

# SCIENTIFIC REPORTS



OPEN

## Mining gut microbiome oligopeptides by functional metaproteome display

Jonas Zantow<sup>1</sup>, Sarah Just<sup>2</sup>, Ilias Lagkouvardos<sup>2</sup>, Sigrid Kisling<sup>2</sup>, Stefan Dübel<sup>1</sup>, Patricia Lepage<sup>3</sup>, Thomas Clavel<sup>2,\*</sup> & Michael Hust<sup>1,\*</sup>

Received: 31 May 2016

Accepted: 12 September 2016

Published: 05 October 2016

Pathogen infections, autoimmune diseases, and chronic inflammatory disorders are associated with systemic antibody responses from the host immune system. Disease-specific antibodies can be important serum biomarkers, but the identification of antigens associated with specific immune reactions is challenging, in particular if complex communities of microorganisms are involved in the disease progression. Despite promising new diagnostic opportunities, the discovery of these serological markers becomes more difficult with increasing complexity of microbial communities. In the present work, we used a metagenomic M13 phage display approach to select immunogenic oligopeptides from the gut microbiome of transgenic mice suffering from chronic ileitis. We constructed three individual metaproteome phage display libraries with a library size of approximately  $10^7$  clones each. Using serum antibodies, we selected and validated three oligopeptides that induced specific antibody responses in the mouse model. This proof-of-concept study provides the first successful application of functional metaproteome display for the study of protein-protein interactions and the discovery of potential disease biomarkers.

Biomarkers are indispensable tools for diagnostics, which can be measured in different body fluids and, thus, are important for the disease management in clinics. Due to the ease of collection, blood serum is the most convenient source for biomarker measurements. Moreover, it circulates through all body regions and tissues and carries immunoglobulins which can be indicative of specific diseases. Countless diagnostic ELISAs and other assays rely on the detection of one specific immunoglobulin. Technological advances such as protein microarrays, peptide microarrays, or Luminex multiplex assays also allow fast and cost-efficient measurement of complex patterns of antibody responses<sup>1–4</sup>. However, the identification of the antigens that trigger these specific systemic antibody responses is still difficult and limits the application of these technologies.

Classical proteomic studies to identify pathogen proteins for diagnostic applications rely on the cultivation of the pathogen followed by 2D-PAGE and immunoblot of the proteome using patient sera and protein identification by mass spectrometry (MS)<sup>5–7</sup>. However, proteome analysis after *in vitro* cultivation does not allow the identification of proteins only expressed during host-pathogen interactions. To identify immunogenic proteins that are specifically important in pathogenesis, the pathogen has to be cultivated in direct contact with the host to ensure gene expression patterns reflecting the true pathogenicity situation. However, an inherent problem is that the pathogen proteome will most likely be overwhelmed by that of the host<sup>8</sup>. Additionally, two other groups of potential biomarkers also usually fail to be identified by classical proteome analysis: Proteins with a molecular mass smaller than 10 kDa or weakly expressed proteins may not be identified by 2D-PAGE/MS<sup>9–11</sup>.

As an alternative, microarrays with several hundred spotted proteins have been used to identify immunogenic proteins from pathogens<sup>12–14</sup>, *E. coli*-derived biomarkers for inflammatory bowel disease<sup>15</sup>, and exposure related antibody responses to proteins from *Plasmodium falciparum*<sup>16</sup>. Even though microarrays allow the screening of several hundreds of target proteins at a time, the technology is laborious and expensive as each protein must be produced in a recombinant manner. Peptide microarrays can be produced with potentially millions of individual sequences<sup>4</sup> allowing to cover entire proteomes, but they are still limited to only detect antibodies binding to short linear epitopes.

<sup>1</sup>Technische Universität Braunschweig, Institute of Biochemistry, Biotechnology and Bioinformatics – Department for Biotechnology, Germany. <sup>2</sup>Technische Universität München, ZIEL Institute for Food and Health, Freising, Germany. <sup>3</sup>Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to T.C. (email: thomas.clavel@tum.de) or M.H. (email: m.hust@tu-bs.de)

Serologic expression cloning (SEREX), a technology developed in the 1990s, relies on the expression of cDNA libraries in *E. coli* using lytic phage vectors<sup>17</sup>. Plaques containing the recombinant proteins are transferred to membranes and stained with patient sera. Immunogenic proteins are identified by DNA sequencing of corresponding clones. The SEREX technology was initially developed for serologic analysis of tumor cDNA expression libraries<sup>18</sup>, but was also used to screen expression libraries from metagenomic DNA<sup>19</sup>. However, handling of lytic phage, maintenance of the phage library, and laborious screening procedures are major drawbacks of this technology.

Instead of using phage for cytoplasmic expression libraries, heterologous proteins can be displayed on the surface of lytic phage (e.g. T7 phage<sup>20</sup>) or non-lytic phage particles<sup>21</sup>. In a subsequent panning procedure using a protein interaction partner immobilized on a solid surface, the phage particles displaying proteins which are specifically bound by the immobilized interaction partner are enriched at the solid surface.

Phage display using non-lytic filamentous M13 phage is enabled by fusing the heterologous protein to a phage coat protein, mostly minor coat protein III (pIII) and is mostly used for the selection of antibodies today<sup>22–24</sup>. However, the ability of peptide phage libraries to be enriched for epitopes of monoclonal antibodies has long been demonstrated<sup>25</sup>. As phage particles contain the genetic information of the displayed proteins, genotype and phenotype are linked. Hence, the technology facilitates not only enrichment but also the simple identification of protein interaction partners by DNA sequencing, e.g. immunogenic proteins which induced systemic antibody responses. Phage display can be used to display oligopeptides and full protein domains derived from a natural source such as genomic DNA or cDNA. However, when randomly fragmented DNA is cloned in a non-directional manner, only one out of eighteen clones result in open reading frames (ORF) and the enrichment of ORFs is a crucial step to improve the library quality. ORF enrichment in phage display libraries was accomplished by several different strategies, such as cloning the DNA fragments upstream of a selection marker like  $\beta$ -lactamase. Only fragments cloned in-frame with the  $\beta$ -lactamase gene and do not hamper correct folding allow the expression of functional  $\beta$ -lactamase and confer ampicillin resistance to the corresponding *E. coli* clone cultivated on selective media. Enriched ORFs are subsequently subcloned into a phage display vector<sup>26,27</sup> or the  $\beta$ -lactamase gene is removed by CRE-mediated recombination<sup>28</sup>. In a T7 phage display approach to identify Calpain substrates, fragmented DNA was cloned upstream of a biotinylation-tag. This way, only ORF fragments could result in tagged and biotinylated gene products, characterizing a procedure similar to using antibiotics for selection<sup>29</sup>. Furthermore, immobilizing produced phage on a surface using streptavidin allowed an ORF selection and the identification of protease substrates by Calpain-mediated release from the surface.

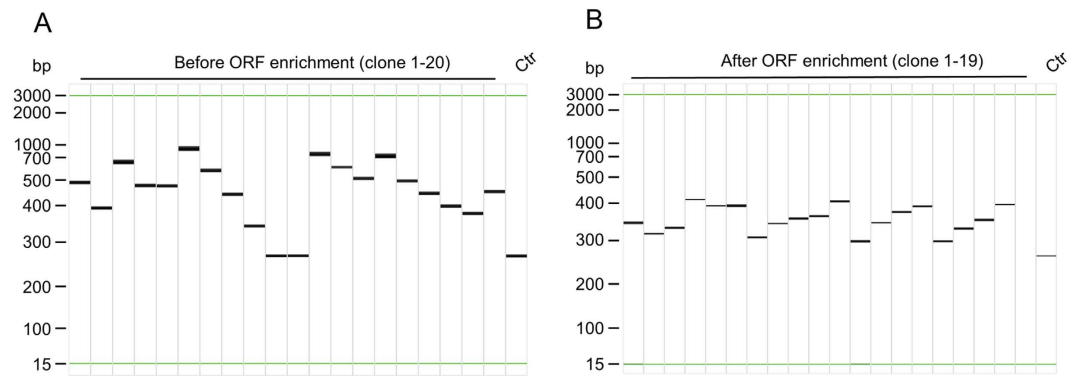
Infectivity of the lytic M13 phage is mediated by pIII whereas each phage particle has up to five pIII copies. The N-terminal domain can be divided in two subdomains (N1 and N2) that mediate infection by binding to the host cell's F pilus, whereas the C-terminal domain (CT) anchors the protein in the phage capsid. These three regions are separated by glycine and serine rich linkers. Introducing trypsin sites into the linker regions renders pIII sensitive to proteolytic cleavage and decreases infectivity after trypsin treatment. Cloning randomly fragmented DNA upstream of a trypsin resistant pIII gene in a phagemid vector and using a trypsin sensitive helper-phage for phage packaging allows ORF enrichment of genomic libraries after trypsin treatment and infection of *E. coli* host cells<sup>30</sup>.

Using another special helper phage which has pIII in its capsid and is infective but has a genomic deletion of the pIII encoding gene *gIII* ("Hyperphage"<sup>31,32</sup>), ORFs can be enriched without further cloning steps or protease treatment (ORFeome phage display<sup>33,34</sup>). Therefore, randomly fragmented DNA is cloned upstream of the pIII gene on a phagemid vector. After "Hyperphage" co-infection, infective phage particles are only produced if the inserted DNA sequence is in-frame with the pIII gene and does not contain any stop codon as the fusion protein is the only pIII source to render the phage particles infective. Consequently, ORFs are strongly enriched during "Hyperphage" packaging.

The ORFeome phage display technology was recently used to identify novel potential biomarker or vaccine candidates from genomic libraries of different *Mycoplasma* species<sup>35,36</sup>, *Salmonella* Typhimurium<sup>37</sup>, *Neisseria gonorrhoeae*<sup>38</sup>, and from a tick salivary gland cDNA library<sup>39</sup>. These studies used functional proteomic display of oligopeptides from single species, but since the ORFeome phage display technology is independent of cultivation, it is in principle usable for the study of any ORF containing DNA mixture. Complex microbial communities such as the intestinal microbiota could be analyzed by functional screening of metaproteome phage display libraries when using metagenomic DNA preparations.

Crohn's disease (CD) and ulcerative colitis (UC) are idiopathic relapsing chronic inflammatory disorders of the gut commonly referred to as inflammatory bowel disease (IBD). Genetic predisposition markers comprise mutations in genes associated with microbial sensing<sup>40–42</sup>. Moreover, shifts in microbiota composition are associated with IBD<sup>43,44</sup>, highlighting the crucial role of the microbiome in IBD pathogenesis. Hence, IBD is closely correlated to aberrant immune responses towards the gut microbiota in a genetically susceptible host. Humoral immune responses in IBD patients include autoantibodies like anti-neutrophil cytoplasmic antibodies (ANCA) and increased immunoglobulin levels against microbial antigens such as cell-wall components of yeast (anti-*Saccharomyces cerevisiae* antibodies, ASCA), bacterial flagellin (anti-Cbir1), *Escherichia coli* outer membrane porin C (anti-ompC), and *Pseudomonas fluorescens* protein I-2 (anti-I2). For IBD diagnostics, ASCA have the best combined sensitivity (31–45%) and specificity (90–100%) (reviewed in ref. 45). However, due to the low prevalence of already known IBD-associated antibody responses, there is an urgent need for additional serologic markers that may support IBD diagnostics and give further insights into host-microbiome interactions.

In order to study molecular mechanisms underlying IBD, numerous mouse models have been established (for review see ref. 46). Transgenic TNF <sup>$\Delta$ ARE/+</sup> mice are characterized by elevated TNF levels and spontaneous development of CD-like ileitis<sup>47</sup>. Ileal inflammation in this model was recently shown to be triggered by the gut microbiota, as germ-free mice did not develop inflammation and mice deficient in MyD88 (an adaptor protein essential for innate recognition of gut microbes) showed attenuated signs of inflammation<sup>48,49</sup>. Moreover, shifts in the



**Figure 1. Representative colony PCR of random clones of a metaproteome library before and after ORF enrichment.** (A) Lane 1–20: randomly analyzed clones before ORF enrichment. (B) Lane 1–19: randomly analyzed clones after ORF enrichment. Ctr: empty pHORF3 vector backbone. Colony PCR was analyzed by capillary electrophoresis and results are displayed as virtual gel image. Insert rates and mean insert sizes were estimated by colony PCR of 20 clones per transformation ( $n = 4$ ) of each library and after Hyperphage packaging of each library.

microbiota associated with inflammation were able to induce ileal inflammation when transferred to germ-free animals<sup>48</sup>. Hence, involvement of both TNF and the gut microbiota are shared features between TNF<sup>ΔARE/+</sup> mice and human CD, suggesting mechanistic similarities in the pathophysiology, making this mouse model well-suited to study host-microbiome interactions.

Mining the gut microbiome for immunogenic bacterial proteins capable of inducing a systemic antibody response in the host may reveal novel biomarker candidates for IBD diagnostics. In the present work, we performed a proof-of-concept study on the development and use of functional metaproteome display for biomarker discovery.

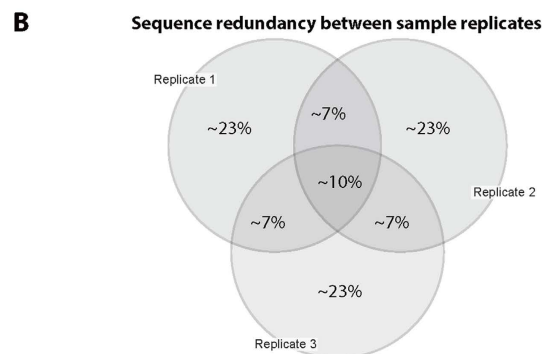
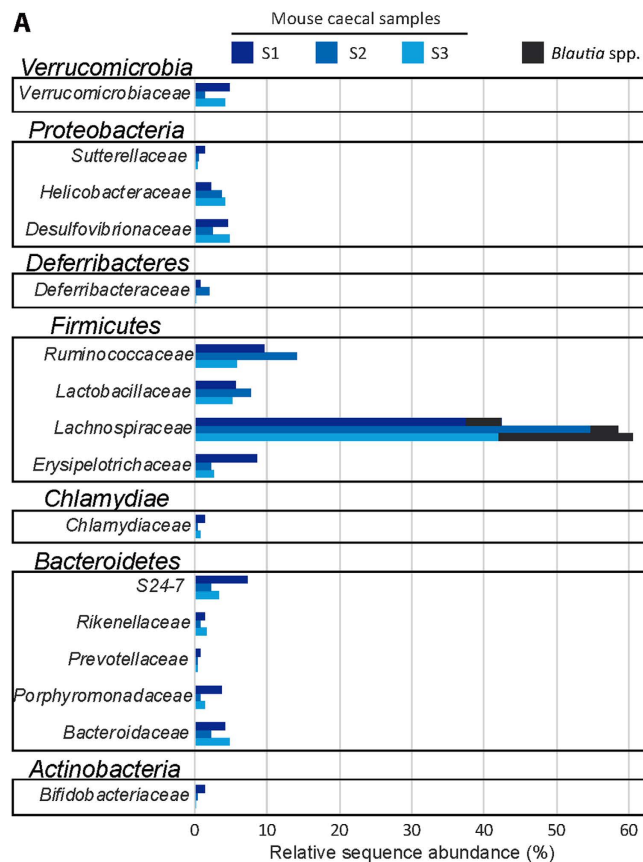
## Results

**Construction of metaproteome phage display libraries.** Three individual metaproteome phage display libraries were constructed from metagenomic DNA isolated from the caecum of three 8-week-old TNF<sup>ΔARE/+</sup> mice (illustrated workflow in Figure S1A). Based on transformation rates, the initial library diversities were  $6 \times 10^7$ ,  $2 \times 10^7$  and  $3 \times 10^7$  independent clones. Insert rates were above 85% for all three libraries with mean insert sizes of 284 to 354 bp as analyzed by colony PCR (Fig. 1A). Library quality was further enhanced by the enrichment of ORFs and the corresponding oligopeptides were displayed on phage particles (illustrated in Figure S1B). ORF enrichment also removed empty vector backbones (100% insert rate) and mean insert sizes were decreased to 92 to 117 bp after ORF enrichment (Fig. 1B).

**Gut microbiota sequencing.** To assess gut bacterial diversity and composition in the three mice used to construct the phage libraries, DNA isolated from caecal contents was analyzed by high-throughput 16S rRNA amplicon analysis. The operational taxonomic unit (OTU) table summarizing sequence counts of each molecular species per sample together with the corresponding OTU sequences are provided in File S1 and S2, respectively. Samples were characterized by low alpha-diversity, i.e. an average species richness of 56 and a Shannon effective species count<sup>50</sup> of 24. The taxonomic composition per sample (Fig. 2A) was in line with data from the literature. The dataset was characterized by marked inter-individual differences (e.g. OTU-1 was abundant in sample S2 and S3, but not in S1). However, more than 50% of the reads were classified as member of the phylum *Firmicutes* in all samples, with relative abundance of OTUs corresponding to members of the genus *Blautia* (see candidate peptides below) ranging between 4 and 18%.

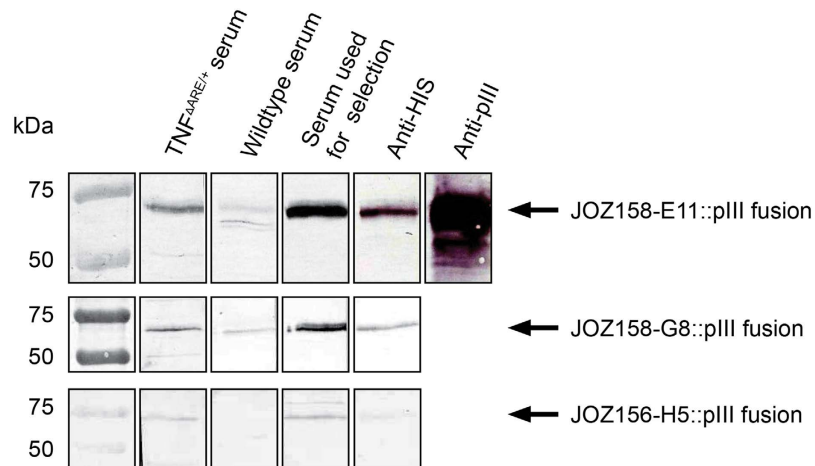
**Phage oligonucleotide diversity and size distribution.** In order to obtain an overview of the starting oligonucleotide diversity prior to panning, triplicate phage libraries of each of the three donor mice were sequenced. An average of 193,500 total reads (about 176,000 after merging and excluding reads with more than 0.1% expected error rate) and 122,000 unique reads were obtained per mouse per replicate. The mean sequence length of each library was  $84.8 \pm 1.9$  nt,  $65.8 \pm 4.3$  nt and  $71.7 \pm 5.7$  nt with maximum insert lengths of 499 nt, 448 nt and 492 nt, respectively (size limit due to sequencing specifications was approximately 500 nucleotides).

Approximately 30% of the reads (54,000 out of 176,000, represented on average by 30,000 unique peptides) corresponded to redundant peptides (i.e., those with more than one copy; median copy number was 2, showing that most of the redundant peptides had low copy numbers; maximum copy number ranged between 44 and 163, depending on the library). The 20 most abundant clones represented less than 0.5% of total reads. Across the reads obtained from the triplicate library preparations of the same sample, only 10% of the sequences were shared between all three replicates and another 7% between each of the replicate pairs (Fig. 2B). These findings suggest that the total diversity of oligonucleotides existing in the libraries was not covered by either the individual library preparation or more likely the number of sequencing reads, and exceeded the 265,000 detected sequences per library. Most of all, these data show that the constructed library displayed a high degree of diversity as unique peptides dominated the libraries, demonstrating their suitability as starting materials for metaproteomic biomarker candidate selection.



**Figure 2. Gut microbiota and phage oligonucleotide sequencing.** (A) Taxonomic composition of the three donor mice at the bacterial family level as per 16S rRNA gene analysis. Reads were classified using SILVA<sup>70</sup> and families were ranked by phyla. Relative abundances refer to the proportion of sequences belonging to the given taxa considering a mean number of sequences per sample of approximately 17,000. (B) Venn's diagram showing the percentage of oligonucleotides that are shared between or unique to the replicate libraries for any given mouse caecal sample. The percentages refer to a total number of approximately 265,000 unique oligonucleotide reads.

Detailed NGS analysis of one replicate of ORF enriched library 1 (Table S1) revealed a size distribution ranging from 8 to 476 nt. Random fragmentation of DNA for library construction results in DNA fragments with sizes of  $3n$ ,  $3n + 1$  and  $3n + 2$  nucleotides. Sequences with a length of  $3n + 1$  were enriched over sequences with  $3n + 2$  and  $3n$  nucleotides (70.3% vs. 27.3% vs. 2.4% of unique reads) (Figure S2). Due to the vector design (Figure S3), only inserts with  $3n + 1$  nucleotides constitute ORFs without a frameshift in the pIII gene. Only 3.4% of inserts with  $3n + 1$  nt contained stop codons (vs. 2.7% and 43.0% for inserts with  $3n + 2$  and  $3n$  nt, respectively). Notably, out-of-frame inserts with  $3n + 2$  nucleotides were less efficiently removed during ORF enrichment than inserts with  $3n$  nucleotides and contained significantly less stop codons. A fraction of 17.4% of inserts with  $3n + 1$  nt and one or more stop codons could potentially be rescued as they also contained a second start codon in the correct frame to translate into a pIII fusion protein. Moreover, alternative start codons for the potential translation of a pIII fusion protein were detected in about 30% (30.0% and 28.4% of inserts with  $3n + 2$  and  $3n$  nt, respectively) of the out-of-frame inserts. Taking alternative start codons into account, a total of 77.2% of analyzed sequences had



**Figure 3. Immunoblot of oligopeptide phage particles stained with serum.** Oligopeptide:pIII fusion protein was stained with serum pools (1:200 dilution) from  $TNF^{\Delta ARE/+}$  and wildtype mice. Bound antibodies were detected with an anti-mouse (IgA, IgG, IgM) secondary antibody. Control staining of mouse anti-hexahistidine and exemplary mouse anti-pIII (JOZ158-E11) staining was included to identify the pIII fusion protein.

a potential gene product comprising an oligopeptide:pIII fusion protein. Hence, these sequencing data demonstrate the efficiency of the ORF enrichment step.

**Selection of immunogenic oligopeptides from metaproteome phage display libraries.** To select microbiota-derived immunogenic oligopeptides from the metaproteome phage display libraries, three independent pannings using immobilized serum antibodies (IgA, IgG and IgM in combination) were performed. The complete workflow is illustrated in Figure S1C. Paired samples (i.e. one individual library constructed from a given mouse gut content and the corresponding serum sample) were used. After three consecutive rounds of panning (output of selection rounds is given in Figure S4), 92 clones from each panning were screened by monoclonal phage ELISA aiming to identify specific antibody response of the corresponding sera towards the displayed oligopeptides. In total, 16 individual clones reactive with the designated mouse serum were selected (Table 1). Some selected clones were found in multiple copies among the 276 screened clones. In particular, clone JOZ156-A6 was dominantly enriched during the panning procedure, but was later found to be a universal immunoglobulin binding protein (data not shown). The selected fragments ranged from 31 to 328 bp in size. Protein homologies (blastp) were only found for clones JOZ157-E1 and JOZ158-E11. Clones JOZ156-A6, JOZ156-A8, JOZ156-F7, and JOZ157-C10, were annotated as predicted proteins from *Blautia* species.

**Validation of immunogenicity by immunoblot.** Blotted proteins of the selected oligopeptide phage particles were incubated with pools of each 15  $TNF^{\Delta ARE/+}$  and 19 wt C57BL/6N mouse sera at the age of 18 weeks. All  $TNF^{\Delta ARE/+}$  mice were inflamed in the distal ileum and proximal colon, as confirmed by histological scoring ( $3.9 \pm 1.3$  vs.  $0.03 \pm 0.08$  (distal ileum) and  $4.3 \pm 1.7$  vs.  $0.05 \pm 0.13$  (proximal colon)). Detection of bound serum antibodies resulted in a stronger staining of the oligopeptide::pIII fusion protein for three of the 16 candidate oligopeptide phage clones (Fig. 3). These peptides (JOZ156-H5, JOZ158-E11, and JOZ158-G8) were considered to be the most promising candidates for detection of specific circulating antibodies in  $TNF^{\Delta ARE/+}$  mice and were synthesized as biotinylated peptides.

**Validation of  $TNF^{\Delta ARE/+}$  specific immune responses by ELISA.** Even though immunogenicity was validated using oligopeptide phage and serum pools, further characterization was performed using synthetic peptides and individual sera of the  $TNF^{\Delta ARE/+}$  and control cohorts. Peptides were immobilized via streptavidin and incubated with the individual sera. Bound anti-peptide serum antibodies were detected by ELISA using a secondary anti-mouse antibody. Some sera of mostly the  $TNF^{\Delta ARE/+}$  cohort reacted with streptavidin alone and competition with streptavidin in solution did not decrease the background binding of streptavidin. Therefore, a background subtraction step was implemented to exclude false positive reactive sera. The measured antibody responses against all three peptides were significantly higher in the  $TNF^{\Delta ARE/+}$  than wildtype mice, despite marked inter-individual differences (Fig. 4). Compared to peptide JOZ156-H5, antibody responses against the peptides JOZ158-E11 and JOZ158-G8 were generally stronger and reactivity of some control sera was increased. Nevertheless, the serum reactivity of the  $TNF^{\Delta ARE/+}$  cohort against both peptides (JOZ158-E11 and JOZ158-G8) was significantly increased compared to the control sera. Consequently, all three isolated peptides can be considered as biomarker candidates for the  $TNF^{\Delta ARE/+}$  mouse model.

The sequence JOZ158-E11 derived from a protein of *Blautia* sp. with predicted function of a phosphoribosyl aminoimidazole succinocarbox amide synthase (SAICAR synthase) based on the identification of the conserved domain and the similarity to the already annotated protein (Table S2). The identification of the closest homolog was performed with a blastp search over the nr protein database in NCBI.

Clone	Origin mouse	Insertsize (bp)	Hit redundancy	Protein homology (blastp)
JOZ156-A3	1	43	1	—
JOZ156-A6	1	298	32	Predicted protein ( <i>Blautia</i> sp.) (E-value 7e-24, 66 % identity)
JOZ156-A8	1	319	1	Predicted protein ( <i>Blautia</i> sp.) (E-value 3e-27, 57 % identity)
JOZ156-B11	1	58	6	—
JOZ156-F7	1	328	1	Predicted protein ( <i>Blautia</i> sp.) (E-value 3e-29, 58 % identity)
JOZ156-G7	1	85	2	—
JOZ156-H5	1	58	1	—
JOZ157-C10	2	211	3	Predicted protein ( <i>Blautia</i> sp.) (E-value 7e-18, 72 % identity)
JOZ157-E1	2	202	1	ABC transporter permease ( <i>Eubacterium</i> sp. ER2) (E-value 2e-21, 62 % identity)
JOZ157-G6	2	58	1	—
JOZ158-C11	3	76	1	—
JOZ158-C12	3	46	1	—
JOZ158-D10	3	31	2	—
JOZ158-E11	3	121	2	Phosphoribosylaminoimidazol-succinocarboxamide synthase ( <i>Blautia</i> sp.) (E-value 4e-20, 100 % identity)
JOZ158-G8	3	163	1	—
JOZ158-H8	3	52	1	—

**Table 1. Summary of selected clones.**

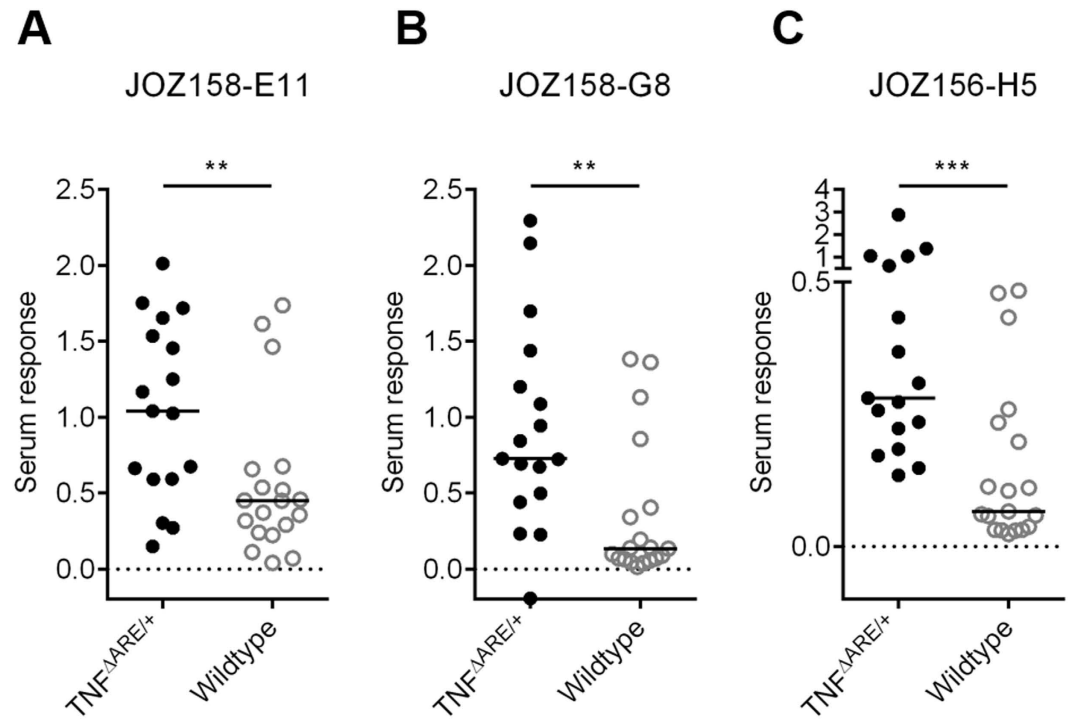
**Time-dependent occurrence of the biomarker candidates.**  $\text{TNF}^{\Delta\text{ARE}/+}$  mice are known to develop intestinal inflammation gradually from the age of 4–6 weeks on with plateauing around 12 weeks of age<sup>47,51</sup>. To test whether the systemic antibody response towards the three identified gut microbiome-derived peptides correlated with age and inflammatory state, serum reactivity was monitored using ten mice of each genotype sacrificed at week 4, 6, 8, 12, and 18 of age. Intestinal inflammation was assessed by histopathological scoring and anti-JOZ156-H5, anti-JOZ158-E11, and anti-JOZ158-G8 serum antibody titers were measured by ELISA (Fig. 5). Intestinal inflammation was observed by week 6 and the inflammatory state increased with older age of the  $\text{TNF}^{\Delta\text{ARE}/+}$  mice, whereas no inflammation was observed in wildtype controls (Fig. 5A). Inter-individual differences in serum response were observed for all three analyzed oligopeptides. Levels of anti-JOZ158-E11 antibodies were relatively high in 8-week-old  $\text{TNF}^{\Delta\text{ARE}/+}$  mice, but were not further increased thereafter and were not abundant in the serum of 18-week-old mice (Fig. 5B). Anti-JOZ158-G8 antibodies were only present in a fraction of  $\text{TNF}^{\Delta\text{ARE}/+}$  samples with only a trend towards increased response from week 8 on ( $p = 0.079$  and  $0.078$  for 12 weeks and 18 weeks, respectively) (Fig. 5C). Serum levels of anti-JOZ156-H5 antibodies were increased in 4-, 6-, 8-, 12- and 18-week-old  $\text{TNF}^{\Delta\text{ARE}/+}$  mice when compared to the matched wildtype controls (Fig. 5D).

## Discussion

The present work identified three novel immunogenic oligopeptides from gut microbiota-derived metagenomes in a mouse model of Crohn's disease and thus demonstrates that ORFeome phage display is suitable to analyze metaproteomes. To date, ORFeome phage display technology was only used to identify immunogenic proteins from single organisms<sup>35–37,39</sup>. Ciric *et al.* 2014 used a similar technology for the enrichment for DNA inserts that contained an endogenous signal peptide and analyzed this secretome enriched library by next generation sequencing, but this approach did not include any functional screening<sup>52,53</sup>. Here, we present an approach that allows to identify immunogenic proteins from ORF-enriched phage display libraries based on specific protein interactions including functional selection from proteins of a metaproteome independent of cultivation, expression level or antigen size.

Compared to phage display libraries covering genomes of one species, metagenomic approaches require larger libraries. Estimating the necessary library diversity to cover the whole metagenome remains difficult as the microbiome composition and relative abundance of the different species is not completely known and is highly individual-specific. The libraries constructed in the present work contained more than  $10^7$  individual clones. In a previous study, serological expression cloning (SEREX) of randomly fragmented metagenomic DNA from the caecum of C3H/HeJ Bir mice and screening of  $6 \times 10^5$  pfu of the lambda expression libraries was sufficient to identify bacterial flagellin as immunogenic protein<sup>19</sup>. Moreover, SEREX relies on non-directional cloning of randomly fragmented DNA, so only a small fraction (typically less than 6%) of inserts usually results in open reading frames. Employing non-lytic M13 phage for the ORFeome phage display technology as in the present study not only allows the screening by monoclonal ELISA but also an ORF enrichment prior to selection. Hence, ORFeome phage display libraries contain a much lower number of junk fragments which can hamper functional selection by their stickiness. Although our libraries contained more than  $10^7$  clones and up to 20 Gbp, they presumably still not covered the entire metagenome. The human gut microbiome harbors several hundred different bacterial species<sup>54</sup> with more than 10 million genes<sup>55,56</sup>. Genes from species with low abundance may be underrepresented. Additionally, the metagenomic DNA used for library construction may have also contained DNA of other microorganisms such as fungi and protozoa as well as viruses, host, and food-derived DNA.

Nevertheless, the constructed phage libraries had a substantial degree of diversity that allowed their use as starting materials for peptide selection. Moreover, the ability to perform culture-independent proteomic studies

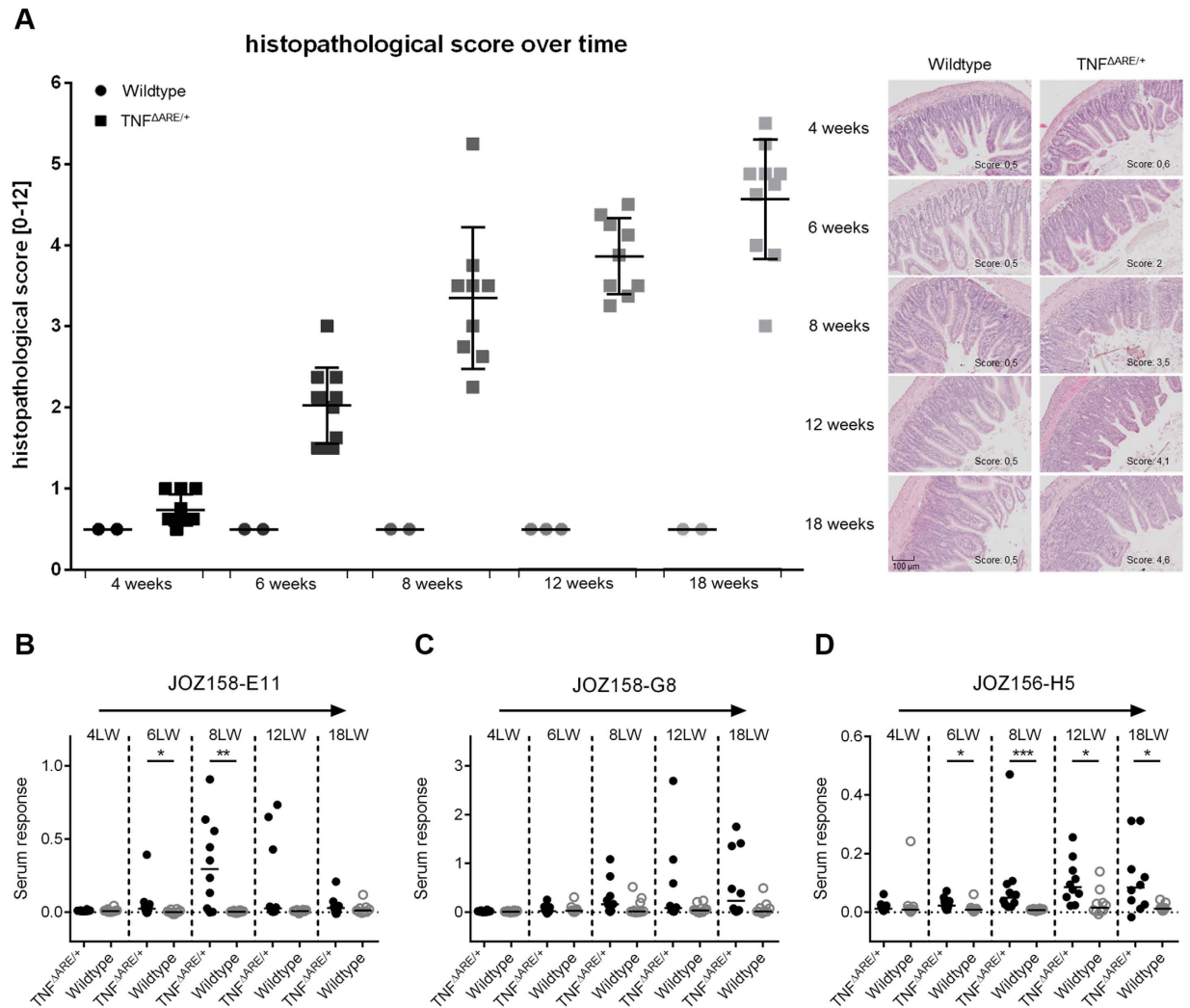


**Figure 4. Specific serum reactivity against selected peptides in ELISA.** Reactivity of TNF $\Delta$ ARE<sup>+/+</sup> and wildtype sera with (A) JOZ158-E11 oligopeptide (B) JOZ158-G8 oligopeptide (C) JOZ156-H5 oligopeptide. The median cohort reactivity is indicated by a black line. Selected peptides were immobilized by streptavidin and bound serum antibodies from single sera (1:100 dilution) were detected with an anti-mouse (IgA, IgG, IgM) antibody. (Non-parametric Mann-Whitney test; \*\* =  $p < 0.01$ ; \*\*\* =  $p < 0.001$ ).

is one of the superior characteristics of the technology, clearly demonstrated as only 56% of 16S rRNA reads detected by sequencing have been related to cultured bacteria at the genus level<sup>57</sup>.

In the present work, the ORF-enriched libraries were analyzed for the first time by NGS analysis and revealed that few clones were overrepresented in the libraries but the majority of sequences were equally distributed. In another NGS study analyzing a commercial peptide phage library (Ph.D.<sup>TM</sup>-12) after amplification, the authors acquired  $2 \times 10^7$  reads to analyze a library of  $10^6$  unique sequences<sup>58</sup>. These are about 50-times more reads compared to the present study suggesting that we did not cover the full library diversity especially considering that only 10% of sequences were shared between NGS analysis of replicate library preparations. Nevertheless, we were able to gain insights into the size distribution and relative abundance of individual clones in the libraries. In the previous study with the commercial Ph.D.<sup>TM</sup>-12 peptide phage library, the authors reported a strong bias towards 150 clones that dominated 20% of the library (with 20 clones accounting for 8%)<sup>58</sup>. We expected an even stronger bias in the present study due to the different insert sizes ranging from 8 to 476 bp having different impact on cell growth. However, the 20 most abundant clones accounted for less than 0.5% of total reads, indicating the lack of a respective bias in our libraries. The size distribution of nucleotide inserts revealed a strong enrichment (70.3% of total sequences) of clones with inserts of  $3n + 1$  nucleotides and only in a minor fraction (3.4% of sequences with  $3n + 1$  nt) remained stop codons. The remaining 29.7% of the packaged ORFeome phage display library had  $3n$  and  $3n + 2$  nt inserts, which were out-of-frame with the pIII gene and in theory they should not result in infective phage particles. However, about 30% of each of these two types of out-of-frame inserts contained alternative start codons as potential translational start points for a pIII fusion protein which renders the phage clone infective. This phenomenon was described in a previous study when oligopeptide phage clones with out-of-frame inserts and alternative start codons coding for an immunogenic protein were selected and validated<sup>38</sup>. From insert size distribution in combination with detailed stop codon analysis of the obtained NGS data we concluded that in total more than 77% of the library contained potential ORFs demonstrating the efficacy of the ORF enrichment using the “Hyperphage” technology. The degree of enrichment is in line with previously reported ORF enrichment using the ORFeome phage display technology<sup>34,38</sup>. Further, functionally displayed oligopeptides derived from inserts with frameshifts were described before<sup>34,59,60</sup> and were proposed to result from RNA secondary structures or a selection pressure against oligopeptides which may be toxic for the *E. coli* host. Therefore, an enrichment for 100% ORFs may not be possible at all.

We selected two immunogenic oligopeptides of 20 and 56 amino acids that could not be identified by BLAST analysis. The third selected oligopeptide comprised 40 amino acids of a phosphoribosylaminoimidazole-succinocarboxamide synthase (SAICAR synthase) of *Blautia* sp. and is very similar to SAICAR synthases from other highly abundant members of the mammalian gut (e.g. *Dorea* sp. and *Clostridium* sp.). SAICAR synthase is an enzyme involved in the *de novo* biosynthesis of purine nucleotides<sup>61</sup> and



**Figure 5. Time course of serum reactivity.** Antibody response of mice sacrificed at life week 4, 6, 8, 12 and 18. (A) Histologic scoring of ileal inflammation (only 2 out of 10 wildtype samples were scored); serum reactivity with (B) JOZ158-E11 oligopeptide (C) JOZ158-G8 oligopeptide (D) JOZ156-H5 oligopeptide. The median cohort reactivity is indicated by a black line. Selected peptides were immobilized by streptavidin and bound serum antibodies from single sera (1:100 dilution) were detected with an anti-mouse (IgA, IgG, IgM) antibody. (Non-parametric Mann-Whitney test; \* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ; \*\*\* $p < 0.001$ ).

essential for all living organisms. Due to structural and functional differences between microbial and vertebrate SAICAR synthases, selective inhibitors have been proposed as antimicrobial agents<sup>62</sup> and structural differences may confer immunogenicity to bacterial SAICAR synthase. As intracellular protein, SAICAR synthase may have been exposed to the mouse immune system after bacterial cell lysis. In a previous study using the ORFeome phage display technology, the selection procedure allowed the identification of intracellular proteins from *Neisseria meningitidis* that have been described to be immunogenic before<sup>38</sup>.

Except from SAICAR synthase, protein function and genetic origin of the selected clones remain unknown. The fact that BLAST analysis of all 16 initially selected clones did not return any annotation from the NCBI nt/nr database with high homology on nucleotide level indicates that the gut microbiome may be more diverse than assumed and motivates to look for more so far unknown species in the mammalian gut<sup>63</sup>.

In conclusion, ORFeome phage display technology has demonstrated promising features for the identification of novel biomarkers in IBD and other microbiome-related diseases. Libraries may be constructed from any type of genomic or metagenomic DNA. Thus, ORFeome phage display may also facilitate the use of environmental metagenomes to identify protein interaction partners of microbial communities that may be of further interest for biotechnological or industrial processes.

## Methods

**Mouse experiments.** Animal use was approved by 'Landratsamt Freising' (animal welfare authorization no. 32–568), following the guidelines of the 'Deutsches Tierschutzgesetz' (German Animal Welfare Act) and the 'Deutsche Tierschutz-Versuchstierverordnung' (German Animal Welfare of Experimental Animals Regulation)



under supervision of a veterinarian and an animal welfare officer. Mice were sacrificed for scientific purpose only and were not included in any specific treatment protocols. Heterozygous  $TNF^{\Delta ARE/+}$  and wildtype mice on a C57BL/6N genetic background were housed in the mouse facility at the Life Sciences faculty of the Technische Universität München under conventional conditions with a 12 h light/dark cycle at 24–26 °C. All mice were fed a standard diet (R/M-H, Ssniff, Soest, Germany) *ad libitum* and were sacrificed by CO<sub>2</sub> inhalation at the indicated age (4 to 18 weeks).

**Histological scoring.** Formalin fixed and paraffin embedded tissue sections of the distal ileum and proximal colon from  $TNF^{\Delta ARE/+}$  and WT mice were stained with hematoxylin and eosin (H&E) and scored in a blinded manner by assessing lamina propria mononuclear cell infiltration, crypt hyperplasia, goblet cell depletion and architectural distortion. Scoring resulted in a score from 0 (non-inflamed) to 12 (highly inflamed) per section as previously described<sup>64</sup>.

**Isolation of metagenomic DNA.** Frozen caecal samples were mixed with 600 µl DNA stabilization solution (STRATEC biomedical), 400 µl phenol:chloroform:isoamyl alcohol (25:24:1; Sigma-Aldrich) and 500 mg autoclaved zirconia/silica beads (0.1 mm; BioSpec). Cells were disrupted by mechanical lysis using a FastPrep<sup>®</sup>-24 (3 × 30 sec at maximum speed) (MP Biomedicals). Samples were then heat-treated (95 °C, 5 min) and centrifuged (16,000 × g/5 min/4 °C). Supernatants were treated with RNase (1 µg/µl) for 30 min at 37 °C. Metagenomic DNA was purified using gDNA columns (Macherey-Nagel) following the manufacturer's instructions. Concentrations and purity were measured using NanoDrop<sup>®</sup> (Thermo Scientific) and samples were stored at –20 °C.

**Construction of metaproteome phage display libraries.** Metagenomic DNA (100 ng) was amplified by linear unbiased multiple displacement amplification according to the manufacturer's instructions (illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit, GE Healthcare). The amplified DNA was diluted in 2 mL H<sub>2</sub>O and fragmented by sonication (6 × 2 min, 50% intensity, MS72 sonotrode, HD2200 Sonopuls). Fragmented DNA was analyzed on 1% agarose gels to ensure fragment sizes between 200 and 1500 bp. The fragmented DNA was concentrated using Amicon Ultra Centrifugal Filters (30K) (Millipore). Cohesive ends were blunted and blunt ends were phosphorylated according to manufacturer's instructions (Fast DNA End Repair Kit, Thermo Scientific). DNA was then purified using a spin column according to the manufacturer's instructions (NucleoSpin Gel and PCR Clean-UP Kit, Macherey-Nagel).

The libraries were constructed by blunt-end cloning of 1,400 ng fragmented metagenomic DNA into 1,000 ng PmeI (NEB) linearized and dephosphorylated (Calf Intestine Phosphatase, NEB) pHORF3<sup>35</sup> library vector (16 h at 16 °C, T4 DNA Ligase, Promega). The ligation was inactivated for 10 min at 65 °C and the buffer was exchanged 4 times with 500 µL H<sub>2</sub>O using Amicon Ultra Centrifugal Filters (30K) (Millipore). The ligation was splitted into 4 separate transformations and 25 µL of electrocompetent *E. coli* TOP10F' (TOP10F' Electrocomp<sup>™</sup> Kit, Life Technologies) were transformed by electroporation (1.8 kV, MicroPulser<sup>™</sup>, BioRad). After 1 h of incubation at 37 °C and 650 rpm in 1 mL SOC medium (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.05 (w/v) NaCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub>, 20 mM glucose), transformation rates were determined by plating dilutions on 2 × YT agar (1.6% (w/v) tryptone, 1% (w/v) yeast extract, 0.05% (w/v) NaCl, 1.2% (w/v) agar) supplemented with 100 mM glucose and 100 µg/mL ampicillin (2 × YT-GA). Cells were plated on 2 × YT-GA agar plates (25 × 25 cm) and incubated at 37 °C overnight. The cells were scraped using 20 mL of 2 × YT medium, flash frozen in liquid nitrogen and stored at –80 °C in 20% (v/v) glycerol. Insert rates and mean insert sizes were determined by PCR and capillary gel electrophoresis (Qiaxcel Advance, Qiagen, Germany) of randomly analyzed colonies (n = 20 per transformation).

**Packaging of oligopeptide phage library and ORF enrichment.** Four hundred mL of 2 × YT-GA medium were inoculated with frozen cells (see above) to an OD<sub>600</sub> < 0.1 and cultivated at 37 °C and 250 rpm (Infors HT) until an OD<sub>600</sub> of 0.5 was reached. In order to complement the missing coat proteins and ensure enrichment for ORFs, 25 mL of the culture were infected with a 20-fold excess (2.5 × 10<sup>11</sup> colony forming units (cfu) of the M13K07ΔgIII helperphage “Hyperphage”<sup>31</sup>) for 30 min at 37 °C. The infected cells were incubated at for another 30 min at 37 °C and 250 rpm (Infors HT) to express antibiotics resistance. The cells were pelleted (3,200 × g, 10 min) and subsequently resuspended in 400 mL 2 × YT medium supplemented with 100 µg/mL ampicillin and 50 µg/mL kanamycin (2 × YT-AK).

Oligopeptide phage particles were produced for 24 h at 30 °C and 250 rpm (Infors HT). The cells were pelleted and phage particles were precipitated from the supernatant at 4 °C overnight after adding 1/5 volume precipitation buffer (20% (w/v) PEG 6,000, 2.5 M NaCl). The precipitated phage particles were pelleted for 1 h at 12,000 × g and 4 °C (Sorvall Centrifuge RC5B Plus, Rotor F9S) and resuspended in 10 mL phage dilution buffer (10 mM Tris-HCl pH7.5, 20 mM NaCl, 2 mM EDTA). In order to confer higher purity, the resuspended phage were filtered (Whatman syringe filter, 0.45 µm) and a second precipitation step was performed for 2 h at 4 °C with 1/5 volume precipitation buffer. The phage particles were pelleted for 30 min at 20,000 × g and 4 °C (Sorvall Centrifuge RC5B Plus, Rotor SS34) and resuspended in 1 mL fresh phage dilution buffer. Remaining bacteria were pelleted for 2 min at 16100 × g (Eppendorf Centrifuge 5415 D) and supernatants containing the oligopeptide phage libraries were collected and stored at 4 °C.

**Library panning (selection of candidate immunogenic oligopeptides from complex libraries).** *Immobilization of serum antibodies.* Panning allows the selection of displayed oligopeptides based on specific affinity binding to serum antibodies. In order to reduce background signals, the serum used for panning was pre-cleared from antibodies reactive with phage particles alone. Therefore, two wells of a 96-well ELISA plate (Costar) were coated with 4 × 10<sup>10</sup> cfu Hyperphage in 300 µL phosphate buffer saline (PBS). The protein binding capacity

of the well surface was saturated with 350  $\mu$ L blocking solution (PBS supplemented with 2% skim milk powder (Roth) and 0.1% Tween 20) per well. The blocking solution was removed and the mouse sera (1:500 dilution in 300  $\mu$ L blocking buffer) were incubated for 1 h in each of the two wells (total 2 h) with immobilized “Hyperphage” to preclear anti-phage serum antibodies.

The precleared serum was transferred into 2 wells (150  $\mu$ L each) previously coated with anti-mouse (IgA, IgG and IgM specific) antibodies (antikoerper-online, ABIN376851) and saturated with blocking buffer in order to capture the serum antibodies used for the panning procedure (2 h at room temperature). Excess serum antibodies and other serum proteins were removed by washing 3 times with washing buffer (PBS supplemented with 0.05% Tween 20).

**Selection of specifically bound oligopeptides.** Panning is an enrichment process of specifically reactive oligopeptide phage clones over several consecutive rounds (a schematic illustration is provided in Figure S1). Therefore, three rounds of panning were performed using  $2 \times 10^{10}$  cfu oligopeptide phage library (in 150  $\mu$ L blocking buffer) as input in round one and  $3 \times 10^{11}$  cfu of amplified oligopeptide phage as input in rounds two and three.

For the selection of specifically bound oligopeptides, the oligopeptide phage were incubated with the immobilized serum antibodies for 2 h at room temperature. Unbound oligopeptide phage were removed by stringent washing (10 times with washing buffer). Bound oligopeptide phage particles were eluted with 200  $\mu$ L elution buffer (10  $\mu$ g/mL trypsin in PBS) per well for 30 min at 37 °C. The eluted phage titer was determined using 10-fold dilutions of 10  $\mu$ L of the elution followed by *E. coli* TOP10F<sup>+</sup> infection and plating onto  $2 \times$  YT agar supplemented with ampicillin.

**Amplification of eluted phage.** In order to amplify the eluted phage for input in the next panning round, the remaining 390  $\mu$ L of elutions (see above) were used to infect 5 mL of an *E. coli* TOP10F<sup>+</sup> culture ( $OD_{600} = 0.5$ ) for 30 min at 37 °C. Cells were plated on 15 cm  $2 \times$  YT-GA agar plates and incubated at 37 °C overnight in order to allow amplification of eluted phage for the next panning round. The cells were scraped using 5 mL of  $2 \times$  YT-GA. For the production of phage, 30 mL of  $2 \times$  YT-GA were inoculated ( $OD_{600} < 0.1$ ) with the scraped cells and cultivated up to  $OD_{600} = 0.5$ . Phage particles were produced following the same procedure as described above for packaging of oligopeptide phage libraries. Amplified phage were precipitated only once (1 h on ice) after adding 1/5 volume precipitation buffer. After pelleting (1 h at 4 °C,  $3,220 \times g$ ), the oligopeptide phage were resuspended in 500  $\mu$ L phage dilution buffer. Remaining bacteria were pelleted for 2 min at  $16,100 \times g$ , the amplified oligopeptide phage containing supernatant was collected and stored at 4 °C until used as input phage for the next panning round.

After the third panning round, 10-fold dilutions of eluted phage were used to infect 50  $\mu$ L *E. coli* TOP10F<sup>+</sup> ( $OD_{600} = 0.5$ ), the cells were plated on  $2 \times$  YT-GA agar plates and incubated at 37 °C overnight in order to obtain single colonies that allow screening of monoclonal oligopeptide phage. To screen the enriched clones after the third round of each panning for immunogenic oligopeptides, 92 randomly selected clones were analyzed by monoclonal phage ELISA.

**Monoclonal screening ELISA.** In order to screen the clones selected in high-throughput manner using TNF <sup>$\Delta$ ARE/+</sup> serum antibodies as described above, monoclonal oligopeptide phage were produced. A 96-well microtiter plate (MTP) was supplemented with 180  $\mu$ L  $2 \times$  YT-GA medium, inoculated with single colonies after the third panning round (see above) and incubated (34 °C, 800 rpm, Labnet Vortemp 56) overnight. For phage production, 175  $\mu$ L  $2 \times$  YT-GA per well were inoculated with 10  $\mu$ L of the overnight culture and incubated for 2 h at 37 °C and 800 rpm to reach logarithmic growth. The cells were infected with  $5 \times 10^9$  cfu Hyperphage (M13K07 $\Delta$ gIII) for 30 min at 37 °C. To ensure antibiotic resistance, the cells were incubated for another 30 min at 37 °C and 800 rpm. In order to change the medium for phage production, the cells were pelleted for 10 min at  $3,220 \times g$  and resuspended in 180  $\mu$ L  $2 \times$  YT-AK followed by phage production at 30 °C and 800 rpm overnight. The cells were pelleted and the phage were precipitated by adding 40  $\mu$ L precipitation buffer to 150  $\mu$ L phage containing supernatant.

To capture the oligopeptide phage, rabbit anti-M13 (pVIII specific, 1:5,000 in PBS, PA1-46334, Pierce Biotechnology) was immobilized on an ELISA plate (Costar) at 4 °C overnight. Wells were saturated with blocking buffer for 1 h at room temperature. The plates were washed 3 times with washing buffer before adding the monoclonal phage (40  $\mu$ L in 100  $\mu$ L blocking buffer). Phage particles were captured for 2 h at room temperature. To reduce background phage binding, the mouse serum samples (1:500) were precleared in blocking buffer containing  $10^{10}$  cfu/mL Hyperphage for 2 h at room temperature. Remaining monoclonal phage were removed by 3 washing steps with washing buffer and the precleared sera were transferred to the wells containing immobilized oligopeptide phage particles and incubated for 2 h at room temperature. The serum was removed by 3 washing steps with washing buffer and bound serum antibodies were detected using goat anti-mouse (IgA, IgG, IgM) antibodies conjugated with horseradish peroxidase (HRP) (antikoerper-online, ABIN376237, 1:8,000 in blocking buffer) for 1.5 h at room temperature. Excess detection antibody was removed by 3 washing steps and the ELISA was developed with TMB substrate solution. Signals were detected using an ELISA reader (TECAN Sunrise, 450 nm, reference 620 nm).

**Immunoblot of oligopeptide phage.** In order to validate the immunogenic character of the displayed oligopeptide, staining of the oligopeptide::pIII fusion protein by TNF <sup>$\Delta$ ARE/+</sup> sera and wildtype controls was analyzed by immunoblotting of the selected oligopeptide phage. The oligopeptide phage ( $4 \times 10^8$ – $6 \times 10^9$  cfu) proteins were separated on a 12% polyacrylamide gel and blotted onto PVDF membranes. Membranes were blocked by incubating in blocking solution (PBS supplemented with 2% skim milk powder (Roth, Germany)) at 4 °C overnight. Sera of fifteen 18-week-old TNF <sup>$\Delta$ ARE/+</sup> and 19 C57BL/6N wt male mice were pooled. These pools and the

reference sera used for selection were each diluted 200-fold in blocking solution containing  $1.5 \times 10^{10}$  cfu/mL M13K07 wildtype phage for competition and were incubated at 4 °C overnight. Membranes were stained with the sera, mouse anti-hexahistidine (Qiagen Germany, 34660, 1:500 in blocking buffer) and mouse anti-pIII (MoBiTec, Germany, PSKAN3, 1:1,000 in blocking buffer) controls for 2.5 h at room temperature and excess antibodies were removed by using washing buffer (PBS supplemented with 0.1% Tween 20). Bound serum antibodies were detected with a goat anti-mouse (IgA, IgG, IgM) antibody conjugated with HRP (antikoerper-online, ABIN376237, 1:4,000 in blocking buffer) for 1.5 h at room temperature. Immunoblots were developed using diaminobenzidine substrate.

**ELISA based measurement of antibody response.** The ELISA plates (Costar) were coated with 200 ng/well streptavidin (Serva) at 4 °C overnight and the wells were blocked with  $1 \times$  PBS supplemented with 0.1% Tween 20 and 2% BSA. After three washing steps with washing buffer ( $H_2O$  supplemented with 0.05% Tween 20), 200 ng/well of the three selected peptides produces as synthetic biotinylated peptides (Peps4LS, Heidelberg, Germany) were immobilized for 2 h at room temperature. Excess peptides were removed by three washing steps with washing buffer. The mouse sera were diluted 1:100 in blocking buffer (PBS supplemented with 2% skim milk powder (Roth) and 0.1% Tween 20) and incubated with the immobilized peptides and the streptavidin controls for 2.5 h at room temperature. Unbound serum antibodies were removed by three washing steps with washing buffer. The bound serum antibodies were detected with goat anti-mouse (IgA, IgG, IgM) antibodies conjugated with HRP (antikoerper-online, ABIN376237, 1:8,000 in blocking buffer) for 1.5 h at room temperature and the ELISA was developed with TMB substrate solution as described above. Serum reactivity with the peptides was measured in doublets. The measured background binding to streptavidin was subtracted from the mean serum reactivity to compensate inter-individual differences in background binding.

**High-throughput sequencing.** For 16S rRNA amplicon analysis, frozen caecal samples were processed, sequenced, and analyzed as described previously<sup>65</sup>. Briefly, DNA was purified after mechanical lysis of microbial cells using NucleoSpin gDNA columns (Macherey-Nagel), the V3-V4 region of 16S rRNA genes was amplified (25 cycles) using primers 341F-785R<sup>66</sup>, and purified using magnetic beads (Beckmann). Sequences obtained by paired-end sequencing using a MiSeq Illumina sequencer were analyzed by in-house developed pipelines following the UPARSE approach<sup>67</sup> available via the IMNGS platform (www.imngs.org)<sup>68</sup>. Defaults setting and a minimum/maximum length of 400/450 nt for merged reads were used. Operational taxonomic units (OTUs) were picked at a threshold of 97% sequence similarity, and only those occurring at a relative sequence abundance of at least 0.5% in one sample were kept for further analysis.

For next generation sequence analysis of ORF enriched phage libraries, each library was packaged (see above) in triplicates and phagemid preparations from *E. coli* cells after infection with the library oligopeptide phage were used as template. In order to prevent any infection bias, the displayed oligopeptides were removed from the phage particles by treatment with 10 µg/mL trypsin for 30 min at 37 °C (the oligopeptide::pIII fusion protein contains a trypsin site between oligopeptide and pIII). Twenty mL of an *E. coli* TOP10F<sup>7</sup> culture ( $OD_{600} = 0.5$ ) were infected with  $1 - 6 \times 10^9$  cfu (at least 20-fold excess of initial library diversity) of the oligopeptide phage library (30 min, 37 °C). For plasmid preparation, the cells were plated on 2xYT-GA agar plates, cultured overnight at 37 °C and scraped the next day. The cells were pelleted (10 min,  $3,220 \times g$ ) and plasmid DNA was isolated by NucleoSpin Plasmid EasyPure Kit (Macherey-Nagel, Berlin, Germany) according to manufacturers' instruction.

Nucleotides libraries were constructed by PCR (25 cycles) using primer Fwd-5'-GCTCAGCCGCGATGG and Rev-5'-CAGCTCTGATATCTTTGGATCCC. They were purified and sequenced in paired-modus using a MiSeq Illumina sequencer. Reads were merged, cleaned from primer and vector sequences, and those with > 0.1% expected error rate were discarded using USEARCH v8.1.18<sup>69</sup>. When different samples or replicates were compared, sequences were first dereplicated to allow comparison of unique reads across samples.

In order to analyze the insert size distribution and to estimate the capacity for successful translation to oligopeptides of the metagenomic fragments, the sequence reads were trimmed from adapter and vector sequences and the stop codon content of the inserts was analyzed for one of the libraries. Around 4% of those reads had truncated or expanded vector sequences that would modify the correct reading frame in the analysis and were removed. All inserts were scanned (starting from the 5' end and translating downstream) for stop codons (TAA, TGA or TAG) that would interrupt the translation. The number of sequences which could be rescued by alternative start codons (ATG) downstream of the last stop codon were determined by walking in opposite direction (from 3' end) to ensure alternative start codons have the potential to initiate translation of in-frame products.

**Statistical analysis.** Statistical analysis was performed by non-parametric Mann-Whitney test using the GraphPad Prism software package.

## References

- van den Berg, S. *et al.* A multiplex assay for the quantification of antibody responses in *Staphylococcus aureus* infections in mice. *J. Immunol. Methods* **365**, 142–148 (2011).
- Ravindran, R. *et al.* Plasma antibody profiles in non-human primate tuberculosis. *J. Med. Primatol.* **43**, 59–71 (2014).
- Robinson, W. H. *et al.* Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat. Med.* **8**, 295–301 (2002).
- Schirwitz, C. *et al.* Sensing Immune Responses with Customized Peptide Microarrays. *Biointerphases* **7**, 1–9 (2012).
- Delvecchio, V. G. *et al.* Proteomic profiling and identification of immunodominant spore antigens of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*. *Appl. Environ. Microbiol.* **72**, 6355–6363 (2006).
- Jacobsen, I. D., Meens, J., Baltes, N. & Gerlach, G.-F. Differential expression of non-cytoplasmic *Actinobacillus pleuropneumoniae* proteins induced by addition of bronchoalveolar lavage fluid. *Vet. Microbiol.* **109**, 245–256 (2005).

7. LaFrentz, B. R., LaPatra, S. E., Call, D. R., Wiens, G. D. & Cain, K. D. Identification of immunogenic proteins within distinct molecular mass fractions of *Flavobacterium psychrophilum*. *J. Fish Dis.* **34**, 823–830 (2011).
8. Zhang, C. G., Chromy, B. A. & McCutchen-Maloney, S. L. Host–pathogen interactions: a proteomic view. *Expert Rev. Proteomics* **2**, 187–202 (2005).
9. Beranova-Giorgianni, S. Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: strengths and limitations. *TrAC Trends Anal. Chem.* **22**, 273–281 (2003).
10. Rabilloud, T., Chevillet, M., Luche, S. & Lelong, C. Two-dimensional gel electrophoresis in proteomics: Past, present and future. *J. Proteomics* **73**, 2064–2077 (2010).
11. Urquhart, B. L., Cordwell, S. J. & Humphery-Smith, I. Comparison of predicted and observed properties of proteins encoded in the genome of *Mycobacterium tuberculosis* H37Rv. *Biochem. Biophys. Res. Commun.* **253**, 70–79 (1998).
12. Hoppe, S., Bier, F. F. & Nickisch-Rosenegk, M. v. Rapid Identification of Novel Immunodominant Proteins and Characterization of a Specific Linear Epitope of *Campylobacter jejuni*. *PLoS One* **8**, e65837 (2013).
13. Danckert, L., Hoppe, S., Bier, F. F. & von Nickisch-Rosenegk, M. Rapid identification of novel antigens of *Salmonella* Enteritidis by microarray-based immunoscreening. *Microchim. Acta* **181**, 1707–1714 (2014).
14. Kunnath-Velayudhan, S. *et al.* Dynamic antibody responses to the *Mycobacterium tuberculosis* proteome. *Proc. Natl. Acad. Sci.* **107**, 14703–14708 (2010).
15. Chen, C.-S. *et al.* Identification of Novel Serological Biomarkers for Inflammatory Bowel Disease Using *Escherichia coli* Proteome Chip. *Mol. Cell. Proteomics* **8**, 1765–1776 (2009).
16. Helb, D. A. *et al.* Novel serologic biomarkers provide accurate estimates of recent *Plasmodium falciparum* exposure for individuals and communities. *Proc. Natl. Acad. Sci.* **112**, E4438–E4447 (2015).
17. Chen, Y.-T., Gure, A. O. & Scanlan, M. J. Serological analysis of expression cDNA libraries (SEREX): an immunoscreening technique for identifying immunogenic tumor antigens. *Methods Mol. Med.* **103**, 207–216 (2005).
18. Türeci, O., Sahin, U. & Pfreundschuh, M. Serological analysis of human tumor antigens: molecular definition and implications. *Mol. Med. Today* **3**, 342–349 (1997).
19. Lodes, M. J. *et al.* Bacterial flagellin is a dominant antigen in Crohn disease. *J. Clin. Invest.* **113**, 1296–1306 (2004).
20. Danner, S. & Belasco, J. G. T7 phage display: A novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. *Proc. Natl. Acad. Sci.* **98**, 12954–12959 (2001).
21. Smith, G. P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315–1317 (1985).
22. Frenzel, A., Schirrmann, T. & Hust, M. Phage display-derived human antibodies in clinical development and therapy. *mAbs* **0**, doi: 10.1080/19420862.2016.1212149 (2016).
23. Hoogenboom, H. R. Selecting and screening recombinant antibody libraries. *Nat. Biotechnol.* **23**, 1105–1116 (2005).
24. Dübel, S., Stoevesandt, O., Taussig, M. J. & Hust, M. Generating recombinant antibodies to the complete human proteome. *Trends Biotechnol.* **28**, 333–339 (2010).
25. Scott, J. K. & Smith, G. P. Searching for peptide ligands with an epitope library. *Science* **249**, 386–390 (1990).
26. Faix, P. H. *et al.* Phage display of cDNA libraries: enrichment of cDNA expression using open reading frame selection. *BioTechniques* **36**, 1018–1022, 1024, 1026–1029 (2004).
27. D'Angelo, S. *et al.* Filtering 'genic' open reading frames from genomic DNA samples for advanced annotation. *BMC Genomics* **12**, S5 (2011).
28. Di Niro, R. *et al.* Rapid interactome profiling by massive sequencing. *Nucleic Acids Res.* **38**, e110–e110 (2010).
29. Caberoy, N., Alvarado, G. & Li, W. Identification of Calpain Substrates by ORF Phage Display. *Molecules* **16**, 1739–1748 (2011).
30. Gupta, A. *et al.* A Novel Helper Phage Enabling Construction of Genome-Scale ORF-Enriched Phage Display Libraries. *PLoS One* **8**, e75212 (2013).
31. Rondot, S., Koch, J., Breitling, F. & Dübel, S. A helper phage to improve single-chain antibody presentation in phage display. *Nat. Biotechnol.* **19**, 75–78 (2001).
32. Soltes, G. *et al.* On the influence of vector design on antibody phage display. *J. Biotechnol.* **127**, 626–637 (2007).
33. Kügler, J., Zantow, J., Meyer, T. & Hust, M. Oligopeptide M13 Phage Display in Pathogen Research. *Viruses* **5**, 2531–2545 (2013).
34. Hust, M. *et al.* Enrichment of open reading frames presented on bacteriophage M13 using hyperphage. *BioTechniques* **41**, 335–342 (2006).
35. Kügler, J. *et al.* Identification of immunogenic polypeptides from a *Mycoplasma hyopneumoniae* genome library by phage display. *Appl. Microbiol. Biotechnol.* **80**, 447–458 (2008).
36. Naseem, S. *et al.* Phage display-based identification and potential diagnostic application of novel antigens from *Mycoplasma mycoides* subsp. *mycoides* small colony type. *Vet. Microbiol.* **142**, 285–292 (2010).
37. Meyer, T. *et al.* Identification of immunogenic proteins and generation of antibodies against *Salmonella* Typhimurium using phage display. *BMC Biotechnol.* **12**, 29 (2012).
38. Connor, D. O., Zantow, J., Hust, M., Bier, F. F. & von Nickisch-Rosenegk, M. Identification of Novel Immunogenic Proteins of *Neisseria gonorrhoeae* by Phage Display. *PLoS One* **11**, e0148986 (2016).
39. Becker, M. *et al.* Application of M13 phage display for identifying immunogenic proteins from tick (*Ixodes scapularis*) saliva. *BMC Biotechnol.* **15** (2015).
40. Hugot, J. P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).
41. Duerr, R. H. *et al.* A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. *Science* **314**, 1461–1463 (2006).
42. Jostins, L. *et al.* Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
43. Manichanh, C. *et al.* Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55**, 205–211 (2006).
44. Gevers, D. *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15**, 382–392 (2014).
45. Prideaux, L., De Cruz, P., Ng, S. C. & Kamm, M. A. Serological Antibodies in Inflammatory Bowel Disease: A Systematic Review. *Inflamm. Bowel Dis.* **18**, 1340–1355 (2012).
46. Hörmannspurger, G., Schaubeck, M. & Haller, D. Intestinal Microbiota in Animal Models of Inflammatory Diseases. *ILAR J. Natl. Res. Counc. Inst. Lab. Anim. Resour.* **56**, 179–191 (2015).
47. Kontoyiannis, D., Pasarakis, M., Pizarro, T. T., Cominelli, F. & Kollias, G. Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: implications for joint and gut-associated immunopathologies. *Immunity* **10**, 387–398 (1999).
48. Schaubeck, M. *et al.* Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of failure in antimicrobial defence. *Gut* doi: 10.1136/gutjnl-2015-309333 (2015).
49. Roulis, M. *et al.* Host and microbiota interactions are critical for development of murine Crohn's-like ileitis. *Mucosal Immunol.* doi: 10.1038/mi.2015.102 (2015).
50. Jost, L. Partitioning diversity into independent alpha and beta components. *Ecology* **88**, 2427–2439 (2007).
51. Baur, P. *et al.* Metabolic Phenotyping of the Crohn's Disease-like IBD Etiopathology in the TNF<sup>ΔARE/WT</sup> Mouse Model. *J. Proteome Res.* **10**, 5523–5535 (2011).
52. Ciric, M. *et al.* Metasecretome-selective phage display approach for mining the functional potential of a rumen microbial community. *BMC Genomics* **15**, 356 (2014).

53. Gagic, D., Ciric, M., Wen, W. X., Ng, F. & Rakonjac, J. Exploring the Secretomes of Microbes and Microbial Communities Using Filamentous Phage Display. *Front. Microbiol.* **7**, 429 (2016).
54. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
55. Li, J. *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
56. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
57. Goodman, A. L. *et al.* Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc. Natl. Acad. Sci.* **108**, 6252–6257 (2011).
58. Matochko, W. L. *et al.* Deep sequencing analysis of phage libraries using Illumina platform. *Methods* **58**, 47–55 (2012).
59. Cárcamo, J. *et al.* Unexpected frameshifts from gene to expressed protein in a phage-displayed peptide library. *Proc. Natl. Acad. Sci. USA* **95**, 11146–11151 (1998).
60. Goldman, E., Korus, M. & Mandeckj, W. Efficiencies of translation in three reading frames of unusual non-ORF sequences isolated from phage display. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **14**, 603–611 (2000).
61. Lukens, L. N. & Buchanan, J. M. Biosynthesis of the purines. XXIII. The enzymatic synthesis of N-(5-amino-1-ribosyl-4-imidazolylcarbonyl)-L-aspartic acid 5'-phosphate. *J. Biol. Chem.* **234**, 1791–1798 (1959).
62. Ginder, N. D., Binkowski, D. J., Fromm, H. J. & Honzatko, R. B. Nucleotide complexes of *Escherichia coli* phosphoribosylaminoimidazole succinocarboxamide synthetase. *J. Biol. Chem.* **281**, 20680–20688 (2006).
63. Lagkouvardos, I. *et al.* The mouse intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nat. Microbiol.*, doi 10.1038/nmicrobiol.2016.131 (2016).
64. Katakura, K. *et al.* Toll-like receptor 9-induced type I IFN protects mice from experimental colitis. *J. Clin. Invest.* **115**, 695–702 (2005).
65. Lagkouvardos, I. *et al.* Gut metabolites and bacterial community networks during a pilot intervention study with flaxseeds in healthy adult men. *Mol. Nutr. Food Res.* **59**, 1614–1628 (2015).
66. Klindworth, A. *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**, e1 (2013).
67. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
68. Lagkouvardos, I. *et al.* IMNGS: A massive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci Rep.*, doi: 10.1038/srep33721 (2016).
69. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinforma. Oxf. Engl.* **26**, 2460–2461 (2010).
70. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).

## Acknowledgements

We are grateful to Dr. George Kollias (Fleming Institute) for providing TNF<sup>ΔARE/+</sup> mice. We thank Gustavo Marçal Schmidt Garcia Moreira for carefully reading and commenting on the manuscript.

## Author Contributions

J.Z. designed and performed experiments, analyzed data and drafted the main manuscript text. S.J. and I.L. performed experiments and analyzed data. S.K. performed experiments. S.D. and P.L. analyzed data and were involved in the design of the study. T.C. and M.H. conceived the study, analyzed data and drafted the manuscript. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Zantow, J. *et al.* Mining gut microbiome oligopeptides by functional metaproteome display. *Sci. Rep.* **6**, 34337; doi: 10.1038/srep34337 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016