Celiac disease (CeD) is an autoimmune disorder induced by consuming gluten proteins from wheat, barley, and rye. Glutens resist gastrointestinal proteolysis, resulting in peptides that elicit inflammation in patients with CeD. Despite well-established connections between glutens and CeD, chemically defined, bioavailable peptides produced from dietary proteins have never been identified from humans in an unbiased manner. This is largely attributable to technical challenges, impeding our knowledge of potentially diverse peptide species that encounter the immune system. Here, we develop a liquid chromatographic-mass spectrometric workflow for untargeted sequence analysis of the urinary peptidome. We detect over 600 distinct dietary peptides, of which ~35% have a CeD-relevant T cell epitope and ~5% are known to stimulate innate immune responses. Remarkably, gluten peptides from patients with CeD qualitatively and quantitatively differ from controls. Our results provide a new foundation for understanding gluten immunogenicity, improving CeD management, and characterizing the dietary and urinary peptidomes.
In humans, the prevailing understanding of physiological digestion is that proteins are broken down into single amino acids or di- or tri-peptides before absorption as nutrients. Gluten proteins found in wheat, barley, and rye are exceptions to this tenet, which results in notable health consequences. Unusual biochemical properties of these proteins, such as a high abundance of glutamine (Gln, Q) and proline (Pro, P) residues, render them resistant to degradation by gastrointestinal proteases. Consequently, relatively long gluten peptides with intact immunotoxic epitopes accumulate in the lumen of the small intestine and cross the epithelial barrier, although the mechanism(s) of transport remain controversial. In approximately 1 in 100 individuals, an aberrant immune response to these peptides causes celiac disease (CeD), an autoimmune disorder that causes small intestinal mucosal injury characterized by villous atrophy. Common symptoms include abdominal pain, bloating, nausea, vomiting, and/or diarrhea. Extraintestinal CeD manifestations also occur, including blistering skin rashes and ataxia. At present, the only effective CeD treatment is a strict, lifelong adherence to a gluten-free diet (GFD).

Assays to probe gluten immunogenicity in CeD have typically relied on extracted or recombinant gluten peptides that are digested in vitro with gastrointestinal proteases. The resulting peptides can be used in bioassays to characterize mechanisms of immunotoxicity. In most cases, these peptides elicit adaptive immune responses mediated by HLA-DQ2 or -DQ8 antigen presentation to Th1 cells in the small intestinal mucosa of patients with CeD. Gluten peptides with alternative (i.e., non-T cell-dependent) modes of action have also been reported. Structure-function analyses using synthetic gluten peptides have revealed exquisite sequence specificity for both the HLA and T-cell receptor; single amino acid alterations can result in dramatically altered affinities, thereby altering the strength of the immune response.

Although it is well established that chemical variation in gluten peptides can influence CeD immune responses, measuring their naturally processed forms and connecting them with CeD patients’ health status is an underexplored research area. Over 20 years ago, chromatographic analysis coupled to UV detection implied the existence of gluten-derived peptides in the urine of patients with CeD. This was confirmed more recently by antibody-based methods. Indeed, most current gluten detection methods rely on monoclonal antibodies, which recognize amino acid motifs present in a subset of gluten proteins. Notwithstanding the valuable knowledge that has been gained from analyzing biospecimens with these immunoreagents, they are neither capable of revealing the exact gluten peptide sequences that are present, nor are they comprehensive in that some CeD-relevant peptides may lack the epitopes these antibodies recognize.

To address mounting evidence that many disease-relevant gluten peptides remain to be discovered, new methods are needed to recover and precisely characterize in vivo gluten digestion products in an untargeted fashion. In fact, not a single chemically defined peptide from wheat (or, to our knowledge, from any dietary protein) has ever been identified from the human circulatory or excretory systems. Consequently, very little is known about the physiological absorption, distribution, metabolism, and excretion (ADME) of gluten. Specifically, we lack insight into the chemical structures of peptides formed from in vivo digestion that may stimulate the immune system.

As a step toward filling this knowledge gap, we sought to analyze human urine by liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS). Currently, LC–MS/MS is the most widely used technique for peptide sequencing in complex biological samples. However, established LC–MS/MS methods suffer from technical limitations when applied to urinary peptidome analysis. High concentrations of urinary salts and metabolites, which are not efficiently removed by standard reversed-phase or liquid-liquid extraction procedures, can overwhelm chromatography systems and interfere with peptide detection. Here, we develop a sample preparation and LC–MS/MS method that utilizes mixed cation exchange solid-phase extraction to exclude these interfering molecules in a single step. This workflow overcomes problems we initially encountered with adapting established methods for urinary peptidomics, such as the need for time-consuming strong cation exchange purification and/or limited depth of peptide sampling. With it, we can now efficiently identify dietary gluten peptides and report the precise sequences of such peptides in the urine of human volunteers. We also undertake an exploratory clinical study, which revealed wheat-derived peptides that are substantially different in their chemical and biological properties and are differentially found in patients with CeD versus healthy controls. These peptides are attractive candidates for improving CeD diagnosis and for monitoring patient compliance to GFDs. They also set the stage for elucidating mechanisms underlying the anomalous ADME characteristics of gluten and other dietary proteins. More generally, the successful application of our urinary peptidomic workflow to CeD suggests it should be broadly applicable for the direct measurement of both endogenous and exogenous peptides present in urine.

Results

LC–MS/MS method development. Previous chromatographic and antibody-based detection methods suggested the presence of gluten-derived peptides in human urine but were incapable of directly elucidating their sequences. To determine if an unbiased LC–MS/MS approach could be used for this purpose, we initially analyzed urine samples from volunteers who had consumed a meal rich in dietary gluten by following an extraction and analysis protocol typically used in LC–MS/MS-based proteomics studies (Supplementary Methods). These exploratory efforts yielded just one gluten-derived peptide identification. Furthermore, very few endogenous human peptides were detected (Supplementary Fig. 1). In accordance with challenges previously reported with urinary peptidome experiments, we noted severe degradation of the LC column and MS electrosource spray over consecutive analyses, preventing us from analyzing multiple samples without unacceptable interruptions in instrument operation.

We, therefore, endeavored to systematically optimize each step in our peptidome extraction and LC–MS/MS analytical protocols in order to reliably detect peptides from urine. Starting with urine from an individual challenged with dietary gluten, we first confirmed it had measurable gluten levels by ELISA. The commercially available R5 monoclonal antibody we used targets the pentameric motif QXP(W/F)P found in many gluten proteins and is therefore expected to react with gluten peptides produced by patients with CeD and healthy controls. Using this reference specimen, we tested a variety of methods to enrich and purify urinary peptides. In addition, we optimized the LC gradient and MS acquisition method (Supplementary Methods). These efforts resulted in 29- and 13-fold increases in the number of identifiable wheat- and human-derived peptides, respectively, while eliminating longstanding problems with instrument contamination by background metabolites (Supplementary Methods, Supplementary Figs. 2 and 3, see also Method Specificity section below). Importantly, by combining solid-phase extraction on a mixed cation exchange column with online reversed-phase LC and high-resolution MS/MS
Fig. 1 An efficient chemical extraction and LC–MS/MS workflow for urinary peptideomic analysis. Urine sample volumes are first normalized based on creatinine measurement. To prevent the formation of intra- or intermolecular disulfide bonds that would confound downstream data analysis, cysteine (Cys) residues are reduced with dithiothreitol (DTT) and alkylated with iodoacetamide (IAM). Proteins are depleted using a centrifugal filtration device with a molecular weight cutoff (MWCO) of 10 kDa under acidic denaturing conditions. The filtrate is processed using a mixed cation exchange solid-phase extraction column to remove metabolites and salts that interfere with LC–MS/MS analysis. After extensive washing, peptides are eluted from the column with methanol containing 5% ammonium hydroxide. Eluted peptides are dried in a centrifugal vacuum concentrator and reconstituted in water. Peptides are separated by nano-liquid chromatography (LC) on a reversed-phase column (RP-C18) and analyzed by tandem mass spectrometry (MS/MS) on an Orbitrap Fusion Lumos mass spectrometer operated in the data-dependent acquisition (DDA) mode, with precursor and fragment ions analyzed in high resolution in the Orbitrap (OT). The resulting LC–MS/MS data are searched against the desired databases (e.g., the human and wheat proteomes). The experimental workflow takes ~6 h to complete and results in the identification of thousands of peptides.

Pilot identification of chemically defined urinary wheat peptides. To evaluate our optimized urine peptidome extraction and LC–MS/MS method, four healthy participants were recruited into a pilot study outlined as follows (Fig. 2a): On the first day, participants initiated a GFD. The GFD was maintained on the second day, and a pooled urine sample was collected over 8 h. On the third day, participants underwent a gluten challenge comprised of two wheat bagels (~18 g gluten), and subsequently collected a second pooled urine specimen over the next 8 h. LC–MS/MS analysis of these urine samples yielded an average of 24 unique peptide sequences per participant mapping to the wheat proteome, all of which were identified only in the post-gluten challenge samples (Supplementary Dataset 3). The number of unique wheat peptides varied widely between individual volunteers (Fig. 2b). In contrast, the number of detected human peptides was essentially unchanged in the GFD versus post-challenge samples (Supplementary Fig. 4).

In anticipation of recruiting larger cohorts of human participants, we sought to confirm our initial findings while also simplifying the dietary gluten challenge protocol. The four original participants, along with four additional healthy participants, were asked to fast overnight in place of the 48 h GFD employed in the pilot study (Fig. 2c). Then, a single urine sample was collected before consumption of two wheat bagels for breakfast and pooled urine collection over the next 8 h. Wheat peptides were readily detected in all post-challenge samples (Fig. 2d and Supplementary Dataset 4), while the number of human peptides was similar in both sample groups (Supplementary Fig. 5). These results support our initial finding that a dietary gluten challenge consistently leads to urinary gluten peptides that are measurable by LC–MS/MS. They also indicate that levels of these peptides fall to undetectable levels after overnight fasting, justifying the use of the abbreviated protocol for subsequent studies.

Considering all sequences of the wheat-derived peptides identified from healthy participants (Supplementary Datasets 3 and 4), four noteworthy characteristics were apparent (Table 1). First, although 123 unique peptide sequences were identified, only one peptide was found in common in all 8 participants: GQQQPFPPQYPQPFQPS. Its C-terminally truncated derivative, GQQQPFPPQYPQPFQP, was identified in 7/8 participants; other overlapping variants of these peptides were also detected in more than one urine sample. While these sequences from α-gliadin proteins are not known to be recognized by patient-derived T cells, multiple studies have demonstrated that they can stimulate an innate immune response. Strikingly, these prevalent urinary peptides do not have a motif that is recognized by monoclonal antibodies commonly used for gluten detection.

Second, we identified many peptides harboring known CeD-relevant T-cell epitopes. The DQ2.5-glia-y4c/DQ8-glia-y1a epitope was most common, with 32 distinct peptides harboring the corresponding QQPQPQFQ sequence. The DQ2.5-glia-y5 epitope (QQPQPQFQFQ) also appeared frequently, as it was identified in 14 unique peptides. Other known T-cell epitopes, including DQ2.5-glia-y1/DQ8.5-glia-y1/DQ8-glia-y2 (PQOSFPQF), DQ2.5-glia-y3/DQ8-glia-y1b (QQPQPQFQPF), and DQ2.5-glut-L1/DQ2.2-glut-L1 (PFSQFPQPV) were identified less frequently (Supplementary Table 1). Some T-cell stimulatory peptides, such as FLPQQPQFPQFQFPFQFQ, harbor more than one epitope and were previously identified as being both proteolytically resistant and...
highly inflammatory to CeD patients through in silico and in vitro analysis. Third, some peptides such as PyrQTFPHQPQQVPQPQPPQQPQPQQ had undergone post-translational modification via cyclization of an N-terminal Gln into a pyroglutamic acid residue. The possibility that this modification occurs during sample processing cannot be excluded. However, N-terminal pyroglutamination is known to protect peptides from proteolytic degradation and thus may contribute to their stability in circulation.

Last, we found gluten peptides from human urine that have not been described by any prior study as either being resistant to gastrointestinal digestion or as having any pathophysiological characteristics. One such peptide, SCHVMQQQQCC, is derived from γ-gliadin and from the low-molecular-weight subunits of glutenins B and C. It was identified in 6/8 participants, while its C-terminally extended form was detected in 5/8 participants. Longer and shorter versions were also detected in multiple samples. This result illustrates the power our untargeted approach has for identifying wheat-derived peptides that have eluded prior description despite their seemingly common representation.

Method specificity. To our knowledge, untargeted detection of diet-derived peptides in humans had not been reported prior to our proof-of-concept studies above. We, therefore, sought to confirm that our LC–MS/MS method reliably detects these peptides and accurately identifies their sequences in the complex background of the human urinary peptidome. To do so, we took advantage of the fact that rye and barley also contain proteolytically resistant gluten proteins. For example, wheat gliadins (one of two major protein families that comprise gluten) are homologous to secalin proteins in rye and hordeins in barley (Fig. 3a). We thus anticipated that urine analysis should distinguish the particular grain consumed by volunteers who ate meals rich in either wheat, barley, or rye. To test this hypothesis, two healthy participants fasted overnight, and then, on different days, maintained a grain-free diet, or ate breakfasts prepared with approximately 1.5 cups of wheat, barley, or rye flour. Urine was then collected for 8 h, analyzed by our optimized LC–MS/MS protocol, and the data were searched against a customized database that included the human, rye, wheat, and barley proteomes (see “Methods” section).

When participants maintained a grain-free diet, only a single grain peptide (likely a false discovery) was identified. In contrast, subsequent to wheat, rye, and barley dietary challenges, we respectively detected 51, 43, and 37 unique grain peptide sequences (Supplementary Dataset 5). The human peptide repertoire was similar in all specimens (Supplementary Fig. 6). Notably, all identified peptides were restricted to just one dietary challenge (Fig. 3b), suggesting that our method readily differentiates peptides formed from different dietary proteins. Indeed, when the sequences were mapped onto the grain proteomes, the majority were predominately derived from the corresponding dietary challenge (Fig. 3c and Supplementary Fig. 7). Our observation that fewer sequences could be mapped to the rye proteome from rye challenge urine is a likely artifact of the size of the available rye proteome: the UniProt resource contains ~100× fewer sequences than the wheat or barley proteomes, despite their genomes having similar sizes. Our data suggest that many of the detected sequences in the rye challenge urine may be derived from yet unannotated rye proteins whose sequences are also present in the wheat proteome. Taken together, these results demonstrate that the peptides detected by our method originate from the diet. They also confirm the specificity of our method to distinguish closely related dietary components.

Preliminary analysis of banked urine from patients with CeD. To test our ability to identify wheat-derived peptides in the urine of individuals with CeD, we analyzed a set of nine urine samples banked from a prior study. These urine samples were collected from single voids of patients with confirmed CeD diagnoses and were previously assayed for gluten presence by lateral flow immunoassay. This immunoassay used G12 and A1 monoclonal
Table 1 Examples of wheat-derived peptide sequences identified in human urine.

| Wheat peptide sequence | Uniprot protein accession(s) | Protein name(s) | Mass (Da) | # of participants with peptide |
|------------------------|------------------------------|-----------------|-----------|-------------------------------|
| GQQQPFPPQQPYPQQPFPFS  | GDA0_WHEAT; GDA1_WHEAT; GDA2_WHEAT; GDA9_WHEAT | α/β-gliadin; α-gliadin | 2205.0642 | 7 |
| GQQQPFPPQQPYPQQPFPFP  | GDA0_WHEAT; GDA1_WHEAT; GDA2_WHEAT; GDA9_WHEAT | α/β-gliadin; α-gliadin | 4284.0259 | 6 |
| SQQPEQTISQQQQQPFPQQPQQPQYYGSSL | A0A290XYU3_WHEAT; USUA46_WHEAT; R9X67_WHEAT; R9XUY1_WHEAT | γ-gliadin; LMW glutenin | 1336.4781 | 6 |
| GQQQPFPPQQPYPQQPFPQP  | GDA0_WHEAT; GDA1_WHEAT; GDA2_WHEAT; GDA9_WHEAT | α/β-gliadin; α-gliadin | 2558.0360 | 5 |
| SQQPQQPFPQPHOPQQPYPQ  | GDB1_WHEAT; GLTC_WHEAT; GDB3_WHEAT; GLTB_WHEAT | γ-gliadin; LMW glutenin | 1464.5367 | 5 |
| PyrQTFPHOPQQPQPQQPQPQP  | A0A0K2QJX_WHEAT; A0A0E3Z7F7_WHEAT, A0A1K0INE4_WHEAT | α/β-gliadin; α-gliadin | 2354.1807 | 2 |
| PQPQFPPQQPQQPQPQPQPQP  | A0A290XYU3_WHEAT; USUA46_WHEAT; R9X67_WHEAT | α/β-gliadin; α-gliadin | 2758.0360 | 5 |
| TQPQFPPPPQPQPPQPQPQPQP  | GDB2_WHEAT; B6DQ5_WHEAT; R9XWD0_WHEAT; R9XUB9_WHEAT | γ-gliadin; LMW glutenin | 3098.4998 | 3 |
| TQPQFPPPPQPQPPQPQPQPQP  | GDB2_WHEAT; B6DQ5_WHEAT; R9XVB7_WHEAT; R9XWD0_WHEAT | γ-gliadin | 3457.6763 | 3 |
| PyrQTFPHOPQQPQPQQPQPQP  | GLTA_WHEAT, D2DII3_WHEAT, Q6SPZ1_WHEAT, Q6SPZ3_WHEAT | γ-gliadin | 3416.6174 | 1 |

aResidues in bold indicate CeD-relevant T-cell epitopes, as defined in ref.39. All Cys residues were detected in their carbamidomethylated forms due to derivatization with iodoacetamide during sample workup. "PyrQ" indicates glutamine residues that had undergone cyclization into pyroglutamic acid.

bFor peptides mapping to more than four proteins in the UniProt database, only the first four accession codes are displayed. All identified peptide sequences and full details relating to peptide identification in each participant’s urine are provided in Supplementary Datasets 3 and 4.

cThe UniProt database contained otherwise identical peptides containing Leu and Ile at the same position. These two amino acids are isobaric and indistinguishable by our MS/MS method and thus the sequence is reported as [I/L].

dThe Met in this peptide was detected in its oxidized form.
were excluded from this challenges. Sequences that mapped to more than one of the proteomes mapping to the wheat, barley, and rye proteomes following dietary controls. In accordance with the lateral

tocol. In accordance with the lateral

t filtration, we did not analyze the results of these assays.

Fig. 3 LC-MS/MS analysis of urine from two healthy participants challenged with dietary wheat, rye, and barley. a Illustration of sequence differences between closely related gliadin, secalin, and hordein proteins from wheat, rye, and barley, respectively. The boxes highlight source-specific sequence regions. b All possible intersections of the three sets of peptide sequences identified after the grain challenges are depicted as an UpSet plot. The individual set size is plotted to the left of each row, with the black portion of the bar representing sequences that map uniquely to the respective dietary challenge proteome, and the orange portion representing those that do not. There are no intersections between these sets of peptides, suggesting the peptides depicted in orange are nonetheless likely to have originated from the specific grain used for the dietary challenge. c Number of distinct peptide sequences uniquely mapping to the wheat, barley, and rye proteomes following dietary challenges. Sequences that mapped to more than one of the proteomes were excluded from this figure. Sequences that mapped to more than one of the proteomes were excluded from this figure. In the rye diet, sequences mapping uniquely to the wheat proteome were only found in the rye diet, suggesting they are yet unannotated rye peptides. A full list of observed peptide sequences and their respective proteome mappings is provided in Supplementary Dataset 5. Participants in this experiment were males, ages 27 and 28. Source data are provided as a Source Data file.

antibodies, which specifically recognize the amino acid motifs QPQLP(Y/F) and QLP(Y/F)PQP, respectively. Samples (n = 4) from patients reporting adherence to a GFD and in which the measured gluten peptide concentrations were below the lateral flow assay limit of quantification were analyzed as negative controls. In accordance with the lateral flow results, we did not measure peptides mapping to the wheat proteome by LC-MS/MS.

Clinical study. Adult patients with gastrointestinal symptoms including dyspepsia, bloating and diarrhea who were eating gluten-containing diets and were undergoing evaluation for CeD were recruited within the Celiac Disease Program at the Stanford Digestive Health Center, along with gender-matched healthy controls (Fig. 5a). CeD vs. non-CeD diagnoses were not known at the time of enrollment and were subsequently established by serology and endoscopy with biopsy. All participants undertook a dietary gluten challenge and urine collection according to the optimized study design depicted in Fig. 2c. At the end of the recruitment period, urine samples were processed and analyzed by LC–MS/MS. Compared to healthy controls (n = 8) and patients ultimately diagnosed with non-celiac gastrointestinal disorders (n = 5), patients with CeD (n = 6) had approximately 5 times more unique gluten peptides in their urine (Fig. 5b), while the number of endogenous human peptides was similar in all three groups (Supplementary Fig. 9).

Consistent with our pilot studies, variants of the innate immune response stimulating-peptide GQQQFPFPQPYPQPQFP were detected in all individuals, regardless of clinical status (Supplementary Dataset 7). We analyzed whether patients with CeD had increased diversity of peptides with T-cell epitopes known to stimulate the adaptive immune response. Indeed, the number of peptides with at least one CeD-relevant T-cell epitope was

Fig. 4 LC-MS/MS analysis of banked urine samples from patients with CeD. Urine collected from a single void was tested for the presence of gluten by an A1 and G12 monoclonal antibody-based lateral flow assay. Each data point represents a single urine sample, and the horizontal bar indicates the median. Information associated with the identification of these peptide sequences by LC-MS/MS is provided in Supplementary Dataset 6. Source data are provided as a Source Data file.

In contrast, all samples (n = 5) for which lateral flow assays indicated high apparent gluten peptide concentrations, yielded at least 17 wheat peptide sequences (Fig. 4) by LC-MS/MS. The mean number of identified wheat peptides (74 per sample) was much higher than those described in our proof-of-concept studies with healthy participants (24 per sample; Fig. 2b, d) while the number of human peptides was similar (Supplementary Fig. 8). Strikingly, one sample from a patient with CeD yielded 206 distinct wheat peptides, a quantity four times greater than in any healthy participant analyzed previously. These intriguing observations led us to hypothesize that the urinary wheat peptide repertoires of patients with CeD are much more diverse than non-CeD individuals. We, therefore, initiated a prospective clinical study to formally test this hypothesis, described below.
approximately 4-fold higher in patients with CeD (Fig. 5c). CeD-relevant T-cell epitopes occurred more frequently in participants with CeD, and seven different T-cell epitopes completely absent in controls were found in patients (Table 2 and Supplementary Table 2). Overall, we measured 289 unique wheat peptides from patients with CeD that were not detected in healthy individuals or patients with non-celiac gastrointestinal disorders, whereas only 37 peptides were shared between all three groups (Fig. 5d and Supplementary Dataset 7).

To gain insight into the diversity of peptide sequences found in patients with CeD, we ranked the frequency by which CeD-specific peptide sequences were detected (Supplementary Dataset 8). Just four peptides were found in the majority of (≥4/6) patients (Table 3), whereas most (201) were detected only in single individuals. To understand the chemical space occupied by the peptides that were found in patients with CeD, we aligned the detected peptide sequences to the wheat proteome (Fig. 5e). Strikingly, patients with CeD not only had more variants (i.e., slightly longer or shorter versions) of peptides found in controls, but they also possessed peptides that mapped to distinct regions of the wheat proteome (Fig. 5f–h and Supplementary Fig. 10). For example, three patients had peptides spanning residues 194–199 of γ-gliadin (GDBX) that were absent in controls. Taken together, these results demonstrate that patients with CeD have peptidomes that are more diverse in their chemical and biological properties compared to control participants.

**Discussion**

Despite extensive research into the mechanisms by which gluten proteins play a role in CeD pathogenesis (Fig. 6), the identities of the actual diet-derived molecules that interact with the human
immune system in the gut and other organs have been a long-standing mystery. As a step toward illuminating these molecular interactions, we developed a LC–MS/MS-based assay that directly detects the metabolic products of food grains. In contrast to in vitro gluten preparations that have been standard research tools in the CeD field for several decades, our method is compatible with in vivo dietary challenge. It therefore can reveal the precise chemical structures of dietary peptides that may drive CeD in humans. Furthermore, unlike traditional antibody reagents, the untargeted LC–MS/MS readout readily distinguishes the precise amino acid sequences of digestion-resistant peptides from wheat, barley, and rye. Thus, this method provides a unique window into gluten–immune interactions and immunogenicity. Indeed, we observed more varied wheat peptide repertoires from patients with CeD than was previously appreciated – and ones that were much more diverse than those found in non-CeD individuals.

Initially, our efforts were hindered by LC–MS/MS methods that were unsuited to the unique challenges of urinary peptidome analysis. Urine contains high concentrations of metabolites and salts that are difficult to separate from peptides and are incompatible with LC–MS/MS. Therefore, we developed an extraction technique to remove these interfering compounds. This allowed us to achieve our main goal of wheat peptide identification, while also improving our ability to measure urinary peptides originating from other endogenous and dietary sources (Supplementary Fig. 3). Our workflow (Fig. 1) is compatible with standard reversed-phase LC–MS/MS instrumentation available in most proteomics laboratories. Compared to published approaches employing solid-phase extraction techniques, our method identifies approximately 2 to 10 times more endogenous human peptides from typical 1–10 mL urine samples. Moreover, our sample preparation technique requires only ~6 h, facilitating sufficient throughput for us to undertake a comparative analysis of the wheat-derived urinary peptidomes of patients with CeD and healthy controls. Given its high specificity (Fig. 3), our workflow has potential utility for other studies requiring analysis of the urinary peptidome. Although here we focused on the identification of wheat-derived peptides, we also identified over 30,000 human peptides (Supplementary Datasets 1–7), which is to our knowledge, the largest collection of urinary peptides sequenced by LC–MS/MS to date. Parenthetically, we note that previous surveys of the urine peptidome found that hydroxyproline-modified collagen peptides were among the most abundant. However, here we did not allow hydroxyproline as a variable modification in our database searches, as this modification was not relevant to our overriding goal of detecting biologically relevant, wheat-derived peptides. Therefore, we have deposited our raw data in the PRIDE database to facilitate the identification of additional peptides (e.g., by including other variable modifications or using alternative search engines).

This study has identified, for the first time, the specific amino acid sequences and post-translational modifications of peptides resulting from in vivo digestion of gluten. In our pilot identification of dietary wheat peptides in healthy participants (Fig. 2), we identified sequences with known CeD relevance, as well as others that had never been implicated in CeD or identified as being resistant to gastrointestinal digestion. These findings underscored the value of our untargeted LC–MS/MS method and motivated us to analyze the urine of patients with CeD.

Using banked urine specimens, we found that wheat-consuming patients with CeD had substantially greater gluten peptide diversity than non-CeD individuals (Figs. 2 and 4). We confirmed this finding with a subsequent controlled clinical study (Fig. 5) in which participants followed our well-defined gluten challenge protocol (Fig. 2d). Mechanistically, this peptide

---

**Table 2 Number of detected T-cell epitopes in patients with celiac disease (CeD), non-celiac gastrointestinal disorders (GI), and healthy controls (HC).**

| Epitope name(s) | Epitope sequence | CeD | GI | HC |
|----------------|------------------|-----|----|----|
| DQ2.5-glia-α1a | PFPQQLPY | 2 | 0 | 0 |
| DQ2.5-glia-α1b | PYPPQLPY | 1 | 0 | 0 |
| DQ2.5-glia-α2 | PQPQLPYQ | 2 | 0 | 0 |
| DQ2.5-glia-ε3 | FRQQPYQP | 1 | 0 | 0 |
| DQ2.5-glia-γ1/DQ8.5-glia-γ1 | PQQSPQ | 1 | 0 | 1|
| DQ8-glia-γ2 | | | | |
| DQ2.5-glia-γ3/DQ8-glia-γ1b | QQPQQPYQ | 6 | 1 | 0 |
| DQ2.5-glia-γ4a | SQPQFQ | 1 | 0 | 0 |
| DQ2.5-glia-γ4b | PQQPFQ | 8 | 4 | 3 |
| DQ2.5-glia-γ4c/DQ8-glia-γ1a | QQPQPQPQ | 105 | 29 | 36 |
| DQ2.5-glia-γ4d | PQQPQFQ | 28 | 5 | 4 |
| DQ2.5-glia-γ4e | LQQPQPQ | 37 | 10 | 15 |
| DQ2.5-glia-α1/DQ2.5-hor-1 | PFPQPQ | 3 | 0 | 1 |
| DQ2.5-sec-1 | | | | |
| DQ2.5-glia-α2 | PQQPQPQ | 1 | 0 | 0 |
| DQ2.5-glut-L1/DQ2.2-glut-L1 | PQSQQ | 17 | 2 | 11 |
| DQ2.5-glut-L2 | FSQQQOSP | 1 | 0 | 0 |
| DQ8.5-glut-H1/DQ8-glut-H1 | QGGYYTPQ | 2 | 0 | 1 |

*Epitope sequences and nomenclature are defined in ref. 33.*
diversity may stem from the fact that individuals with active CeD have elevated intestinal permeability, which likely allows wheat-derived peptides to more readily cross the epithelial barrier. This phenomenon could lead to increased peptide concentrations in the systemic circulation and in urine, thereby enhancing their ability to be detected. A different (or complementary) explanation is that patients with CeD digest and/or metabolize wheat differently than other individuals. Such potential differences in wheat metabolism could contribute to the onset of CeD. Future analyses of patients with other non-CeD enteropathies known to cause “leaky gut” such as tropical sprue as well as disorders in which the mucosa remains grossly intact such as irritable bowel syndrome with diarrhea should be insightful in pinpointing the origins of the increased peptide diversity observed in patients with CeD. In contrast, the sequence GQQQPFPPQQPYPQPQPFPS appeared in 2/7 patients with CeD. This peptide is also proteolytically resistant and highly inflammatory to CeD-specific T cells on account of its polyvalency.

Despite the abundance of immunogenic wheat peptides in urine, thus far we have been unable to definitively identify any peptide that underwent regioselective Gln deamidation by transglutaminase 2 (TG2), an important step in CeD pathogenesis (Fig. 6). Because TG2-catalyzed deamidation increases the immunogenicity of gluten-derived peptides by several orders of magnitude, it is possible that a relatively small molar fraction of absorbed gluten is deamidated in patients with CeD, making the concentrations of these peptide species too low to be detectable. Alternatively (or additionally), deamidation may enhance the metabolic lability of circulating gluten peptides. The possibility that their high affinity for HLA-DQ2 or -DQ8 alters their ADME characteristics should not be overlooked. Regardless, future development of targeted LC–MS/MS methods could facilitate the detection of low-abundance deamidated gluten peptides.

There is increasing interest in non-invasive monitoring of GFD compliance via urinalysis, as a large proportion of patients with CeD experience symptoms even when attempting to adhere to a GFD. Available methods predominantly exploit the G12 monoclonal antibody, which preferentially recognizes the immunotoxic QPQLP(Y/F) motif found in several CeD-relevant T-cell epitopes. Our data suggest that this motif is rare in urine from individuals who consume dietary wheat (Table 2). In contrast, the sequence GQQQPFPPQQPYPQPQFPQPS appeared in

Table 3 Wheat-derived peptide sequences detected in the majority of patients with CeD but not in control groups.

| Wheat peptide sequence | Protein name | # of CeD patients with peptide |
|------------------------|--------------|-------------------------------|
| PQOQIPQHQPQQQQQOPQ    | α-gliadin    | 5                             |
| GQQQQQFPQQOFPQQPQPQPQPQ    | α/β-gliadin | 4                             |
| PELQQIPQQPQQPQPQPQQPOQPQPQ    | α/γ-gliadin | 4                             |
| PTSPQSGQGQQQPQGQQPQPQ | HMW-glutenin | 4                             |

Fig. 6 Schematic representation of relevant metabolic products of dietary gluten in humans. Wheat-derived proteins such as gluten are resistant to gastrointestinal proteolysis. Incomplete digestion preserves disease-causing T-cell epitopes in peptides that accumulate in the gut lumen. A portion of these gluten peptides translocate to the lamina propria where transglutaminase 2 (TG2) deamidates selected Gln residues (brown dots) in some sequences, thereby converting them into Glu residues (yellow dots). Absorbed peptides can also enter the bloodstream, and are further metabolized and/or excreted in urine.
the urine of every gluten-challenged individual in this study regardless of CeD status. Most urine samples also contained longer or shorter variants of this sequence. Interestingly, a variant of this peptide has been widely investigated for its ability to actuate the innate immune response by stimulating the production of IL-15, Il-17, and IL-21, a key cytokine in CeD pathogenesis. At minimum, our findings suggest that antibody-based detection methods for monitoring GFD compliance could be vastly improved by targeting detection of this common urinary peptide.

The results reported here pave the way for designing improved diagnostic methods for CeD. Although we observed a statistically significant increase in the number of wheat-derived peptides in patients with CeD compared to healthy controls and individuals with other gastrointestinal ailments, the range of unique peptide sequences was quite broad, suggesting that peptide number alone is unlikely to be sufficient to be sensitive or selective for CeD diagnosis. Only a small subset (~2%) of 289 CeD-unique peptides occurred in the majority of patients, while ~75% of these sequences were uniquely detected in single individuals (Supplementary Dataset 8). This broad distribution of detected peptide sequences likely stems from the intrinsically high variability of protein digestion, as well as the fact that CeD is a highly heterogenous disease in terms of symptoms and extent of gastrointestinal damage. Nonetheless, a subset of diet-derived peptide sequences tended to occur more frequently in CeD patients (Fig. S5f–h and Supplementary Dataset 8). Undoubtedly, future studies using larger sample cohorts will be required to identify and validate strong candidate peptides for diagnostic purposes. Such validation studies should benefit from the development of targeted LC–MS/MS methods that reproducibly sample and quantify the peptides of interest.

Beyond monitoring and diagnosing CeD, our urine peptidomic workflow has the potential to advance our fundamental understanding of gluten immunogenicity. Recent evidence indicates that only half of gluten-reactive T cells recognize a known epitope. Analysis of urine from gluten-challenged individuals with T cells of unknown sequence specificity may reveal candidate epitopes that can be tested in immunoassays. Knowledge of the complete repertoire of disease-relevant wheat peptide sequences that encounter the immune system in vivo will be critical to advance efforts to treat CeD by sequestering, degrading, or blocking the interaction of these peptides with immune cells.

Generally, our sample extraction and LC–MS/MS analysis protocol promises to reveal new biological insights currently hidden in the urinary peptidome. For example, there is a longstanding interest in bioactive peptides derived from dietary proteins other than gluten, but little direct evidence for their existence in humans. More broadly, while it is appreciated that proteins other than gluten, but little direct evidence for their immunogenicity.
extraction of peptide-enriched urine was performed using Oasis Mixed Cation Exchange Prime Columns (3 cm x 100 mg resin, Waters Corporation #186002). The column was washed with 1 mL of an aqueous solution of 2% formic acid containing 100 mM ammonium formate, followed by 1 mL of methanol. The resin was re-equilibrated with 2 mL aqueous 2% formic acid, then washed with 1 mL 95% water/5% ammonium hydroxide followed by 1 mL 95% acetonitrile/5% ammonium hydroxide. The reequilibration and wash steps were repeated one additional time. Peptides were eluted in 1 mL 95% methanol/5% ammonium hydroxide directly into a low-binding microcentrifuge tube (Fisher #3453) at a rate of approximately 1 drop every 2 s and dried using vacuum centrifugation. Dried peptides were stored at 4°C.

LC-MS/MS sample analysis. At the time of analysis, solid-phase extracted urine samples were reconstituted in 25 µL MilliQ water. The resuspended peptides were centrifuged at 16,300 × g for 10 min, and the supernatants were transferred to low-binding ultra-performance liquid chromatography vials (Wheaton #11-0000-100-S) and analyzed on a Fusion Lumos mass spectrometer (ThermoFisher Scientific, San Jose, USA). Peptides were separated by capillary reverse-phase chromatography on a 2 cm × 100 µm inner diameter, packed in-house with ReproSul-Pur C18-AQ 3.0 µm resin from Dr. Maisch GmbH. The Fusion Lumos was equipped with a Dionex Ultimate 3000 LC system and used a two-step linear gradient with 4–25% buffer B (0.1% (v/v) formic acid in acetonitrile) for 50 min followed by 25–50% buffer B for 20 min, where buffer A was 0.1% (v/v) formic acid. Sample analysis was carried out with the Fusion Lumos equipped with a NanoACQUITY UPLC (Waters) and a NanoAcquity System Manager with ThermoFisher IonSpray (Parr, TX) equipped with a nano-electrospray ion source. The mass spectrometer was operated in data-dependent mode with a scan range from 300 to 1700 m/z. Full MS scans were acquired in the Orbitrap mass analyzer with a resolution of 120,000 (FWHM) and m/z scan range of 400–1500. Precursor ions with charge state 2–7 and intensity threshold above 50,000 were selected for fragmentation using electron transfer dissociation (ETD) with a collision energy of 35 eV, activation time of 0.5 ms, and activation Q of 0.25. Precursor ions with charge state 3–7 were also selected for fragmentation using electron transfer dissociation (ETD) supplemented with 25% collision energy (ETD3). Calibrated charge-dependent ETD parameters were enabled. MS2 fragment ions were analyzed in the Orbitrap mass analyzer with a resolution of 15,000 (FWHM) and m/z scan range of 150–2000. Fragmented precursor ions were dynamically excluded from further selection for a period of 30 s. The AGC target was set to 400,000 and 50,000 for full FTMS scans and FTMS2 scans. The maximum injection time was set to “auto” for full FTMS scans and to “dynamic” for FTMS2 scans.

LC-MS/MS data analysis. To identify peptides originating from the human and wheat proteomes, and in some cases, the barley and rye corre- sponding protein sequences were downloaded from UniProt (www.uniprot.org) to generate custom FASTA databases, as detailed below. The raw MS data were searched against these databases using PEAKS software (versions X or X+). Carbohydrates (on N and O), cysteine (on Cys) were fixed with oxidation and deamidation (on Asn and Gln), and were allowed to vary. The mass tolerances were 10 ppm for precursor ions and 0.02 Da for product ions. The allowed for a maximum of 3 variable modifications per peptide. The allowed for the presence of fragment ions that would allow us to assign the deamidation to a particular Gln residue within the peptide sequence. In all of these cases, we were still able to unambiguously identify a particular Gln residue that had undergone deamidation, as assessed by the presence of multiple site-determining fragment ions, and the combined MS/MS- and MS2-level evidence for deamidation. The MS1 spectra were either equally visible, but the experimentally observed mass more closely matched the theoretical non-monoisotopic mass of the native peptide compared to the monoisotopic mass of the deamidated form, then the sequence was also manually reassigned as native. –90% of the peptide sequences assigned as deamidated by PEAKS were corrected to their native sequences on the basis of these two criteria. In Supplementary Dataset 1–7, peptides whose MS2 (product ion) spectra for the presence of fragment ions that would allow us to assign the deamidation to a particular Gln residue within the peptide sequence. In all of these cases, we were unable to unambiguously identify a particular Gln residue that had undergone deamidation, as assessed by the presence of multiple site-determining fragment ions, and the combined MS/MS- and MS2-level evidence for deamidation. The MS1 spectra were not clearly more consistent with either the deamidated or the native peptide sequences, we did not consider these peptides for downstream data analysis. We cannot exclude the possibility that some peptides were indeed deamidated (or a mixture of native and deamidated forms) but produced insufficient quality spectra to unambiguously assign the deamidation site. Alternatively, it is possible that the sequences of these peptides were otherwise misassigned by the PEAKS software. For clarity, all excluded peptides are reported in the PEAKS search output in Supplementary Datasets 1–7 and are denoted “N” in the “Included for Data Analysis” column.

As an additional step, we acknowledge that this approach for validating deamidated in silico assignment. To assign a peptide as deamidated, we required that the isotopic pattern and mass errors of the MS1 spectra to be consistent with the deamidated form, and for the MS2 spectra to be of sufficient quality to allow clear assignment of the deamidation site. Future studies utilizing targeted mass spectrometry approaches should be designed to conclusively determining if deamidated wheat peptides are detectable in urine, and if so, how their concentration compares to the corresponding native forms.
Received: 18 March 2021; Accepted: 20 January 2022; Published online: 16 February 2022

References
1. Erickson, R. H. & Kim, Y. S. Digestion and absorption of dietary protein. *Annu. Rev. Med.* 41, 133–139 (1990).
2. Miner-Williams, W. M., Stevens, B. R. & Moughan, P. J. Are intact peptides absorbed from the healthy gut in the adult human? *Nutr. Res. Rev.* 27, 308–329 (2014).
3. Shan, L. & Khosla, C. Chemistry and biology of gluten proteins. *Immunol. Endocr. Metab. Agents Med. Chem.* 7, 187–193 (2007).
4. Shan, L. et al. Structural basis for gluten intolerance in celiac sprue. *Science* 297, 2275–2279 (2002).
5. Heyman, M., Abed, J., Lebreton, C. & Cerf-Bensussan, N. Intestinal permeability in coeliac disease: insight into mechanisms and relevance to pathogenesis. *Gut* 60, 1355–1364 (2011).
6. Green, P. H. & Cellier, C. Celiac disease. *N. Engl. J. Med.* 357, 1731–1743 (2007).
7. Leffler, D. A., Green, P. H. & Fasano, A. Extraintestinal manifestations of coeliac disease. *Nat. Rev. Gastroenterol. Hepatol.* 12, 561–571 (2015).
8. Plugis, N. M. & Khosla, C. Therapeutic approaches for celiac disease. *Best. Prac. Res. Clin. Gastroenterol.* 29, 503–521 (2015).
9. Molberg, O. et al. Gliadin specific, HLA DQ2-restricted T cells are commonly found in small intestinal biopsies from celiac disease patients, but not from controls. *Scand. J. Immunol.* 46, 103–109 (1997).
10. Sjöström, H. et al. Identification of a gliadin T-cell epitope in coeliac disease: general importance of gliadin deamidation for intestinal T-cell recognition. *Scand. J. Immunol.* 48, 111–115 (1998).
11. van de Wal, Y. et al. Small intestinal T cells of celiac disease patients recognize a natural pepsin fragment of gliadin. *Proc. Natl Acad. Sci. USA* 95, 10050–10054 (1998).
12. Arentz-Hansen, H. et al. The intestinal T cell response to alpha-gliadin in adult celiac disease is focused on a single deamidated gliadin targeted by tissue transglutaminase. *J. Exp. Med.* 191, 603–612 (2000).
13. Iversen, R. et al. Efficient T-cell-B cell collaboration guides autoantibody epitope bias and onset of celiac disease. *Proc. Natl Acad. Sci. USA* 116, 15134–15139 (2019).
14. Høydahl, L. S. et al. Plasma cells are the most abundant gluten peptide MHC-pathogenensis. *Gut* 61, 1355–1364 (2012).
15. de Ritis, G. et al. In vitro (organ culture) studies of the toxicity of specific A-gliadin peptides in celiac disease. *Gastroenterology* 94, 41–49 (1988).
16. Shidrawi, R. G. et al. In vitro toxicity of gluten peptides in coeliac disease assessed by organ culture. *Scand. J. Gastroenterol.* 30, 758–763 (1995).
17. Mauire, L. et al. Association between innate response to gliadin and activation of pathogenic T cells in coeliac disease. *Lancet* 362, 30–37 (2003).
18. Meresse, B. et al. Coordinated induction by IL15 of a TCR-independent NKG2D signaling pathway converts CTL into lymphocyte-activated killer cells in celiac disease. *Immunity* 21, 357–366 (2004).
19. Hye, S. et al. A direct role for NKG2D/MICA interaction in vitrosphere acting during celiac disease. *Immunity* 21, 367–377 (2004).
20. Nanayakkara, M. et al. P31-43, an undigested gliadin peptide, mimics and enhances the innate immune response to viruses and interferes with endocytic trafficking: a role in celiac disease. *Sci. Rep.* 8, 10821 (2018).
21. Gómez Castro, M. F. et al. P31-43 Gliadin peptide forms oligomers and induces NLRP3 inflammasome/caspase-1 dependent mucosal damage in small intestine. *Front. Immunol.* 10, 31 (2019).

Statistical analyses and other software. All statistical tests indicated in the figure legends were performed using GraphPad Prism version 9. Parts of some figures in this manuscript were created using BioRender.com.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (https://www.ebi.ac.uk/pride/) under accession number PXD031048. The processed peptide identifications are available in Supplementary Datasets 1–7. The processed data used to generate figures are provided in the Source Data file. Source data are provided with the paper. Any remaining data are available within the Article or from the authors upon request. Source data are provided with this paper.

Costa, A. F. et al. Gluten immunogenic peptide excretion detects dietary transgressions in treated celiac disease patients. *World J. Gastroenterol.* 25, 1409–1420 (2019).

Ruz-Carnicer, A. et al. Negative predictive value of the repeated absence of gluten immunogenic peptides in the urine of treated celiac patients in predicting mucosal healing: new proposals for follow-up in celiac disease. *Am. J. Clin. Nutr.* 121, 1240–1250 (2020).

Diaz-Amigo, C. & Popping, B. Accuracy of ELISA detection methods for gluten and reference materials: a realistic assessment. *J. Agric. Food Chem.* 61, 5681–5688 (2013).

Ráki, M. et al. Similar responses of intestinal T cells from untreated children and adults with celiac disease to deamidated gluten epitopes. *Gastroenterology* 153, 787–798.e4 (2017).

Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* 537, 347–355 (2016).

Sigdel, T. K. et al. Optimization for peptide sample preparation for urine peptidomics. *Clin. Proteomics* 11, 7 (2014).

Smith, C. R. et al. Deciphering the peptideomic of urine from ovarian cancer patients and healthy controls. *Clin. Proteomics* 11, 23 (2014).

Osman, A. A. et al. A monoclonal antibody that recognizes a potential coeliac-toxic repetitive pentapeptide epitope in gliadins. *Eur. J. Gastroenterol. Hepatol.* 13, 1189–1193 (2001).

Morón, B. et al. Toward the assessment of food toxicity for celiac patients: characterization of monoclonal antibodies to a main immunogenic gluten peptide. *PLoS ONE* 3, e2294 (2008).

Solld, L. M. et al. Update 2020: nomenclature and listing of celiac disease-relevant gluten epitopes recognized by CD4+ T cells. *Immunogenetics* 72, 85–88 (2020).

Shan, L. et al. Identification and analysis of multivalent proteolytically resistant peptides from gluten: implications for celiac sprue. *J. Proteome Res.* 4, 1732–1741 (2005).

Garème, B., Boachie, R. T., Tsopmo, A. & Udenigwe, C. C. Occurrence, properties and biological significance of pyrogglutamyl peptides derived from different food sources. *Food Sci. Hum. Wellness* 8, 268–274 (2019).

Sato, K. et al. Occurrence of indigestible pyrogglutamyl peptides in an enzymatic hydrolysate of wheat gluten prepared on an industrial scale. *J. Agric. Food Chem.* 46, 3403–3405 (1998).

Van Coillie, E. et al. Functional comparison of two human monocye chomatographic protein-2 isoforms, role of the amino-terminal pyrogglutamic acid and processing by CD26/dipeptidyl peptidase IV. *Biochemistry* 37, 12672–12680 (1998).

Biesiekierski, J. R. What is gluten? *J. Gastroenterol. Hepatol.* 32, 78–81 (2017).

Di Mee, A. et al. An integrated proteomic and peptidomic assessment of the normal human urinome. *Clin. Chem. Lab. Med.* 55, 237–247 (2017).

Van, J. A. D. et al. Peptidomic analysis of urine from youths with early type 1 diabetes reveals novel bioactivity of urumodulin peptides in vitro. *Mol. Cell. Proteomics* 19, 501–517 (2020).

Arrieta, M. C., Bistriz, L. & Melding, J. B. Alterations in intestinal permeability, *Gut* 55, 1512–1520 (2006).

Kumar, S. et al. Abnormal small intestinal permeability in patients with idiopathic malabsorption in tropics (tropical sprue) does not change even after successful treatment. *Dig. Dis. Sci.* 56, 161–169 (2011).

Lerner, A. & Matthias, T. Changes in intestinal tight junction permeability associated with industrial food additives explain the rising incidence of autoimmune disease. *Autoimmun. Rev.* 14, 479–492 (2015).

Lerner, A. & Matthias, T. Processed food additive microbial transglutaminase and its cross-linked gliadin complexes are potential public health concerns in celiac disease. *Int. J. Mol. Sci.* 21, 1127 (2020).

Silvester, J. A. et al. Most patients with celiac disease on gluten-free diets consume measurable amounts of gluten. *Gastroenterology* 158, 1497–1499.e1 (2020).
50. Rubio-Tapia, A. et al. Mucosal recovery and mortality in adults with celiac disease after treatment with a gluten-free diet. *Am. J. Gastroenterol.* **105**, 1412–1420 (2010).

51. Abadie, V. & Jabri, B. IL-15: a central regulator of celiac disease immunopathology. *Immunol. Rev.* **260**, 221–234 (2014).

52. Abadie, V. et al. IL-15, gluten and HLA-DQ8 drive tissue destruction in coeliac disease. *Nature* **578**, 600–604 (2020).

53. Walther, B. et al. GutSelf: interindividual variability in the processing of dietary compounds by the human gastrointestinal tract. *Mol. Nutr. Food Res.* **63**, e1900677 (2019).

54. Parzanese, I. et al. Celiac disease: from pathophysiology to treatment. *Nature Communications* **13**, 888 (2022).

55. Vitorino, R. Digging deep into peptidomics applied to body fluids. *Proteomics* **18**, 1700401 (2018).

56. Rubio-Tapia, A. et al. ACG clinical guidelines: diagnosis and management of celiac disease. *Am. J. Gastroenterol.* **108**, 656–676 (2013).

57. Watson, C. T. & Breden, F. The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. *Genes Immun.* **13**, 363–373 (2012).

58. Nepomuceno, A. I., Gibson, R. J., Randall, S. M. & Muddiman, D. C. Accurate identification of deamidated peptides in global proteomics using a quadrupole orbitrap mass spectrometer. *J. Proteome Res.* **13**, 777–785 (2014).

59. Zhang, J. et al. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell. Proteomics* **11**, M111.010587 (2012).

60. Lex, A., Gahlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. UpSet: visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992 (2014).

**Acknowledgements**

We thank Elvi Sanjines and Gotzone Garay for support with patient recruitment, and Dr. Carlos Gonzales, Dr. Niclas Olsson, and Shelley Dutt for assistance with mass spectrometry. We also thank Dr. Emma Chory for helpful discussions on data processing and visualization. C.K. and B.J. received support for this research from the National Institutes of Health under award number NIH R01 DK063158. C.K. gratefully acknowledges seed funding from The Joint Institute for Metrology in Biology, founded by the National Institute of Standards and Technology. J.E.E. is supported by the Chan Zuckerberg Biohub. N.F.B. gratefully acknowledges Iolee and Robert Triebisch and the Division of Gastroenterology and Hepatology for supporting our celiac translational research program. C.S. was supported by a grant from Ministerio de Ciencia e Innovación from Spain and FEDER funds (SAF2017-83700-R).