□,¶}, Katri Kaukinen^{§,⊥}, and Patrick S. Daugherty^{†,*}

[†] Department of Chemical Engineering, University of California, Santa Barbara, California 93106, United States [‡] Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota 55905, United States [□] Laboratory of Mucosal Immunology, Department of Medicine, University of California, San Diego, La Jolla, California 92093, United States [¶] Department of Pediatrics, University of California, San Diego, La Jolla, California 92093, United States [§] Department of Gastroenterology and Alimentary Tract Surgery, Tampere University Hospital, Tampere, Finland [⊥] Medical School, University of Tampere, Tampere, Finland

Abstract

A general strategy to identify serum antibody specificities associated with a given disease state, and peptide reagents for their detection was developed using bacterial display peptide libraries and multiparameter flow cytometry (MPFC). Using sera from patients with celiac disease (CD) (n = 45) or healthy subjects (n = 40), bacterial display libraries were screened for peptides that react specifically with antibodies from CD patients and not those from healthy patients. The libraries were screened for peptides that simultaneously cross-react with CD patient antibodies present in two separate patient groups labeled with spectrally distinct fluorophores, but do not react with unlabeled non-CD antibodies, thus affording a quantitative separation. A panel of six unique peptide sequences yielded 85% sensitivity and 91% specificity (AUC = 0.91) on a set of 60 samples not used for discovery, using leave-one-out cross-validation (LOOCV). Individual peptides were dissimilar with known CD specific antigens tissue transglutaminase (tTG) and deamidated gliadin, and classifier accuracy was independent of anti-tTG antibody titer. These results demonstrate that bacterial display/MPFC provides a highly effective tool for the unbiased discovery of disease-associated antibody specificities and peptide reagents for their detection that may have broad utility for diagnostic development.

Keywords

serum antibody; diagnostic; peptide

***Corresponding Author** To whom correspondence should be addressed: psd@engineering.ucsb.edu, Tel: (805) 893-2610, Fax: (805) 893-4731.

Supporting Information

Quantitative fluorescence histograms of pooled antibody reactivity during library sorting, heatmap of discovery phase patient reactivity with five peptides, cross-validation predictions using the SVM model, relationships between serology and histology with CD prediction probabilities, and CD classification with 28 disease control patients. This material is available free of charge via the Internet at <http://pubs.acs.org>.

The authors declare no competing financial interest.

Introduction

The circulating antibody repertoire provides a rich source of potential diagnostic information. Serum antibodies serve as one of the largest classes of clinical disease biomarkers,^{1–3} owing to their intrinsic affinity and specificity, analytical stability, and amplification by the immune system. Thus, antibody detection assays can enable rapid and accurate diagnosis of diseases with high sensitivity and specificity, and are amenable to point-of-care diagnostic applications. Unique antibody reactivity patterns or signatures, generated by multiple distinct antibody specificities, have been observed in many diseases including cancer,⁴ autoimmune,⁵ neurodegenerative,⁶ and infectious diseases.⁷ Additionally, since some disease-associated antibodies play a role in pathology,⁸ monitoring their concentration in blood can provide valuable information to guide therapy.^{9,10} Despite their established utility, the identification of serum antibody specificities that indicate disease and reagents for their detection has proven remarkably challenging. Consequently, there remains a need for robust and unbiased serum antibody specificity profiling technologies to identify signatures of disease for known and unknown etiology.

Several distinct methodologies have been developed to identify serum antibody specificities associated with disease. Protein arrays have been used to identify antibody specificities associated with breast,¹¹ and colorectal cancer,¹² Alzheimer's disease,^{6,13} and rheumatoid arthritis.¹⁴ However, antigen arrays do not enable a complete search of the human proteome, and are biased towards a limited subset of human antigens comprising the array.¹⁵ For example, short peptides and membrane proteins are not represented by current array platforms. An alternative approach to access a larger number of candidate antigens, uses random peptide, cDNA or genomic libraries displayed on the surface of M13 or T7 bacteriophage. The panning of phage display cDNA libraries constructed from prostate and colon cancer tissue^{4,16} and an ovarian cancer cell line¹⁷ has enabled identification of peptide sequences with potential diagnostic utility. Even so, phage cDNA libraries are biased towards soluble proteins that are expressed efficiently in *E. coli*, and their effective use generally requires knowledge of which tissues are most relevant to the pathology. In an effort to overcome these problems, phage display peptide libraries have been constructed using overlapping 36-mer linear fragments spanning the human proteome.¹⁸ Nevertheless, since a majority of antibody specificities are likely directed towards structural rather than linear epitopes,¹⁹ small libraries of linear peptides do not generally yield mimotopes with sufficient affinity and specificity to mimic structural epitopes. Both cDNA and proteome derived peptide libraries are less useful when non-human antigens are involved in disease.^{20–22} Furthermore, when compared to random peptide libraries, the use of protein antigen arrays and cDNA libraries may not provide adequate diversity to effectively mimic post-translationally modified peptides.

To address these limitations, and enable identification of antibody biomarkers or signatures for arbitrary pathological processes, we sought to develop a robust and quantitative screening method using unbiased bacterial display peptide libraries in conjunction with quantitative library screening using fluorescence-activated cell sorting (FACS). We demonstrated previously that very large random peptide libraries displayed on *E. coli*^{23,24} can be screened quantitatively on the basis of both affinity and specificity enabling identification of those mimotopes exhibiting the highest specificity for a mock disease sample.²⁵ Here, we investigated the utility of specificity based screening of bacterial display libraries for the identification of disease-specific antibody signatures and peptide reagents for their detection. Our results demonstrate that bacterial display coupled with FACS provides a highly effective route to identify diagnostically useful antibody signatures of disease, and may enable development of effective and versatile diagnostic tests for a broad range of unmet medical needs.

Experimental Section

Reagents and strains

All experiments were performed with *E. coli* strain MC1061.²⁶ A pool of three bacterial display libraries with the format X₁₅, X₁₂CX₃, or X₄CX₇CX₄ displayed at the N-terminus of the eCPX display scaffold²⁷ were used for peptide discovery. Bacterial cultures were grown at 37°C with vigorous shaking in LB media supplemented with chloramphenicol (CM) (34 µg/mL) for expansion. The medium was supplemented with arabinose (final concentration 0.02-0.04% w/v) to induce peptide display. Reagents were supplied as follows: streptavidin-R-phycoerythrin (SA-PE) (Invitrogen), EZ-Link NHS-PEG₄-biotin (Thermo Scientific), Alexa Fluor® 488 reactive dye (Invitrogen), Dynabeads MyOne streptavidin C1 1-µm magnetic beads (Invitrogen).

Study subjects and samples

Sera used in this study were provided by the University of Tampere (Finland) and the Mayo Clinic (Rochester, MN) (**Table 1**). The discovery phase patient cohort contained 25 CD and 20 healthy patient samples from the University of Tampere and 20 CD and 20 healthy samples from the Mayo Clinic. The entire 88 patient training and cross-validation cohort was provided by the University of Tampere. Each sample was from a distinct individual, and the discovery phase cohorts were independent of the training and cross-validation cohort. Samples were shipped frozen at -20°C and 50 µL aliquots were made upon initial thawing. All CD sera were collected at initial diagnosis while patients were still on a gluten-containing diet. Celiac patients from Finland underwent a complete small intestine mucosal biopsy and their Marsh classification was reported. Patients from the Mayo clinic cohort did not undergo biopsies but tested positive for both anti-endomysial antibodies (EMA) and anti-tissue transglutaminase (tTG) antibodies, which gives a high predictive value (~99%) for CD. Healthy controls tested negative for anti-tTG antibodies (and anti-EMA in borderline cases) and were asymptomatic. Disease control samples from the University of Tampere were tTG negative and underwent a biopsy to rule out CD.

Antibody fractionation and depletion

Antibody purification was carried out using the Thiophilic Adsorbent kit (Thermo Scientific). Pooled serum purification was performed after pooling 50 µL of serum from each patient. Serum samples (University of Tampere) were divided into pools of 12 and 13 for the CD patients and 20 for the healthy patients. Serum samples (Mayo Clinic) were divided into two pools of 10 for the tTG and EMA double positive patients and a pool of 20 for the tTG-negative samples. Individual patient antibody purifications were performed with 50 µL of serum per patient. Eluted antibody fractions were concentrated by using 100 kDa molecular weight cutoff filters (Millipore). To deplete antibody fractions of *E. coli* binding antibodies, overnight cultures of cells over expressing the library scaffold eCPX without a displayed peptide were diluted (1:50) in fresh LB/CM media and grown until the optical density at 600 nm (OD₆₀₀) was between 0.45 and 0.65 (~2 hours). Cells were induced with arabinose for 1 hour at 37°C. Cells were centrifuged at 3000 rcf for 5 min. The supernatant was removed and the cell pellet was resuspended with antibody solution (>2 × 10⁹ cells per 1 mg of antibody). Samples were incubated at room temperature on a rotary shaker for 2 hrs, were centrifuged as above, and the supernatant was recovered.²⁵

Library screening

To reduce the library size to one allowing fluorescence-activated cell sorting (FACS), a pre-enrichment was performed by magnetic cell separation (MACS)²⁸ with the following change to the protocol: unlabeled pooled healthy patient antibodies (50 µM) were pre-

incubated with the library for 30 min prior to the addition of biotinylated CD antibodies (5 μ M). The overnight culture (post-MACS) was sub-cultured and induced as above. A volume representing a 10-fold oversampling of the number of cells retained after magnetic separation was centrifuged and resuspended in pooled healthy patient Ig. Following 30 min incubation at 4°C, disease Ig was added to the cell and Ig suspension. One pool of the CD Ig was biotin-labeled and the other pool was labeled with Alexa Fluor 488 dye. Total Ig ratio was 5:1 healthy:CD for the first round of FACS and 10:1 for subsequent rounds. Samples were incubated 45 min at 4°C, washed, and labeled with 10 nM SA-PE. Cells were washed by centrifugation at 4°C, resuspended in ice-cold PBS, and sorted or analyzed on a FACSARIA™ cell sorter (Becton Dickinson). Sorted cells were grown overnight and the procedure was repeated for enrichment.²⁸ Three rounds of sorting against the first cohort of samples were performed, followed by a single round of enrichment against a new cohort of samples. Similarly, for single clonal analysis by flow cytometry, biotinylated Ig (200 nM) was used.

Statistical analysis and classification algorithm

Statistical analysis and optimization of the classification algorithm were performed using R environment version 2.14.2.²⁹ The e1071,³⁰ pROC,³¹ and ROCR³² packages were used to train a support vector machine classification algorithm, calculate area under the curve (AUC) and 95 percent confidence intervals, and obtain receiver operating characteristic (ROC) curves, respectively. The raw fluorescent signals from single patient flow cytometry assays were normalized to a “Z-score”.^{4,33} The fluorescence signal of each sample's antibodies binding to *E. coli* that do not display peptides was subtracted from the raw fluorescence signal and then a Z-transformation was applied so that each clone had a mean reactivity of zero and standard deviation of 1.0. To determine statistical significance between CD patient antibody and healthy sample antibody reactivity, the Wilcoxon rank-sum test was performed. Support vector machines (SVMs) were used for peptide mimotope classification training.^{34,35} Between 27 and 33 support vectors, depending on the random seed, were required for the highest classifier accuracy using a sigmoid kernel function. The algorithm was optimized by using the tune.svm algorithm in R to select the cost, gamma, and coef0 parameters.³⁰ The classification accuracy was determined by leave-one-out cross-validation (LOOCV) within the training set of samples from 60 CD and healthy patients. The final reported classifier accuracy, area under the curve (AUC), sensitivity and specificity are averages from four trials of LOOCV, each trial having a different random seed. Sensitivity was calculated by dividing the true positives by the sum of the true positives and false negatives. Specificity was calculated by dividing the true negatives by the sum of true negatives and false positives. The peptide classifier was further validated using an independent test set of 28 disease control patient sera. Heatmap representations were created using a heatmap builder program.³⁶

Results and discussion

Discovery of celiac disease specific antibody signatures using bacterial display

To enable the identification of antibody signatures of disease with diagnostic utility, we developed a quantitative, specificity-based screening method to identify peptide mimotopes from random peptide libraries that capture disease specific serum antibodies (**Figure 1**). Celiac disease (CD) was selected to demonstrate the utility of this approach since it involves a relatively well understood pathology³⁷ and is known to involve disease specific serum antibodies against tissue transglutaminase (tTG)³ and deamidated gliadin peptides (dGP).²⁰ The recent identification of tTG and dGP as CD specific antigens with diagnostic utility required roughly 40 years of basic research. Celiac disease is a complex systemic disease propagated by an abnormal immune response to peptides in dietary wheat gluten and similar

prolamins in barley and rye. Diagnosis of CD is facilitated by serological testing for tTG and/or dGP antibodies and confirmed by histological examination of a biopsy of the small intestine. Prompt diagnosis is important to prevent substantial morbidity and possibly mortality associated with untreated CD.^{37,38} Yet, definitive diagnosis has been estimated to require 11 years on average from the first presentation,³⁹ and many individuals remain undiagnosed.⁹ Antibody specificities other than those against tTG and dGP have been reported to be present in patients with CD,⁴⁰ but their diagnostic utility has not been systematically examined. A previous attempt to profile CD patient antibodies using a small peptide display library and a non-quantitative panning method did not identify diagnostically useful mimotopes.⁴¹ Given this, we sought to determine whether the coupling of quantitative specificity-based screening via FACS with high quality bacterial display peptide libraries could provide an effective route to identify antibody signatures of CD.

To increase the probability that the peptides recovered would be both disease specific and cross-reactive with multiple patients' Ig, four pools of CD patient sera from two geographically distinct cohorts, exhibiting diverse severity as assessed by Marsh staging and tTG titers, were used for library screening (**Table 1**). Healthy controls were those without symptoms and with negative tTG serology. A bacterial display peptide library comprised of three distinct pooled libraries of the form X_{15} , $X_{12}CX_3$, and $X_4CX_7CX_4$ with 1.2×10^{10} members was used for mimotope discovery.²⁷ To facilitate quantitative specificity-based screening via FACS (**Figure 2**), in an Ig isotype independent manner, antibody fractions were conjugated with spectrally distinct fluorophores. Antibodies that recognize native antigens on *E. coli* were depleted from the pool by incubation with *E. coli* cells prior to antibody labeling with biotin or Alexa488 dye.²⁵ Use of two-color screening enabled direct enrichment of clones that display peptides that cross-react with Ig present in two distinct CD patient pools, and that are not blocked by unlabeled Ig from healthy patients. The specificity of the enriched library pools for CD patient Ig over healthy patient Ig was explicitly quantified by during each cycle of screening (**Figure S1, Supporting Information**). Thus, unlike reported phage-based profiling methods that involve sequential panning with one or two separate disease samples,^{7,42} our method recovered only clones that bind simultaneously to Ig from two distinct CD patient groups.

Reported antibody profiling strategies typically rely upon the use of defined protein arrays^{11,12} that provide fractional coverage of the human proteome or use cDNA libraries derived from specific organs or tissues in conjunction with phage display.^{4,16,17} In contrast, the present approach using a diverse random peptide library coupled with quantitative library screening requires no prior knowledge about the disease of interest or tissue origin of the antibody biomarkers. Furthermore, this random peptide library approach is not limited to autoantigen discovery since antigens derived from pathogens and allergens can be effectively mimicked. Although not a prerequisite for diagnostic development, the use of a high quality random peptide library can enable direct antigen identification,^{43,44} without bias toward human open reading frames. Here, both linear and disulfide-constrained peptide libraries were pooled to provide a broad range of potential structures for antigen epitope mimicry.

The capability to quantify mimotope specificity during MPFC screening offers important advantages over panning of phage libraries. In the latter approach, each round of positive selection is based primarily on affinity between ligand and antibody, and specificity is not explicitly accounted for despite its obvious importance. Consequently, peptides identified using panning may not possess sufficient disease Ig specificity or desired cross-reactivity to improve diagnostic sensitivity. Both affinity and specificity were favored during MPFC screening because disease and control patient antibodies were incubated simultaneously with the peptide library. Using MPFC, the stringency of each round of screening can be adjusted

quantitatively in response to a quantitative measurement of cross-reactivity (green and red fluorescence) and disease specificity (signal intensity) (**Figure 2**).

Down-selection of library isolated peptides

The final enriched library population was plated and peptide sequences from individual colonies were determined (**Figure 1B**). Roughly 200 peptide sequences were obtained from the sorted library population, yielding 44 unique peptide sequences. To identify clones exhibiting maximal cross-reactivity and specificity for CD Ig, the unique clones were first assayed for binding to Ig from at least four disease patients from the discovery phase using flow cytometry (**Figure 1C**). A total of 22 clones with mean fluorescence intensities greater than two-fold above background (i.e., *E. coli* that do not display peptides) were retained for training. In addition, five of the best performing clones were further assayed with Ig fractions from 14 CD patients and 22 healthy controls from the discovery cohorts to assess whether the peptides could cross-react with Ig from CD patients and not bind appreciably to healthy patient Ig (**Figure S2, Supporting Information**). Antibody reactivity with these clones was significantly higher for CD patients than for healthy patients (Wilcoxon rank-sum test p -value < 0.001), as measured by flow cytometry.

To identify individual peptides from the set of 22 with the highest utility for CD patient diagnosis, clones were assayed for binding to new patient samples that were not used for discovery. First, each clone was assayed for binding to ten healthy patient samples, and clones exhibiting fluorescence signals less than three-fold above background were retained. Ten clones that met this condition were further assayed with 10 CD samples. Six peptides yielded significant (p -value < 0.05) differences in reactivity between CD and healthy patient Ig and were retained for combined array training and cross-validation phases. The remaining peptides (**Figure 3**) were dissimilar with the primary amino acid sequence of tTG and deamidated gliadin. Four peptides (9, 10, 55, 56) possessed a Cys-Lys consensus motif with a preference for an aromatic residue (W/Y) N-terminal to cysteine. However, these sequences did not enable unambiguous candidate antigen identification using protein database search algorithms. Three clones (9, 55, 56) originated from the $X_{12}CX_3$ library, two clones (10, 168) originated from the $X_4CX_7CX_4$ library, and one (7) was derived from the X_{15} library. Assaying these six clones with 16 additional CD patients and 24 additional healthy patient antibody fractions revealed that these peptides maintained desired CD patient Ig cross-reactivity and specificity for CD over healthy patients in the training set (**Figure 3A**). These peptides, with the exception of peptide 7, retained statistically significant reactivity differences between CD and healthy patients with the inclusion of more patient samples (**Figure 3C**).

Classification model development and cross-validation of the peptide panel

Because no single peptide identified provided high specificity alone and because each peptide's reactivity pattern with the 60 new samples was unique, the peptides were used in combination to improve classification accuracy. A support vector machine (SVM) algorithm was applied to classify patients as either Celiac or non-Celiac.^{30,33,34} A sigmoid kernel function was used to project the data to higher dimensions and allow for separation between classes, and model parameters were tuned to optimize classification. Future sample prediction accuracy was estimated using leave-one-out cross-validation (LOOCV) in quadruplicate using random sampling (**Table S1, Supporting Information**). Thus, the model was trained using 59 of the 60 samples and used to predict the class of the remaining sample. This process was repeated 60 times for each of the four random samplings so that each sample was used as the validation sample exactly once per sampling. The predictions for each sample were averaged over the four trials and prediction error was calculated (**Table S1, Supporting Information**).

The true positive rate (sensitivity) was plotted against the false positive rate (1-specificity) in a receiver operating characteristic (ROC) curve as a function of varying thresholds of class prediction probabilities (**Figure 4**). Use of a six peptide panel enabled discrimination of CD patients from healthy controls with an area under the curve (AUC) of 0.91 (95% confidence interval range, 0.88 to 0.95). Using a predicted class probability threshold of 0.45, sixty patients were classified with 88% overall classifier accuracy, with 85% sensitivity and 91% specificity. A cutoff of 0.5 similarly classifies patients with 87% accuracy, but higher specificity of 94% and lower sensitivity of 77%. Since the prevalence of CD is approximately 1 in 100 Caucasians on the population level but may be as high as 1 in 10 in first degree relatives,^{45–47} multiple prediction probability thresholds may be useful depending on the type of screening being performed. Thus, using a relatively small panel of non-optimized peptide reagents, we were able to obtain sensitivity and specificity values that approach those of current serological tests. The accuracy of anti-tTG IgA ELISA tests has demonstrated variable diagnostic sensitivity (85–98%) and specificity (95–98%) in several different studies,^{48–50} but may be lower in routine clinical use.⁵¹ Similarly, tests for the presence of IgA antibodies against deamidated gliadin have been reported to yield average sensitivities of 88% and specificities of 95%.⁵⁰

Classifier accuracy was independent of tTG titer suggesting the peptides were not mimotopes of tTG (**Figure S3, Supporting Information**). Interestingly, each of three CD patients in the test set that were *negative* for either EMA or tTG antibodies when both tests were performed were classified correctly using the peptide panel. It is possible that additional tTG/dGP negative but biopsy positive CD patients could be correctly classified using the peptide panel; however, the scarcity of such samples makes evaluation problematic for this application. Diagnostic sensitivity obtained with the panel was apparently independent of the severity of villous atrophy since there was no trend between Marsh stage of the 26 untreated CD patients and the peptide classifier predictions (**Figure S4, Supporting Information**).

Independent validation set of disease control patients

To investigate whether the peptide classifier was specific to CD and not related conditions, twenty-eight disease control patients were classified using the peptide panel. This group included 13 patients that possessed the necessary CD genotype (HLA-DQ2 or -DQ8) but were confirmed by biopsy not to have the characteristic histology required for CD diagnosis. To rigorously test our peptide panel against patients that suffer from symptoms that overlap with those occurring in CD, patients with irritable bowel syndrome (IBS, $n = 10$), dyspepsia ($n = 10$), lactose intolerance ($n = 3$), collagenous colitis ($n = 1$), colitis ulcerosa ($n = 1$), and non-specific discomfort ($n = 3$) were classified as CD or non-CD using our SVM model. With a probability threshold of 0.5, 22/28 patients (79%) were correctly classified as not having CD (**Figure S5, Supporting Information**). The six misclassifications were from four dyspepsia patients (two with HLA-DQ2), one patient with IBS, and one collagenous colitis patient, although no disease showed a statistically significant difference of reactivity in comparison to the other groups. Additionally, we cannot definitively rule out the possibility that biopsy results were falsely negative.⁵² Interestingly, patients with dyspepsia and IBS have been reported to exhibit positive tTG serology more frequently than the general population.⁵³

Conclusions

In summary, our results demonstrate that cell display technology provides a remarkably effective route to identify diagnostically useful serum antibody signatures of disease and peptide capture reagents for their detection. The coupling of multicolor screening with

bacterial display peptide libraries provides an alternative to phage cDNA and peptidome libraries, enabling an unbiased search of peptide sequence space. Consequently, the present approach may allow the identification of both self and non-self antigens from heterologous organisms without prior knowledge of the autoantigens' tissue expression patterns or organisms involved in disease onset or progression. For these reasons, specificity based screening of bacterial display libraries may be generally useful to identify novel antibody biomarkers and reagents for diagnostic assay development.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the National Institutes of Health (NIAID and NIDDK grant R01 23628 to P.S.D. and DK057892 to J.A.M.) for supporting this work.

ABBREVIATIONS

CD	celiac disease
FACS	fluorescence-activated cell sorting
MACS	magnetic-activated cell sorting
eCPX	enhanced circularly permuted OmpX
CM	chloramphenicol
SA-PE	streptavidin phycoerythrin
Ig	immunoglobulin
Ab	antibody
tTG	tissue transglutaminase
dGP	deamidated gliadin peptides
EMA	endomysial antibody
LOOCV	leave-one-out cross-validation
SVM	support vector machine
ROC	receiver operating characteristic
AUC	area under the curve
IBS	irritable bowel syndrome

REFERENCES

1. Anderson KS, LaBaer JJ. *Proteome Res.* 2005; 4:1123–1133.
2. Schellekens GA, Visser H, De Jong BAW, Van Den Hoogen FHJ, Hazes JMW, Breedveld FC, Van Venrooij WJ. *Arthritis & Rheumatism.* 2000; 43:155–163. [PubMed: 10643712]
3. Dieterich W, Ehnis T, Bauer M, Donner P, Volta U, Riecken EO, Schuppan D. *Nature Medicine.* 1997; 3:797–801.
4. Wang X, Yu J, Sreekumar A, Varambally S, Shen R, Giacherio D, Mehra R, Montie JE, Pienta KJ, Sanda MG, Kantoff PW, Rubin MA, Wei JT, Ghosh D, Chinnaiyan AM. *N. Engl. J. Med.* 2005; 353:1224–1235. [PubMed: 16177248]

5. Binder SR, Hixson C, Glossenger J. Autoimmunity Reviews. 2006; 5:234–241. [PubMed: 16697963]
6. Nagele E, Han M, DeMarshall C, Belinka B, Nagele R. PLoS ONE. 2011; 6:e23112. [PubMed: 21826230]
7. Kouzmitcheva GA, Petrenko VA, Smith GP. Clin. Diagn. Lab. Immunol. 2001; 8:150–160. [PubMed: 11139210]
8. Caja S, Mäki M, Kaukinen K, Lindfors K. Cellular & Molecular Immunology. 2011; 8:103–109. [PubMed: 21278768]
9. Kagnoff MF. Journal of Clinical Investigation. 2007; 117:41–49. [PubMed: 17200705]
10. Brennan DJ, O'Connor DP, Rexhepaj E, Ponten F, Gallagher WM. Nature Reviews Cancer. 2010; 10:605–617.
11. Anderson KS, Sibani S, Wallstrom G, Qiu J, Mendoza EA, Raphael J, Hainsworth E, Montor WR, Wong J, Park JG, Lokko N, Logvinenko T, Ramachandran N, Godwin AK, Marks J, Engstrom P, LaBaer JJ. Proteome Res. 2011; 10:85–96.
12. Pedersen JW, Blixt O, Bennett EP, Tarp MA, Dar I, Mandel U, Poulsen SS, Pedersen AE, Rasmussen S, Jess P, Clausen H, Wandall HH. International Journal of Cancer. 2011; 128:1860–1871.
13. Reddy MM, Wilson R, Wilson J, Connell S, Gocke A, Hynan L, German D, Kodadek T. Cell. 2011; 144:132–142. [PubMed: 21215375]
14. Hueber W, Kidd BA, Tomooka BH, Lee BJ, Bruce B, Fries JF, Sønderstrup G, Monach P, Drijfhout JW, van Venrooij WJ, Utz PJ, Genovese MC, Robinson WH. Arthritis & Rheumatism. 2005; 52:2645–2655. [PubMed: 16142722]
15. Kodadek T. Chemistry & Biology. 2001; 8:105–115. [PubMed: 11251285]
16. Ran Y, Hu H, Zhou Z, Yu L, Sun L, Pan J, Liu J, Yang Z. Clin Cancer Res. 2008; 14:2696–2700. [PubMed: 18451234]
17. Chatterjee M, Mohapatra S, Ionan A, Bawa G, Ali-Fehmi R, Wang X, Nowak J, Ye B, Nahhas FA, Lu K, Witkin SS, Fishman D, Munkarah A, Morris R, Levin NK, Shirley NN, Tromp G, Abrams J, Draghici S, Tainsky MA. Cancer Res. 2006; 66:1181–1190. [PubMed: 16424057]
18. Larman HB, Zhao Z, Laserson U, Li MZ, Ciccio A, Gakidis MAM, Church GM, Kesari S, LeProust EM, Solimini NL, Elledge SJ. Nature Biotechnology. 2011; 29:535–541.
19. Luzzago A, Felici F, Tramontano A, Pessi A, Cortese R. Gene. 1993; 128:51–57. [PubMed: 7685301]
20. Kaukinen K, Collin P, Laurila K, Kaartinen T, Partanen J, Mäki M. Scandinavian Journal of Gastroenterology. 2007; 42:1428–1433. [PubMed: 17852878]
21. Peeters M, Joossens S, Vermeire S, Vlietinck R, Bossuyt X, Rutgeerts P. The American Journal of Gastroenterology. 2001; 96:730–734. [PubMed: 11280542]
22. Mabey D, Peeling RW, Ustianowski A, Perkins MD. Nature Reviews Microbiology. 2004; 2:231–240.
23. Bessette PH, Rice JJ, Daugherty PS. Protein Engineering, Design and Selection. 2004; 17:731–739.
24. Rice JJ, Schohn A, Bessette PH, Boulware KT, Daugherty PS. Protein Science. 2006; 15:825–836. [PubMed: 16600968]
25. Hall SS, Daugherty PS. Protein Science. 2009; 18:1926–1934. [PubMed: 19610073]
26. Casadaban MJ, Cohen SN. J. Mol. Biol. 1980; 138:179–207. [PubMed: 6997493]
27. Rice JJ, Daugherty PS. Protein Engineering, Design and Selection. 2008; 21:435–442.
28. Kenrick S, Rice J, Daugherty P. Current Protocols in Cytometry. 2007; 4:1–4.
29. Team R. R Foundation for Statistical Computing Vienna Austria. 2010
30. Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A, Meyer MD. Package “e1071”. 2011
31. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. BMC Bioinformatics. 2011; 12:77. [PubMed: 21414208]
32. Sing T, Sander O, Beerenwinkel N, Lengauer T. *R package version* 1.0–1.
33. Stekel, D. Microarray Bioinformatics. Cambridge University Press; 2003.

34. Noble WS. *Nature Biotechnology*. 2006; 24:1565–1567.
35. Agranoff D, Fernandez-Reyes D, Papadopoulos MC, Rojas SA, Herbster M, Loosemore A, Tarelli E, Sheldon J, Schwenk A, Pollok R, Rayner CF, Krishna S. *The Lancet*. 2006; 368:1012–1021.
36. King JY, Ferrara R, Tabibiazar R, Spin JM, Chen MM, Kuchinsky A, Vailaya A, Kincaid R, Tsalenko A, Deng DX-F, Connolly A, Zhang P, Yang E, Watt C, Yakhini Z, Ben-Dor A, Adler A, Bruhn L, Tsao P, Quertermous T, Ashley EA. *Physiol. Genomics*. 2005; 23:103–118. [PubMed: 15942018]
37. Rostom A, Murray JA, Kagnoff MF. *Gastroenterology*. 2006; 131:1981–2002. [PubMed: 17087937]
38. Rubio-Tapia A, Kyle RA, Kaplan EL, Johnson DR, Page W, Erdtmann F, Brantner TL, Kim WR, Phelps TK, Lahr BD, Zinsmeister AR, Melton LJ; III, Murray JA. *Gastroenterology*. 2009; 137:88–93. [PubMed: 19362553]
39. Green PHR, Stavropoulos SN, Panagi SG, Goldstein SL, McMahon DJ, Absan H, Neugut AI. *The American Journal of Gastroenterology*. 2001; 96:126–131. [PubMed: 11197241]
40. Alaedini A, Green PH. *R. Autoimmunity*. 2008; 41:19–26.
41. Zanoni G, Navone R, Lunardi C, Tridente G, Bason C, Sivori S, Beri R, Dolcino M, Valletta E, Corrocher R, Puccetti A. *PLoS Med*. 2006; 3:e358. [PubMed: 16984219]
42. Somers V, Govarts C, Hellings N, Hupperts R, Stinissen P. *Journal of Autoimmunity*. 2005; 25:223–228. [PubMed: 16256306]
43. Frulloni L, Lunardi C, Simone R, Dolcino M, Scattolini C, Falconi M, Benini L, Vantini I, Corrocher R, Puccetti A. *New England Journal of Medicine*. 2009; 361:2135–2142. [PubMed: 19940298]
44. Mintz PJ, Kim J, Do K-A, Wang X, Zinner RG, Cristofanilli M, Arap MA, Hong WK, Troncso P, Logothetis CJ, Pasqualini R, Arap W. *Nature Biotechnology*. 2002; 21:57–63.
45. Mäki M, Mustalahti K, Kokkonen J, Kulmala P, Haapalahti M, Karttunen T, Ilonen J, Laurila K, Dahlbom I, Hansson T, Höpfl P, Knip M. *The New England journal of medicine*. 2003; 348:2517–2524. [PubMed: 12815137]
46. Rubio-Tapia A, Ludvigsson JF, Brantner TL, Murray JA, Everhart JE. *Am. J. Gastroenterol*. 2012; 107:1538–1544. [PubMed: 22850429]
47. Murray JA, Moore SB, van Dyke CT, Lahr BD, Dierkhising RA, Zinsmeister AR, Melton LJ III, Kroning CM, El-Yousseff M, Czaja AJ. *Clinical Gastroenterology and Hepatology*. 2007; 5:1406–1412. [PubMed: 17919990]
48. Rubio-Tapia A, Rahim MW, See JA, Lahr BD, Wu T-T, Murray JA. *The American Journal of Gastroenterology*. 2010
49. Vermeersch P, Geboes K, Mariën G, Hoffman I, Hiele M, Bossuyt X. *Clinica Chimica Acta*. 2012; 413:1761–1767.
50. Leffler DA, Schuppan D. *The American Journal of Gastroenterology*. 2010; 105:2520–2524. [PubMed: 21131921]
51. Rashtak S, Ettore MW, Homburger HA, Murray JA. *Clinical Gastroenterology and Hepatology*. 2008; 6:426–432. [PubMed: 18304884]
52. Dewar DH, Ciclitira PJ. *Gastroenterology*. 2005; 128:S19–S24. [PubMed: 15825122]
53. Locke GR, Murray JA, Zinsmeister AR, Melton LJ, Talley NJ. *Mayo Clinic Proceedings*. 2004; 79:476–482. [PubMed: 15065612]

Figure 1. Antibody repertoire profiling methodology using bacterial display and multiparameter flow cytometry (MPFC)

A. Pools of CD patient antibodies are labeled with spectrally distinct fluorophores and incubated with the peptide display library simultaneously. Pooled unlabeled healthy patient antibodies in excess are used as competition for peptide binding sites. Cells exhibiting both high red AND high green fluorescence were collected using FACS. **B.** Peptides sequences are identified by DNA sequencing from colonies and grouped into consensus families. **C.** Unique bacterial clones are assayed with individual patient's serum antibodies by flow cytometry to assess each peptide's diagnostic sensitivity and specificity.

Figure 2. Multicolor FACS enables direct screening for library clones binding to multiple CD patients' antibody fractions

Sorting gates are indicated by the red lines. Contour plots of the library population after **A.** MACS, **B.** one, **C.** two, and **D.** three cycles of FACS with patient cohort 1 from the University of Tampere. **E.** The library population obtained from D, after incubation with the second cohort of CD and healthy patients from the Mayo Clinic. **F.** The final enriched library population after one cycle of FACS with the second patient cohort.

Figure 3. Reactivity of the peptides with 60 training phase samples

A. Heatmap representation of 6 peptides assayed for binding with 26 CD patient antibodies and 34 healthy controls not used in the discovery phase by flow cytometry. Individual peptides are represented in rows and patients' antibodies are represented in columns. Red indicates high reactivity, dark green indicates signal above background, and light green indicates non-reactivity. **B.** Amino acid sequences of the six peptides. **C.** Box plot representation of the training phase sample response to the six peptides. Red boxes represent the CD samples and grey boxes represent the healthy controls. *: $P < 0.005$; **: $P < 0.0001$; ns: not significant. P-values were calculated from the Wilcoxon-rank-sum test with $P < 0.05$ considered statistically significant.

Figure 4. Receiver operating characteristic (ROC) curve for the six peptide CD classifier with 60 patients not used in reagent discovery

The ROC curve is calculated using the true positive rate and false positive rate from the average class probabilities from four trials of leave-one-out cross-validation (LOOCV) predicted by the support vector machine (SVM) algorithm. The area under the curve (AUC) for the peptide classifier is 0.91 (95 percent confidence interval, 0.88 to 0.95). A classification probability threshold of 0.45 yields a sensitivity of 85% and a specificity of 91%.

Table 1

Clinical information about the patient samples used in the study

Variable	Discovery	Training & Validation
Summary		
No. of patients	85	88
Age (yr)	44.6±10.2	45.9±16.2
Male	26	29
Female	59	59
Celiac disease patients		
No. of patients	45	26
Age (yr)	45.6±11.0	45.4±16.9
Male	14	6
Female	31	20
Biopsy results available	25	26
Marsh 3a	0	2
Marsh 3b	14	15
Marsh 3c	11	9
Double seropositive (tTG&EMA)	45	23
tTG titer		
Mean (U/mL)	44.4±38.0	51.1±38.6
< 10 U/mL	11	5
10-50 U/mL	15	8
50-80 U/mL	8	5
> 80 U/mL	11	8
Healthy controls		
No. of patients	40	34
Age (yr)	43.5±9.2	50.5±15.9
Male	12	14
Female	28	20
tTG serology available	40	34
tTG positive	0	0