Microbial ecology perturbation in human IgA deficiency

Jehane Fadlallah, Hela El Kafsi, Delphine Sterlin, Catherine Juste, Christophe Parizot, Karim Dorgham, Gaëlle Autaa, Doriane Gouas, Mathieu Almeida, Patricia Lepage, et al.

Title: Microbial ecology perturbation in human IgA deficiency

Authors: Jehane Fadlallah[1,2,†], Hela El Kafsi[1,2,†], Delphine Sterlin[1,2], Catherine Juste[3], Christophe Parizot[2], Karim Dorgham[1], Gaëlle Autaa[1], Doriane Gouas[1,2], Mathieu Almeida[4], Patricia Lepage[3], Nicolas Pons[5], Emmanuelle Le Chatelier[3,5], Florence Levenez[5], Sean Kennedy[5], Nathalie Galleron[5], Jean-Paul Pais de Barros[6,7], Marion Malphettes [6,7], Lionel Galicier[8], David Boutboul[8], Alexis Mathian[2], Makoto Miyara[2], Eric Oksenhendler[8], Zahir Amoura[1,2], Joel Doré[3], Claire Fieschi[8], S. Dusko Ehrlich[3,5], Martin Larsen[1,2,‡,*], Guy Gorochov[1,2,‡,*]

Affiliations:
[1]Sorbonne Universités, UPMC Université Paris 6, INSERM, CNRS, Centre d'Immunologie et des Maladies Infectieuses (CIMI Paris UMRS 1135), 75013, Paris, France.
[2]AP-HP, Groupement Hospitalier Pitié-Salpêtrière, Département d'Immunologie, 75013, Paris, France.
[3]UMR1319 Micalis, INRA, Jouy-en-Josas, France.
[4]Center for Bioinformatics and Computational Biology, University of Maryland, Paint Branch Road, College Park, MD 20742, USA.
[5]INRA, Institut National de la Recherche Agronomique, US1367 MetaGenoPolis, 78350 Jouy en Josas, France.
[6]INSERM, LNC UMR866, University Bourgogne Franche-Comté, F-21000 Dijon, France
[7]LipSTIC LabEx, Fondation de Coopération Scientifique Bourgogne-Franche Comté, F-21000 Dijon, France
[8]Université Paris Diderot Paris 7, Department of Clinical Immunology, Hôpital Saint-Louis, Assistance Publique Hôpitaux de Paris (APHP), EA3518, 75010, Paris, France
[†]These authors contributed equally to this work
[‡]These authors jointly directed this work

*To whom correspondence should be addressed: Martin Larsen (martin.larsen@upmc.fr) or Guy Gorochov (guy.gorochov@upmc.fr).

One Sentence Summary: IgA acts as an immune *avant-garde* influencing bacterial fitness in the intestinal lumen, away from the intestinal barrier, a role that can only partly be rescued by IgM in selective IgA deficient patients.

Word Count: Abstract: 150 words. Text: 7942 words. Legends: 1378 words.
Figures: 8 + 5 supplementary
Tables : 1 + 1 supplementary
Supplementary data files: 3
References: 72

Abstract:

Paradoxically, loss of IgA, one of the most common antibodies, is not irrevocably leading to severe infections in humans, but is rather associated with relatively mild respiratory infections, atopy and autoimmunity. IgA might therefore play covert roles, not uniquely associated with control of pathogens. We show that human IgA deficiency is not associated with massive quantitative perturbations of gut microbial ecology. Metagenomic analysis rather highlights an expected pathobiont expansion, but a less expected depletion in some typically beneficial symbionts. Gut colonization by species usually present in the oropharynx is also reminiscent of spatial microbiota disorganization. IgM only partially rescues IgA deficiency, as not all typical IgA targets are efficiently bound by IgM in the intestinal lumen. Altogether IgA appears to play a non-redundant role at the forefront of the immune/microbial interface, away from the intestinal barrier, ranging from pathobiont control and regulation of systemic inflammation, to preservation of commensal diversity and community networks.

[Main Text:]

Introduction

Eukaryotes have developed spectacular ways not only to protect themselves from pathogens, but also to benefit from unique and essential features of surrounding organisms (symbionts). Mammals are indeed highly dependent on their consortia of symbionts (microbiota) that serve both to optimize processing of nutrients and to protect from opportunistic agents by competition. Innate immune mechanisms controlling host-microbiota mutualism are physically localized, activated by bacterial contact culminating into breach of the gut mucosal firewall. Immediate non-specific host responses involve the secretion of defensins and the intra-luminal recruitment of innate immune cells such as neutrophils that encapsulate commensals and limit their contact with surrounding gut epithelium[1]. Such a potent response comes with a fitness cost to the commensal community and the benefits this brings to the host. It is therefore postulated that an evolutionary pressure took place to acquire mechanisms acting in the lumen, which would regulate microbial communities, thereby reducing the frequency of breaching the gut barrier and activation of innate immunity. This role could be played by antibodies, and most likely by secretory IgA[2, 3]. However, whereas antibody responses to pathogens have been intensively studied, much less interest has been devoted to the study of antibody relations with symbionts. Murine models of IgA deficiency have been studied and display modifications of the microbiome-immune interface. In such models, IgA deficiency was obtained by inducing either: (i) defects in IgA class switch recombination (CSR)[4-6], (ii) defects in IgA transport into the gut lumen (pIgR$^{-/-}$ mice[7, 8], and J-chain$^{-/-}$ mice[9]), or (iii) reduction of IgA repertoire diversity without altering IgA levels. (AID knock-in mutation[10], PD1$^{-/-}$ mice[11] and FoxP3^{+}CD4^{+} depleted mice[12]). Models impairing CSR mechanisms (i) are associated with a gut dysbiosis defined by an expansion of anaerobes, predominantly SFB and *Clostridiales,* as well as nodular hyperplasia secondary to hyper-activation of germinal center B cells induced by microbial antigens[4, 5]. Models impairing IgA secretion (ii) into the gut lumen were associated with altered microbiota composition, increased susceptibility to induced colitis, higher bacterial translocation to mesenteric lymph nodes after *Salmonella typhimurium* challenge, and lack of protective immunity against cholera toxin[7, 8]. Of note, alterations of gut microbiota ecosystems were observed in the small intestine, while large intestine communities were much less affected by the absence of IgA[5], possibly because IgA predominantly target commensal bacteria in the small

intestine, but not in the colon, as shown by Bunker *et al.,* both in mice and humans[13]. Altogether, altered microbiome composition, increased susceptibility to microbial translocation, reduced microbial diversity and reduced microbial fitness are shared features of IgA deficiency models.

These results are in line with an early conception of IgA function, mainly presented as a neutralising antibody, whose role would mainly be to exclude potentially harmful microbes and toxins from intimate contact with intestinal epithelia, thereby preserving mucosal barrier integrity[14]. IgA deficiency is relatively common in human adults, occurring in about 1 in 500 Caucasian individuals[15, 16]. Although human selective IgA deficiency (SIgAd, i.e.: patients deficient for IgA, but sufficient for all other antibody isotype) was for a long time considered asymptomatic, recent longitudinal studies have revealed that 80% of patients are symptomatic, when assessing complications more broadly[17], or when follow-up is extended[18]. Indeed, human IgA deficiency is associated with recurrent mucosal infections, auto-immunity and intestinal disorders such as inflammatory bowel disease and lymphoid hyperplasia[19, 20], To explain the mild phenotype observed in SIgAd patients, it is proposed that IgM may effectively replace IgA as the predominant antibody in secretions[21, 22]. But if IgA indeed represents a merely redundant component of the immune system, it appears paradoxical that it was so well conserved in evolution, and that it is massively produced at the individual level (about 66mg/kg/day of IgA are secreted everyday[23]). Indeed, IgA also appear to orchestrate the beneficial tolerance established between the host and its gut commensal microbiome. These mutualistic host-microbial relationships were emphasized in animal models with an immune system reduced to a single monoclonal antibody of known bacterial specificity challenged by a limited microbial diversity. Indeed, while antibody binding was reducing bacterial fitness, it also resulted in reduced intestinal production of pro-inflammatory signals, hence allowing bacterial tolerance by the host[24]. Similarly, flagellin-specific sIgA may promote tolerance by reducing bacterial motility through modulation of flagellin transcription[25].

While these models have provided important examples and possible mechanisms through which antibodies can imprint specific microbes, the targets of polyclonal IgA and their global impact on the microbiome remain poorly defined. Although it was recently shown in an animal model that IgA-coated bacteria include pro-inflammatory elements[26], it is not known whether IgA

preferentially bind potentially harmful bacteria, and/or commensals[27]. It is not known either whether IgM can indeed replace IgA at no expense to host/bacterial homeostasis. Finally, the relations between SIgAd and systemic autoimmunity are not well understood.

Here, we studied the composition of SIgA-bound gut microbiota in healthy individuals and evaluated alterations of this bacterial consortium in IgA deficient patients. In order to get insights on the specific contributions of IgA on host/microbial symbiosis we also explored systemic immune responses in these patients.

Results

Patients lacking IgA-producing B cells and seric IgA also lack free digestive IgA

Selective IgA deficiency represents a bio-clinical entity that is defined by serologic means, namely undetectable seric IgA titers (<0.07mg/mL) with normal IgG levels[28]. SIgAd patients are known to present low or undetectable salivary IgA levels[21, 22], but their gut IgA status had not been assessed. We serologically confirmed a status of IgA deficiency in 21 patients that could be included in this study because, among other exclusion criteria, they did not receive antibiotic treatment 3 months prior to inclusion. Compared to age- and sex-matched healthy donors (HD, n=34), these SIgAd patients had undetectable seric (Supplementary Figure S1A), and scarce digestive IgA levels, (43[0-206] in HD vs 0[0-21] µg of free IgA per g of stool in SIgAd, $p<0.0001$, Figure 1A), while their seric IgG levels were preserved (Supplementary Figure S1B).

Circulating IgA^+ B cells were undetectable in all patients except one (0.1[0-1.3]% in SIgAd vs 7.1[2.4-14.4]% in HD $p<0.0001$), whereas proportion of $CD19^+IgG^+$ cells among B cells were similar in both groups (12.55[5.32-29.6]% vs 13.1[0.793-37.6]%, p=0.8004, Figure 1B). Compared to controls, SIgAd patients are also characterized by a depletion of $CD19^+CD27^+IgD^-$ switched memory B cells among total $CD19^+$ B cells (20[5.17-34.1]% vs 14.9[3.3-38.1]%, p=0.0328, Figure 1C). These data show that secretory IgA deficiency affects both peripheral blood and distant organs, such as the intestine, both at the cellular and the protein level.

IgA bound bacteria are usually absent from SIgAd microbiota

The clinical spectrum of digestive SIgAd-associated symptoms varies from very mild to severe and it remains unknown whether residual digestive IgA would account for pauci-symptomatic presentations. We therefore used a flow-cytometry assay, derived from our previously published technology[29], in order to test whether traces of IgA might be detected at the surface of the fecal microbiota, although free digestive IgA is usually not detectable in SIgAd, as shown above. The protocol was modified to assess levels of mucosal antibodies targeting colonic microbiota *in vivo* (Figure 1D). IgA, as the main mucosal antibody in healthy donors, binds a median percentage of 7.6 [0.8-17.6]% of the whole fecal microbiota (Figure 1D). Close examination of flow cytometry profiles also reveals that IgA-bound microbiota can be subdivided into IgA^{dim} and IgA^{bright} bacterial populations (Figure 1D, left panel). None of these IgA^+ subsets were significantly

detected in SIgAd patients, except one (Figure 1D, right panel). Of note, this patient also had detectable IgA^+ circulating B cells (Figure 1B). This interesting case confirms that some patients diagnosed SIgAD by serologic means indeed retain secretory IgA, that can be detected by the very sensitive bacterial flow cytometry technique. These data nevertheless establish that microbiota-bound IgA are usually undetectable in SIgAd patients.

Preserved global microbiome structure in IgA deficiency

In order to study the global impact of digestive IgA deficiency on the gut microbiome we performed shotgun sequencing[30, 31] of fecal samples in 34 HD and 17 SIgAd patients. High quality reads from each sample were mapped to a reference catalogue of 3.9 million genes[32]. Taxonomic abundances were computed at the level of co-abundance gene groups (CAGs), and subsequently binned at broader taxonomic levels (genus, family, order, class and phylum). CAGs contain at least 50 different genes. Metagenomic species (MGS) are defined as larger CAGs with very high connectivity and a defined minimal size of at least 700 genes. This approach to study microbiome composition presents the advantage to overcome the limited resolution of previous methods used for metagenomic or 16S amplicon data analysis that rely on comparisons to reference genomes, offering the possibility to comprehensively profile the diversity of a clinical sample, and to potentially identify previously uncharacterized microbes[32].

Metagenomic analysis revealed a similar representation of the dominant phyla in the two groups (HD vs SIgAd, Bacteroidetes: 44.55% vs 45.52%, p=1, Firmicutes: 50.52% vs 48.35%, p=0.7774, Proteobacteria: 2.19% vs 3%, p=1, Actinobacteria: 0.65% vs 0.84%, p=0.99 and unclassified phyla: 1.5% vs 1.41%, p=1, Figure S2A). Microbiota diversity was not different between the two groups, either when including all MGS regardless of their phylum (median Shannon index, HD vs SIgAd: 4.008 [2.501-4.698] vs 3.946 [1.690-4.511], p=0.6344), or when comparing MGS diversity within each phylum (HD vs SIgAd, 2.529 [1.225-3.380] vs 2.564 [0.8064-3.234], p=0.8180 for Bacteroidetes, 3.978 [3.067-4.619] vs 3.779 [2.735-4.286], p=0.05 for Firmicutes, 1.688 [1.170-2.223] vs 1.699 [0.5898-2.131], p=0.8421 for Actinobacteria, 1.432 [0.1518-2.132] vs 1.432 [0.3349-2.014], p=0.9042 for Proteobacteria, 2.283 [0.6089-2.830] vs 2.032 [1.161-2.687], p=0.1357 for unclassified phyla, Figure S2B). Furthermore, no significant difference was observed in terms of median MGS richness between the two groups (454 [215-

694] different MGS detected in healthy donors, vs 451 [160-641] in patients, p=0.7553, Figure S2C), nor in gene count (395841 [236166-636394] genes in healthy donors vs 379482 [248232-514298] genes in patients, p=0.4167, Figure S2C). To estimate the stability of the gut microbiota profiles we analyzed the gut microbiota composition of three healthy subjects sampled longitudinally (two samples 12 months apart). Hierarchical cluster analysis outlines strong structural intra-individual sample proximity, suggesting overall temporal stability of individual gut microbiota profiles, at least in healthy subjects (Figure S2C). Altogether, these data reveal that the diversity of SIgAd and control fecal bacterial repertoires do not differ significantly. Our data also suggest that an individual's profile remains stable over time (Figure S4).

Gut microbiota signatures of IgA deficiency.

We reasoned that IgA deficiency might affect relatively discrete bacterial populations, without impacting global microbiome structure at the analysis level applied above. A gene biomarker approach[33] was used to determine which bacteria were expanded or depleted in IgA deficiency, comparing the whole unsorted microbiota of 34 healthy donors with 17 SIgAd patients (Figure S2B). We found 31 differential MGSs between the two groups, 17 being over-represented, whereas 14 were under-represented in IgA deficient patients. Differential MGSs, between controls and patients, were ranked according to increasing statistical significance (Figure 2A). Most MGSs depleted in IgA deficiency (13/14) belong to the Firmicutes phylum, whereas only one belongs to the Bacteroidetes phylum. Among depleted Firmicutes, more than half of them (7/13) belong to the *Lachnospiraceae* family, and 2 are *Ruminococcaceae* (*Faecalibacterium* genus, n=2/2) (Figure 2B).

Conversely, among the 17 expanded MGSs in IgA deficiency, 10/17 are Firmicutes, 4/17 are Bacteroidetes and 3 are Proteobacteria (Gamma-proteobacteria exclusively, including *Escherichia coli*). These 17 MGSs belong to 11 different families and thus are more diverse than depleted species. Interestingly, 3/17 expanded MGSs are usually present in the oropharynx flora (*Streptococcus sanguinis, Veillonella parvula and Haemophilus parainfluenzae*). In addition, we found 2 different species of *Prevotella* over-represented in IgAd (cf. Figure 2A).

IgA targets more likely decline than expand in the absence of IgA

We then postulated that microbial ecology perturbation *in the absence of IgA* might be appreciated in a different manner if we could focus on bacteria more specifically IgA-targeted in healthy controls. We previously verified that metagenomic analysis could reliably be performed on sorted bacterial subsets with a determined lowest analyzable sample size of 10^8 bacteria[34]. IgA^+ fractions were enriched by magnetic sorting in 30 healthy controls, allowing IgA^+ and IgA^- fractions sequencing and differential metagenomic analysis (Figure 3A). The same gene biomarker identification approach as above (Figure S3) allowed the identification of 24 different MGS overrepresented in the IgA+ fraction of healthy controls. These "IgA^+ MGSs" were ranked according to increasing statistical significance at the lowest taxonomic level available (Figure 3B). Among the 24 IgA^+ MGSs, 19 belong to the Firmicutes phylum (among which 12 belong to the Clostridia class, one of these 12 belonging to the *Faecalibacterium* genus), 2 are bacteria from unclassified phylum, 1 is Proteobacteria (one *Escherichia coli* species), 1 belong to the Actinobacteria phylum (*Bifidobacterium bifidum*) and the last identified MGSs belong to the Bacteroidetes phylum (Figure 3C). We finally compared the prevalence of IgA^+ MGSs between SIgAd patients and controls and found that the prevalence of only 4 of them was significantly altered (Figure 3D). While *Escherichia coli* (CAG 4) was found expanded in patients, we rather observed that *Coprococcus comes* (CAG 19), *Clostridium* sp. (CAG 138) and *Dorea* sp. (CAG 73) were depleted in SIgAd (Figure 3E). Altogether, the prevalence of most IgA^+ MGSs does not vary significantly in SIgAd. Furthermore, an IgA^+ MGS does not systematically expand in the context of SIgAd.

IgM memory B cell expansion and IgM digestive secretion in IgAd patients

We then postulated that compensatory immune mechanisms might explain why IgA deficiency is not associated with massive perturbations of gut microbial ecology, as previously suggested[21, 22, 35]. Microbial flow cytometry analysis was used to detect other antibody isotypes on the surface of IgAd microbiomes. IgM was indeed detected at the surface of IgAd microbiota in all IgAd patients analysed (Figure 4A). IgM bound 6.26[0.625-45]%, (n=21) of the whole microbiome in patients, whereas IgM binding was observed in minimal amounts in only two healthy controls out of 34 (0.05[0-2.4]%, $p<0.0001$) (Figure 4B). Measured free fecal IgM levels were consistent with IgM bound levels (0.83 [0-28.1] and 39.9 [0-436.1] µg/g of feces in HD and IgAd patients,

respectively, p=0.0004, Figure 4B). In blood, so-called "IgM only" B cells[36] ($CD19^+CD27^+IgM^+IgD^-$) were increased in patients (5.67[2.88-12.73]% in HD vs 9.14[3.83-43.75]% in patients, p=0.0048), whereas marginal zone cells ($CD19^+CD27^+IgM^+IgD^+$) are similar in both groups (40.2 [17.8-66.3]% and 38.85 [9.29-76.8]% in HD and SIgAd patients, respectively, p=0.7588, Figure 4C). Therefore, in the absence of IgA, IgM is secreted to the digestive tract, where it binds gut commensals. Our data also suggest that $CD19^+CD27^+IgM^+IgD^-$ B cell expansion could account for this compensatory measure.

IgM only partially supply IgA deficiency

We then asked whether IgM could be looked upon as a surrogate for IgA at the immune-microbiota interface. In order to test whether IgM and IgA bind overlapping repertoires of bacteria we separated IgM^+ and IgM^- bacteria from patient's microbiomes (n=10) using flow cytometry. Metagenomic analysis could not be performed on sorted samples as the lowest analyzable sample size of 10^8 bacteria was not reached for all IgM+ fractions. Bacterial fractions were then identified by 16S rRNA sequencing (Figure 4D). Hierarchical clustering of n=33 dominant taxa abundance ratios [$\log_2$ (IgM^+/IgM^-)] displays no evidence of a common pattern of IgM recognition (Figure 4D). Thus, much like IgA responses (Figure S4), also IgM responses seem to be highly variable inter-individually.

We then focused our analysis on the taxa bound by SIgA, or differing between HD and IgA deficient patients, cf. Figures 3B and 2A, respectively). Using the median $\log_2$ (IgM^+/IgM^-) ratios of all the donors, we found that: (i) *Veillonellaceae* family, *Prevotella* genus, *Porphyromonadaceae* family, *Pseudomonas* genus, *Lachnospira* genus, *Faecalibacterium* genus, *Clostridium* genus, *Bifidobacterium* genus and *Bacteroides* genus are overrepresented in IgM^+ fraction, (ii) *Enterobacteriaceae* family, *Streptococcus* genus, *Ruminococcus* genus, *Dorea* genus, *Coprococcus* genus, *Blautia* genus are overrepresented in IgM^- fraction and, (iii), *Acinetobacter* genus, *Erysipelotrichaceae* family, *Anaerostipes* genus, *Eubacterium* genus are equally present in both fractions (Figure 4E). Paired analysis performed on these 19 taxa revealed that *Clostridium* and *Pseudomonas* genera are significantly enriched in the IgM^+ fraction (p=0.013). There is also a trend toward *Veillonellaceae* family and *Bifidobacterium* genus overrepresentation in this fraction (p=0.0625 and p=0.0938), whereas statistical

significance was not reached for the 15 remaining taxa (Figure 4F). Importantly, bacteria belonging to the *Enterobacteriaceae* family are poorly bound by IgM. Altogether, not all typical IgA targets are bound by digestive IgMs in IgAd patients.

IgM responses are correlated with commensal diversity in IgA deficiency

Given that the recent literature substantiates that IgA shapes microbiota composition and diversity in mice[12], we wanted to know whether IgM could play this role in IgA deficiency. Median gut IgM^+ enrichment ratios based on OTU abundances calculated at phylum level in IgM^+ and IgM^- sorted fractions were correlated with the Shannon diversity index (Figure S2A) within each dominant phylum in 9 SIgAd patients (Figure 5A). As shown, IgM^+ enrichment ratio is positively correlated with Actinobacteria phylum diversity (Spearman coefficient, r=0.7167, p=0.0369), whereas no statistic correlation is observed for the three other dominant phyla. Of note, the very narrow range of IgM binding to Firmicutes, and to some extent Proteobacteria, might have precluded identification of potential correlations between IgM binding and diversity in these phyla. Serum IgM responses against two strains belonging to the Actinobacteria phylum (*Bifidobacterium adolescentis and Bifidobacterium longum*) were confirmed in 16 healthy donors by flow cytometry (Supplementary Figure S5A). We then postulated that Actinobacteria diversity could be further reduced in the absence of IgM. We therefore explored common variable immunodeficiency (CVID) patients with total IgA deficiency, and very low or undetectable IgM digestive levels (0[0-115] µg of free IgM/g of stool, Supplementary Figure S5B). As expected, IgM gut microbiota binding is minimal (0.297 [0.08-3.84]%, n=15, Supplementary Figure S5C) in CVID. Shotgun sequencing of the whole CVID gut microbiota and metagenomic analysis (Figure 5B) suggested a significant global loss of MGS richness (454 [305-590] in HD (n=34), vs 375[294-491] in CVID (n=7), p=0.0615, Figure 5C). Interestingly, a significant loss in Actinobacteria phylum diversity was observed in these patients (Shannon diversity index, 1.688 [1.170-2.223] in HD vs 1.342[0.319-1.827] in CVID, p=0.05). A loss, yet more moderate, of the Firmicutes phylum diversity was also observed (p=0.065, Figure 5D). Taken together, the data suggest that in the absence of intestinal IgA, IgM binding preserves Actinobacteria diversity, although this conclusion needs to be validated on a larger CVID cohort.

IgA deficiency is associated with systemic inflammation

We then asked whether the lack of secretory IgA could induce perturbations in host systemic inflammatory versus regulatory responses, in spite of the presence of mucosal IgM responses. Cytokine-secreting circulating $CD4^+$ T cells were measured in both groups (Figure 6A). Proportions of $IFNg^+CD4^+$ T cells were not different (14.6[3.3-25.7]% in HD vs 16.75[2.28-47.8]% in SIgAd, p=0.3932), whereas $IL\text{-}17^+CD4^+$ and $IL\text{-}22^+CD4^+$ T cells were increased in IgA deficiency (0.422[0.04-1.96]% vs 1.49[0.06-3.65]%, p=0.0137 and 0.136[0.06-0.769]% vs 0.866[0.02-4]%, p=0.0104, respectively). Double-positive $IL\text{-}17^+IL\text{-}22^+$ $CD4^+$ T cells were also increased in IgA deficiency (0.05[0.02-0.44]% vs 0.2[0-0.75]%, p= 0.0058). Serum IL-6, IL-10 and IL-17 levels were all elevated in patients (0,6 [0.33-2.4] vs 1 [0.25-34.37] pg/mL, p=0.0315, 0.47[0-1.41] vs 0.87[0.37-5.2] pg/mL, p=0.0001, and 0.06[0-1.47] vs 0.21[0.007-0.92] pg/mL, p=0.0215, respectively) (Figure 6B). Soluble CD14 (sCD14) levels were increased in patient's sera (2063[1147-4283] vs 2841[1399-5187] pg/mL, p=0.0023) (Figure 6C), although LPS concentration was not significantly increased in the same samples (54.15[23.40-77.92] vs 48.10[34.57-97.33]) (Figure 6C). We also observed an increase in $CD4^+PD\text{-}1^+$ cells (7.11[1.86-16.9]% vs 14[3.13-31.6]%, p=0.0093, Figure 6C) in SIgAd patients. The circulating T regulatory cell compartment was not altered by IgA deficiency, since levels of naïve Tregs ($CD45RA^+FoxP3^+$ $CD4^+$ T cells), effector Tregs ($CD45RA^-FoxP3^{bright}$ $CD4^+$ T cells), and activated conventional T cells ($CD4^+CD45RA^-FoxP3^{low}$) were similar between patients and controls (Figure 6D). Thus, IgA deficient patients display a circulating skewed $CD4^+$ T cell phenotype towards Th17 differentiation associated with an increase of serum sCD14, which is a marker of monocyte activation. Importantly, patients with malabsorption were excluded from the study since mucosal defects might grossly impact gut microbial ecology. We stratified patients in an effort to determine whether other clinical features such as infection, autoimmunity or historic antibiotic treatments (at least 3 months prior to sampling) might be preferentially associated with inflammation. We found no correlations between clinical status and any of the elevated immunological markers mentioned above. Altogether, the immunological and inflammatory modifications associated with SIgAd are not contributed by isolated patients presenting "extreme" clinical phenotypes.

IgA deficiency is associated with a disturbed bacterial dependency network

Bacterial symbiosis in the human gut notably implies that some bacteria depend on other bacteria for their persistence. Within such networks, and by definition, a dependent bacterium, called "satellite", never occurs independently of another, coined "host", in a given sample. Conversely, the same host may occur in a given sample independently of its satellites (Figure 7A). This concept was initially promoted by Nielsen *et al.*[32], who identified a minimal obligatory network of 45 MGS-MGS dependency-associations involving 60 MGS (the same MGS can make several associations). To investigate the potential impact of IgA deficiency on bacterial dependency associations, we tested whether this minimal obligatory network was disturbed in IgA deficiency.

Main links defined by Nielsen *et al.* were confirmed in healthy donors (90-100%), while the confirmed link percentage was more dispersed in IgAd patients ranging from 75% to 100% (Figure 7B). Based on our healthy donors MGS-MGS co-presence distribution, the 99% confidence interval was calculated according to a gamma distribution defining a lower-threshold of 84% of MGS-MGS co-presence, below which the population-level link was considered absent. In our healthy donor cohort, 41 links out of 45 were maintained at population level, while only 30 links were confirmed in IgAd patients. The 11 links lost in IgAd patients involved 21 MGS: 7 *Faecalibacterium* sp., 4 Firmicutes, 2 Clostridiales, 2 *Ruminococcus* sp., 2 *Prevotella*, 2 *Butyricimonas virosa* and 2 unknown MGS (Supplementary Data file S1).

The resulting network of remaining bacterial dependencies for healthy donors and IgAd patients is shown in Figure 7C. In summary, IgA deficiency is associated with a disturbed bacterial dependency association network.

Discussion

Similar to what was observed in murine models of IgAd[5], we found that IgA deficiency does not strikingly alter global fecal microbiota composition in affected patients. But our observations are limited to feces, and a more severe dysbiosis may be present in the small intestine, as shown in mouse models of IgA deficiency[5, 11-13]. The combination of flow cytometry and metagenomic analysis proved useful in order to increase the resolution of the analysis on human fecal samples by monitoring alterations of commensals specifically bound by secretory IgA in healthy controls. More work will be necessary to determine whether IgA^{dim} and IgA^{bright} bacterial populations (Figure 1D, left panel), correspond to bacteria bound by high versus low affinity IgA, respectively, and whether these subsets have overlapping bacterial repertoires or not. Of note, very low affinity IgA interactions are also suggested and could account for the very discrete, but global, bacterial flow cytometry profile shift consistently observed in healthy donors (Figure 1D, left panel), but not observed in SIgAd patients (Figure 1D, right panel).

It could have been expected that commensals bound by SIgA in healthy subjects would all tend to expand in IgAd patients. Indeed, in murine models of immune deficiency, IgA targets such as *SFB*[26] expand in the absence of an effective IgA response[4, 12]. We rather observed that MGSs defined in Figure 3B as main IgA targets in controls do not systematically bloom in IgAd patients. Most notably, *Coprococcus comes* (CAG 19) is under-represented in IgAd patients compared to controls (Figure 2A). In contrast, another typical IgA target such as *Escherichia coli* (CAG 4) is over-represented in IgAd patients by an order of magnitude, compared to controls (Figure 2A). These data suggest that *E. coli* (CAG 4) expansion in the microbiota could indeed be negatively influenced by SIgA, while *Coprococcus comes* thrives in the presence of SIgA. More generally, at family level, we observe that IgA^{+} bacteria are more likely found under-represented than over-represented in IgAd patients (Figure 2B and Figure 3E). These data emphasize the protective role that IgA directly (or indirectly) play in humans on commensal ecology, as previously observed in animal models[12].

Using the gene biomarker approach described in Qin *et al.*[33], on flow-sorted bacterial subsets we identified bacterial targets of secretory IgA in healthy donors by metagenomic analysis of SIgA-bound commensals. We show that IgA targets preferentially Firmicutes, Actinobacteria and

Proteobacteria, while Bacteroidetes are largely underrepresented compared to total microbiota composition. Only 1 out of the 24 MGSs identified as preferred IgA targets belong to the Bacteroidetes phylum, despite its dominance in the human colon. Our results are consistent with previous studies[12, 26, 37-40].

While preferred IgA targets belong to the Firmicutes phylum, under-represented bacteria in IgAd patients are also Firmicutes, such as *Faecalibacterium*, a genus well known to exert anti-inflammatory effects on the gut mucosa[41] and which is notably depleted in inflammatory bowel diseases[42, 43]. Conversely, potentially pro-inflammatory Gamma-proteobacteria and *Prevotella*[44] are over-represented in IgAd patients. Over-represented bacteria are more diverse than under-represented bacteria. Three MGSs usually belonging to the oral flora, *Veillonella parvula, Streptococcus sanguinis* and *Haemophilus parainfluenzae* are over-represented in gut microbiomes of IgAd patients. These bacteria are involved in biofilm formation and could exert inflammatory effects[45, 46]. *Veillonella parvula* has also been described as responsible of sepsis in an X-linked agammaglobulinemia patient[47]. Finally, the genera usually described as pathogenic in patients (*Salmonella* and *Campylobacter*) were not found over-represented.

Overall, our data show that, at steady state, IgAd patients present a quantitatively modest lower digestive tract dysbiosis, nevertheless consisting in a depletion of anti-inflammatory species along with an expansion of pro-inflammatory species and lower digestive track localization of constituents of the oral flora. Interestingly, it was recently shown that ectopic localization of human salivary microbiota can elicit severe gut inflammation in susceptible host animals[48]. It is therefore possible that ectopic oral microbiota could also exert a pro-inflammatory role in the context of IgAD.

Mechanisms underlying these findings are unknown but one could speculate that IgA, on the one hand, excludes[14, 49] and facilitates elimination of fast growing pathobionts[50]. On the other hand, SIgA could protect commensals by agglutination[51] and by localizing these bacteria in a favorable habitat like the mucus (Figure 8). Finally, IgA-mediated “intraluminal trapping” effects should prevent intimate bacterial proximity with the epithelial barrier, and therefore host inflammation and associated negative effects on commensals.

In IgA deficient patients, we show the microbiota remains significantly bound by IgM. This observation, together with the recent characterization of dually coated (IgA+ IgM+) mucus-embedded bacteria with increased richness compared to IgA-only-coated bacteria[52], suggest that IgM may compensate IgA deficiency. We show that microbial IgM-binding is highly variable between patients. When focusing the analysis on commensals bound by IgA in healthy subjects, or on differential genera between healthy controls and IgAd patients, we noticed that IgM, like IgA, tend to preferentially bind *Clostridium*, *Bifidobacterium* and *Faecalibacterium* (all Gram positive bacteria). However, we show that IgM does not bind *Enterobacteriaceae*, nor *Prevotellaceae* families, which are over-represented in IgA deficient patients and involved in pro-inflammatory events[44]. These data suggest that IgM are only partially efficient at controlling pathobionts, accounting for the susceptibility to enteropathogens in IgAd. Our data suggest that IgA has the same positive impact at least on Firmicutes and Actinobacteria diversity in humans; while IgM would only favor the latter. Of note, the very narrow range of IgM binding to Firmicutes, and to some extent Proteobacteria, might have precluded identification of potential correlations between IgM binding and diversity in these phyla.

In an effort to determine whether Actinobacteria diversity would also be lost in the absence of IgM, we studied CVID gut microbiomes. We observed that Actinobacteria diversity is drastically decreased in CVID, an "IgM-deficient" model. Moreover, a similar trend is observed for Firmicutes. These results were confirmed in an independent CVID study. Jorgensen *et al.* recently showed that CVID patients display dysbiotic gut microbiota with reduced alpha-diversity, reduced abundance of Actinobacteria and increased abundance of gamma-proteobacteria[53]. Re-analysis of the same data further underlined that diversity loss was particularly pronounced in Actinobacteria, but also observed in other phyla (Jorgensen et al. personal communication). It should be underlined that although we took care to include CVID that did not receive antibiotics within 3 months prior to sampling, the analysis of CVID patients is potentially hampered by several confounders such as previous antibiotic courses and other treatments. It is however interesting to note that Actinobacteria diversity is preferentially decreased in two independent CVID studies, while other phyla are comparatively less affected in this "IgM-deficient" model, while, allegedly, confounding factors might have similarly affected

all phyla. There is clinical relevance for this concept as it is well established that patients lacking selectively IgA rarely develop IBD, while this complication is instead more frequent and severe in those lacking both IgA and IgM[54].

The impact of IgA deficiency on bacterial symbiosis is further evidenced by the perturbation of bacterial networks (Figure 7C). The presence in IgAd patients of satellite bacteria (such as CAG 97, CAG 133, CAG 488, CAG 206 and CAG 328), in the absence of their previously described host[32], most probably reflect the establishment of novel dependency links. Such modifications underscore the role played by IgA well beyond its recognized neutralizing activity.

IgAd has also profound systemic repercussions. In an attempt to explain the Th17 bias observed in the patients we compared aryl hydrocarbon receptor (AHR) agonists[55] in fecal water and sera from IgAd patients and controls, but observed no differences (data not shown). The Th17 bias we observed is nevertheless potentially connected to intestinal dysbiosis since Klemola *et al.*[56] have already shown the presence of activated T cells in gut mucosa of IgA deficient patients. As described by Perreau *et al.*[57], in CVID, $CD4^{+}PD\text{-}1^{+}$ cells are increased in selective IgA deficient patients. Although LPS was not found elevated in SIgAD serum, increased soluble CD14 levels and PD-1 up-regulation could reflect T-cell exhaustion induced by repeated bacterial translocation[58]. Finally, and from a therapeutic perspective, a much larger cohort of patients would be needed in order to extract a potentially beneficial microbial signature associated with elevated IL-10 levels observed in some patients.

Altogether, we show for the first time that selective IgA deficiency in humans is associated with a mild intestinal dysbiosis, characterized by expansion of pro-inflammatory bacteria, depletion of anti-inflammatory commensals, and a perturbation in the "obligatory" bacterial network. Dysbiosis could be partly explained by the fact that IgA deficiency is not totally compensated by IgM secretion, and by the loss of a non-redundant chaperon-like effect of IgA on microbial diversity.

Materials and Methods

Study design

Patients and controls: We conducted a cross-sectional study of patients with IgA deficiency compared to healthy controls. Fresh stool and blood samples were collected simultaneously at a single time point in 21 patients with selective IgA deficiency (Table S1) and compared with 34 age- and sex-matched healthy donors (Table 1). We furthermore recruited 7 CVID patients with IgA, IgG and/or IgM deficiency thus displaying a global antibody production defect. Patients were recruited from two French referral centers of clinical immunology (Department of Clinical Immunology in Saint Louis hospital and Department of Internal Medecine in Pitié Salpêtrière hospital, Paris), where they are followed for clinical manifestations associated with antibody deficiencies. IgA deficient patient's inclusion criteria was undetectable seric IgA levels (<0,07 mg/mL) in at least three previous samples in the past year. SIgAd is defined by serological means, namely undetectable seric IgA titers (<0.07 mg/mL) with normal IgG levels. CVID patients are characterized by a marked decrease of seric IgG (at least 2 SD below the age-dependent mean) and a marked decrease in at least one of the isotypes IgM or IgA. For our study we furthermore requested that the patients should be deficient in seric IgA (<0.07 mg/ml). Exclusion criteria: antibiotics therapy and laxative drugs use in the last three months prior to stool collection (inclusion for metagenomic analysis: 17 SIgAd patients, 7 CVID, 34 healthy donors). Clinical and biological data were collected at inclusion time. Oral and written consent were obtained from patients before inclusion in the study.

PBMC and plasma

30 mL of blood were collected in ACD tubes (BD Vacutainer®) and PBMC were isolated by density gradient procedure (Ficoll 400, Eurobio) and then stored in liquid nitrogen. Supernatants were collected as plasma and immediately stored at -80°C.

Bacterial Flow Cytometry

Bacterial species-specific antibody levels against culturable strains were assessed by a flow cytometry assay described in Moor *et al.*[29]. Briefly, 10^5 paraformaldehyde (PFA)-fixed bacteria (4% (wt/v) PFA in PBS) were incubated for 25 min at 4°C in serum diluted in bacterial flow cytometry buffer (1x-PBS, 2% wt/v BSA, 0.02% w/v Sodium azide) at a final IgG concentration

of 10µg/mL in a 96-well V-bottom plate. Cells were then washed twice with 200µL/well of bacterial flow cytometry buffer (4000g, 10min, 4°C) and then incubated for 20 min at 4°C with appropriate secondary staining reagent (isotype controls, goat anti-human IgA-FITC, donkey anti-human IgM-PE, goat anti-human IgG-A647, all from Jackson Immunoresearch Laboratories), then washed twice as above, before re-suspension in bacterial flow cytometry buffer. Acquisition of cells events was performed using a FACS CANTO II (Becton Dickinson) and analysis was performed with Flow-Jo software (Treestar). Medians of fluorescence were used to measure bacterial species-specific serum antibody levels.

Quantification of in vivo antibody-binding levels to commensals

Purified microbiota (10^7 bacteria/condition) was thawed and then fixed in 500 µL of 4% (wt/v) PFA in PBS including a cell proliferation dye (eFluor 450, eBiosciences, CA, USA) for 20 min at 4°C. After washing with 1mL PBS (4000g, 10min, 4°C), microbiota was resuspended in bacterial flow cytometry buffer in order to obtain a final concentration of $4x10^8$ bacteria/mL. 10^7 bacteria/well are then incubated in a 96-V bottom well plate with appropriate secondary staining reagents, washed and analyzed as above.

Cytokines quantification

Serum IL-6, IL-10 and IL-17 were measured using Single Molecule Array (Simoa) technology[59]. Working dilutions were 1/4 for all sera, in working volumes of 25µL. Measures were based on 3-step digital assay (see supplementary methods) using the HD-1 Analyzer (Quanterix Corporation).

Soluble CD14 quantification

Soluble CD14 was quantified in serum by ELISA (Quantikine® ELISA kit, R&D). Experimental procedure was performed according to the manufacturer's recommendations.

LPS quantification

LPS concentration was determined by direct quantitation of 3-hydroxytetradecanoic acid (or 3HM) by LC/MS/MS as we have previously described[60]. Briefly, 50 µl of plasma spiked with 50 ng of internal standard (3β-hydroxytridecanoic acid 10 mg/ml in ethanol) was hydrolysed (total

3HM) or not (unesterified 3HM) with 75 µl of NaCl 150mM and 300 µl of HCl 8M for 4 h at 90°C. LPS-derived/esterified 3HM was calculated as the difference between total 3HM (as assessed after HCl hydrolysis) and unesterified/free 3HM (as assessed without HCl hydrolysis).

Peripheral blood mononuclear cell phenotyping

B cell phenotyping was performed using the following markers: CD19-BV510, CD27-APC, IgG-PE-Cy7, IgM-PerCP Cy5.5, IgD-BV421 (all BD Biosciences), IgA-PE (Jackson Immunoresearch). Conventional T cell phenotyping was performed using CD3-H500, CCR7-PE-Cy7, CD4-APC-Cy7 (all BD Biosciences), TCRγδ-PE (Beckman Coulter), CD45RA-PercP Cy5.5 (e-Bioscience), CD8-A405 (Invitrogen), CD279-APC (BioLegend). Regulatory T cell phenotyping was performed according to Miyara *et al.*[61, 62] with the following markers: CD4-PerCP, CD15s-A647, CD25-PE-Cy7, CD45RA-APC-H7, CD127-BV421, Ki67-FITC, and FoxP3-PE (all BD Biosciences). FoxP3 and Ki67 were stained after cell permeabilization with the FoxP3 Fixation/Permeabilization buffer set (eBiosciences, cat#00-5521). Cytokines secreting T cells were detected after a 6-hour Phorbol 12-Myristate 13-Acetate (PMA) - Ionomycin calcium salt (Sigma Aldrich) stimulation and Golgi blocking (Golgi stop, BD Biosciences) and stained with: CD4-APC-Cy7 (BD Biosciences), CD3-PerCP (BD Biosciences), CD8-A405 (Invitrogen), Aqua (LIVE/DEAD® Fixable Dead Cell Stain Kit, Invitrogen), IL-22-A647 (e-Bioscience), IFN-γ-FITC (BD Biosciences), IL-17A-PE (eBioscience).

Stool collection and whole microbiota purification

Stool were collected immediately after emission in a container allowing anaerobic bacteria preservation (Anaerocult band, Mikrobiologie), aliquoted in a CO_2 rich-0_2 low atmosphere and stored at -80°C. Fecal bacterial were extracted by gradient purification in anaerobic conditions (Freter chamber) as previously described[63]. Briefly, thawed feces were diluted in 1x-PBS (Eurobio), 0,03% w/v sodium deoxycholate (NaDC). An aliquot was centrifuged twice (+21000g, 10 min) and supernatant, so called fecal water, was collected and stored at -80°C. The remaining material was diluted in 1xPBS, 60% w/v Nycodenz, 0,03 % w/v NaDC, and loaded on a continuous density gradient obtained by a freezing-thawing cycle of a Nycodenz solution. Bacterial extract was obtained after ultracentrifugation (9000 rpm, 45 min, +4°C) (Beckman Coulter ultracentrifuge, swinging rotor SW28), and washed three times in 1x-PBS (Eurobio),

0,03% w/v sodium NaDC. The final pellet was diluted in 1xPBS-10%Glycerol, immediately frozen in liquid nitrogen and then stored at -80°C(*1*).

Free intestinal Ig titers

IgA, IgM and IgG levels were measured by enzyme linked immunosorbent assay (ELISA) in the fecal water samples obtained as described above. The assays were performed in a 96-flat bottom plate (Maxisorp, Mume Immunoplate) according to the manufacturer's recommendations (Human Ig ELISA Quantification Set, Bethyl Laboratories). Immune complexes were revealed by incubation with the 3,3', 5,5' Tetramethylbenzidine (TMB, Rockland) and the reaction was stopped by a 1M HCl solution. Plates were read at 450 nm (Multiscan Ex Thermofisher Saint Herblain). Final concentrations were calculated using the 4-Parameter Logistic curve equation of the standard dilutions normalized to the feces weight.

Enrichment of SIgA-bound intestinal microbes

Gradient purified gut microbiota (10^{10} bacteria per sample) were thawed and washed twice in 1ml of 1xPBS. Bacteria were enriched using MACS cell separation. Bacteria were incubated with PE-conjugated anti-human IgA (Jackson Immunoresearch Laboratories) for 20 minutes in the dark at 4°C, washed twice and then incubated with anti-PE MicroBeads (Miltenyi, Paris, France) for 20 minutes in the dark at 4°C. Labeled bacteria were then enriched on two consecutive MACS columns. First flow-through (IgA^- fraction) and final eluted fraction (IgA^+ fraction) were collected and cryopreserved as dry pellet at -80°C. Purity of sorted bacterial fractions was systematically verified by flow cytometric analysis.

Fluorescence-activated cell sorting of IgM-bound gut microbes

Purified microbiota (10^9/condition) was thawed, washed in 1x-PBS and stained with isotype control antibody (R-phycoerythrin-conjugated Donkey IgG, Jackson Immunoresearch Laboratories,) as a negative control, and the anti-human IgM-PE (Jackson Immunoresearch Laboratories). Acquisition and sorting were performed on a 2 lasers- 2 ways Fluorescent-activated cell sorter (S3 cell sorter, Bio-Rad Laboratories, California, USA) with a minimum rate of 80% of purity for both fractions. Purity of the fractions was systematically verified after sorting. $1x10^6$ bacteria per fraction were collected and immediately stored at -80°C.

DNA extraction

Genomic DNA was extracted using two distinct but overlapping protocols for whole stool and sorted microbiota. Whole stool samples were treated as previously described[64]. Briefly, 200 mg of faecal sample was lysed chemically (guanidine thiocyanate and N-lauroyl sarcosine) and mechanically (glass beads) followed by elimination of cell debris by centrifugation and precipitation of genomic DNA. Finally, genomic DNA was RNase treated. DNA concentration and molecular size were estimated by Nanodrop (Thermo Scientific) and agarose gel electrophoresis. Sorted commensal samples were treated similar to whole stool samples with the exception that DNA precipitation was performed in smaller volumes and with extra-long incubation times due to the scarcity of bacterial content.

Whole genome metagenomic library construction

Libraries were constructed according to manufactures protocol (Life Technologies). Briefly, extracted genomic DNA was sheared by sonication, size-exclusion purified by Agencourt beads (Beckman Coulter), ligated to P1 and P2 adaptor oligonucleotides with appropriate barcodes, PCR amplified and loaded onto the flow-chip for downstream SOLiD sequencing.

DNA sequencing and Gene abundances profiling

Microbiota gene content was determined by high-throughput SOLiD sequencing of total and IgA-enriched faecal DNA[65]. An average of 70.5 million ± 23.6 million (mean ± s.d.) and 54.4 million ± 20.8 million 35-base-long high-quality single reads were determined for each whole stool and IgA-enriched sample, respectively (a total of 7.7 Gb of sequence). 119 metagenomic samples (34 healthy donors, 17 selective IgA deficient patients, 7 CVID patients, 30 healthy IgA plus fraction and 30 healthy IgA minus fraction) were analyzed. High-quality reads were mapped to the reference 3.9 million gene catalogue[32] using bowtie package[66] with a maximum of 3 mismatches and length above 35bp. Reads mapping at multiple positions were discarded. An average of 21.0 million ± 9.5 million and 16.0 million ± 6.8 million uniquely mapped reads per whole stool and IgA-enriched sample, respectively, were retained for estimating the abundance of each reference gene by using METEOR software. Abundance of each gene in an individual was normalized with the method coined Reads Per Kilobase per Million (RPKM) as previously described[67]. Briefly, gene abundance was determined as the number of reads that uniquely

mapped to a defined gene. Subsequently, normalized gene abundances were transformed in frequencies by dividing them by the total number of uniquely mapped reads for a given sample. The resulting microbial gene profile was used for further analyses. For samples where 6 million or more sequence reads were obtained (*n*=58 (none sorted data), n=56 (sorted data)), 6 M sequence reads were drawn randomly (without replacement). These randomly drawn reads were mapped to the 3.9 million genes catalog and the number of reads counted to form a downsized depth or abundance matrix. The 6 million downsized depth matrix was used to estimate gene abundances and MGS richness.

Bioinformatical analysis

CAGs matrix construction: The projection of genes into co-abundance gene groups was performed using a pre-defined list of 7,381 CAGs[32]. Every pre-defined CAG is a vector of genes ordered by increasing connectivity. For a bacterial genome, the most connected genes correspond to the marker genes present in all individuals carrying this organism (core genome). Starting with the gene frequency matrix, we projected the list of connected genes in each pre-defined CAG and extract the corresponding frequency profile of the 50 most connected ones. Provided that at least 10% of marker genes are found, each frequency profile will be used to compute the mean vector corresponding to the CAG frequency.

Differential co-abundance gene groups identification: Genes from the gene-profile matrix were used to identify those that were differentially abundant between the SIgAd patients and healthy donors groups as described in[30]. Briefly, Wilcoxon tests were employed to compute the probabilities that genes frequency profiles did not differ between SIgAd patients and healthy donors groups by chance alone. Benjamini and Hochberg multiple test correction was applied to the P-values. By performing a selection based on a significance threshold of $P < 0.01$, we identified 18025 genes that were differentially abundant between the two groups. The same method and P-value were applied to identify the differentially abundant genes between the IgA^+/IgA^- fractions in healthy donors. We identified 110558 genes that were differentially abundant between the two groups. In a second step, the differentially abundant genes were projected into co-abundance gene groups (CAGs) defined in the 3.9 million genes catalogue. To validate a differential CAG, the threshold used was a minimum of 50 differentially abundant

genes per CAG (CAG are kept only if they are equal or exceed 50 connected genes) including at least 10% of marker genes. The frequency profile of the 50 most connected genes was used to compute the mean vector corresponding to the CAG frequency. Of the 18025 differential genes between SIgAd patients and healthy donors, 8191 fell into 33 CAGs composed of 53–997 genes after the projection step (Supplementary Figure S2B and supplementary Data file S2). 31 of the 33 CAGs are considered as Metagenomic species (MGS) according to their gene size (>=700 genes). For the IgA^+ and IgA^- fractions comparison, 80415 out of 110558 differential genes fell into 119 CAGs of 50–3105 genes. 30 CAGs are over-represented in the IgA^+ fraction and 89 over-represented in the IgA^- fraction. 24 of the 30 IgA^+ CAGs are considered MGS according to their gene size (Supplementary Figure S3 and supplementary Data file S3).

Wilcoxon tests were employed to compute the probabilities that CAGs frequency profiles did not differ between the two compared groups (SIgAd/HD, IgA^-/IgA^+ fractions) by chance alone. Benjamini and Hochberg multiple test correction was applied to the P-values. The CAG taxonomical annotation was performed as previously described[68].

Dependency associations: Dependency associations between CAGs were defined as described previously[32]. This analysis relies on the detection of CAGs that are systematically associated with other CAGs in n=396 healthy donors. Briefly, they consider dependencies between CAGs based on their sample-wise overlapping detection. First, they used a Fisher exact test to identify statistically significant CAGs detection overlap. Secondly, they validated the dependencies only when the dependent CAG were systematically observed with their associated host CAG. Thus, a dependent CAG, called satellite CAG, should never occur independently of the host CAG. Using these criteria, they identified 886 dependencies, most of them involving a small CAG as a satellite, corresponding to phage/bacteria or pan-genome/core-genome relationship. They also detected 45 MGS-MGS dependency-associations, involving MGS as satellite and host microorganisms that may reflect bacterial symbiosis events. In this study, we have considered the 45 MGS-MGS dependency-associations detected by Nielsen *et al.*[32] as a minimal obligatory network that should be retrieved in healthy donors and we hypothesized that it could be disturbed in IgA deficiency. The presence of the 45 links was subsequently investigated in our cohort, and for each link, we have determined the frequency of its presence in both groups. Based on our

healthy donors frequencies distribution, we have determined the 99% confidence interval according to the gamma distribution law defining a threshold of frequency of 84% below which the link was considered significantly decreased at the population level. The investigation of the 45 links presence-absence profiles in our cohort allowed us to distinguish four possible situations. For each link and each individual, if the MGS satellite is present in the absence of the host MGS we consider the link as lost. If the MGS satellite is present in the presence of the host MGS we consider the link as confirmed. If the MGS host is present in the absence of the satellite MGS or if the two are absent, the situation is not taken into account (no information about the link).

16S rRNA metagenomic analysis of sorted IgM-bound gut microbiota

Microbiota composition was analysed by pyrosequencing of the V3–V4 region of 16S rRNA gene[69]. The 16S rRNA gene was amplified with a semi-nested PCR using primers V3fwd (+357): 5' TACGGRAGGCAGCAG 3', V4rev (+857): 5' ATCTTACCAGGGTATCTAATCCT 3' and X926_Rev (+926) 5' CCGTCAATTCMTTTRAGT 3'. Polymerase chain reaction amplicon libraries were sequenced using a MiSeq Ilumina platform (Genotoul, Toulouse, France). The resulting sequences were analyzed using the open source software package Quantitative Insights Into Microbial Ecology (QIIME)[70].

Following removal of the primers and barcodes, sequences were filtered as follows: (i) minimum and maximum read length of 250 bp and 500 bp respectively, (ii) no ambiguous base calls, (iii) no homopolymeric runs longer than 8 bp and (iv) minimum average Phred score>27 within a sliding window of 50 bp. Sequences were aligned with NAST against the GreenGenes reference core alignment set (available in QIIME as core_set_aligned.fasta.imputed) using the 'align_seqs.py' script in QIIME. Sequences that did not cover this region at a percent identity>75% were removed. Operational taxonomic units were picked at a threshold of 97% similarity using cd-hit[71] from 'pick_otus.py' script in QIIME. Picking workflow in QIIME with the cd-hit clustering method currently involves collapsing identical reads using the longest sequence-first list removal algorithm, picking OTU and subsequently inflating the identical reads to recapture abundance information about the initial sequences. Singletons were removed, as only OTU that were present at the level of at least two reads in more than one sample were

retained (9413±5253 sequences per sample). The most abundant member of each OTU was selected through the 'pick_rep_set.py' script as the representative sequence. The resulting OTU representative sequences were assigned to different taxonomic levels (from phylum to genus) using the GreenGenes database (release August 2012), with consensus annotation from the Ribosomal Database Project naïve Bayesian classifier [RDP 10 database, version 6[72]]. To confirm the annotation, the resulting OTU representative sequences were then searched against the RDP database, using the online program seqmatch (http://rdp.cme.msu.edu/seqmatch/seqmatch_intro.jsp) and a threshold setting of 90% to assign a genus to each sequence.

Statistical Analysis

Log2 ratio of IgM^+/IgM^- bacterial abundances was performed at the lowest taxonomic identification level (genus or family) to analyze IgM^+ relative binding. Only dominant taxa were taken into account, *i.e.:* taxa that are present in at least 50% of the patients. Zero was normalized in pairs and adjusted to the lowest abundance of all taxa in the paired samples.

Flow cytometry analysis was performed using Flow Jo (9.3.2) Treestar® software and datamined with Funky Cells ToolBox software (Version 0.1.2, www.FunkyCells.com).. Statistical non parametric tests were used whenever necessary: Mann-Whitney was used when comparing two groups, Kruskall Wallis with multiple comparison post-test of Dunn's when comparing three groups or more, Fisher's exact test was used for contingency, Wilcoxon paired rank test for paired analysis, and Spearman coefficient for correlations. R v3.2.1 and Graphpad Prism® version 6 were used to perform statistical analysis.

Supplementary Materials

Materials and Methods

Table S1: Individual clinical description of selective IgA deficient patients.

Figure S1: Serum IgA and IgG levels in IgA deficient patients.

Figure S2A: Global gut microbiota composition in IgA deficient patients.

Figure S2B: Differential CAG abundance analysis between healthy subjects and IgA deficient patients.

Figure S2C: Longitudinal stability of gut microbiota composition in 3 healthy subjects.

Figure S3: Differential CAG abundance analysis between IgA^+ and IgA^- microbes.

Figure S4: Gut microbiota IgA-binding patterns.

Figure S5A: Serum IgM responses against *Bifidobacterium adolescentis and Bifidobacterium longum.*

Figure S5B: Free fecal IgA and IgM levels in CVID patients.

Figure S5C: Gut microbiota IgA and IgM binding in CVID patients.

Data file S1: MGS-MGS dependency network confirmation in IgA deficient patients.

Data file S2: 33 CAGs differentiating healthy donors and IgA deficient patients.

Data file S3: 119 CAGs differentiating gut microbiota *in vivo* bound by sIgA or not in healthy donors.

Acknowledgments: The authors wish to thank Cyrielle Fougeroux for aiding initial setup of immune phenotyping of lymphocyte and stool samples. **Funding**: The study was financed by the following funding bodies: Institut national de la santé et de la recherche médicale (Inserm), Agence Nationale de la Recherche (MetAntibody, ANR-14-CE14-0013), Fondation pour l'Aide a la Recherche sur la Sclerose En Plaques (ARSEP) and UPMC Emergence (MycELIA). **Author contributions**: J.F, L.G., D.B., A.M., M.M., E.O., Z.A. and C.F. recruited patients and collected clinical data. J.F., D.S., D.G., C.J. and M.L. prepared biospecimens (aliquoting and purification of live stool commensals) and cryopreserved them to generate biobank. J.F., D.S. C.P., C.J. and M.L. enriched Ig-bound commensals. F.L., S.K., N.G. and N.P. generated whole genome metagenomic data. G.A., M.L. and P.L. generated 16S data. J.F., C.P., D.G., D.S. and M.L. performed flow cytometric analysis of patient lymphocytes and commensals. J.F., D.S. and K.D. measured cytokine and sCD14 levels in donor serum. JPPB and LL measured LPS levels in donor serum. J.F, H.E., M.A., E.L.C. and M.L performed data mining and biostatistical analysis. J.F., H.E., M.L and G.G. designed the study, prepared the figures and wrote the manuscript. S.D.E., J.D., P.L. and C.J. reviewed the manuscript. S.D.E., J.D. and C.J. provided support for the design of the study. **Competing interests**: The authors declare no competing interests **Data and materials availability**: Data will be available online upon acceptance.

References:

1. M. J. Molloy *et al.*, *Cell Host Microbe* **14**, 318 (Sep 11, 2013).
2. S. Hapfelmeier *et al.*, *Science* **328**, 1705 (Jun 25, 2010).
3. L. V. Hooper *et al.*, *Nat Rev Immunol* **10**, 159 (Mar, 2010).
4. K. Suzuki *et al.*, *Proc Natl Acad Sci U S A* **101**, 1981 (Feb 17, 2004).
5. S. Fagarasan *et al.*, *Science* **298**, 1424 (Nov 15, 2002).
6. G. R. Harriman *et al.*, *J Immunol* **162**, 2521 (Mar 1, 1999).
7. D. H. Reikvam *et al.*, *Eur J Immunol* **42**, 2959 (Nov, 2012).
8. O. L. Wijburg *et al.*, *J Exp Med* **203**, 21 (Jan 23, 2006).
9. N. Lycke *et al.*, *J Immunol* **163**, 913 (Jul 15, 1999).
10. M. Wei *et al.*, *Nat Immunol* **12**, 264 (Mar, 2011).
11. S. Kawamoto *et al.*, *Science* **336**, 485 (Apr 27, 2012).
12. S. Kawamoto *et al.*, *Immunity* **41**, 152 (Jul 17, 2014).
13. J. J. Bunker *et al.*, *Immunity* **43**, 541 (Sep 15, 2015).
14. C. R. Stokes *et al.*, *Nature* **255**, 745 (Jun 26, 1975).
15. L. Yel, *J Clin Immunol* **30**, 10 (Jan, 2010).
16. N. Wang *et al.*, *Curr Opin Allergy Clin Immunol* **12**, 602 (Dec, 2012).
17. G. H. Jorgensen *et al.*, *J Clin Immunol* **33**, 742 (May, 2013).
18. S. Koskinen, *J Clin Immunol* **16**, 165 (May, 1996).
19. A. Aghamohammadi *et al.*, *J Clin Immunol* **29**, 130 (Jan, 2009).
20. J. F. Ludvigsson *et al.*, *J Clin Immunol* **34**, 444 (May, 2014).
21. L. Mellander *et al.*, *J Clin Immunol* **6**, 284 (Jul, 1986).
22. P. Brandtzaeg *et al.*, *Clin Exp Immunol* **67**, 626 (Mar, 1987).
23. W. B. Chodirker *et al.*, *Science* **142**, 1080 (Nov 22, 1963).
24. D. A. Peterson *et al.*, *Cell Host Microbe* **2**, 328 (Nov 15, 2007).
25. T. C. Cullender *et al.*, *Cell Host Microbe* **14**, 571 (Nov 13, 2013).
26. N. W. Palm *et al.*, *Cell* **158**, 1000 (Aug 28, 2014).
27. N. J. Mantis *et al.*, *Mucosal Immunol* **4**, 603 (Nov, 2011).
28. W. Al-Herz *et al.*, *Front Immunol* **2**, 54 (2011).
29. K. Moor *et al.*, *Nat Protoc* **11**, 1531 (Aug, 2016).
30. J. Qin *et al.*, *Nature* **464**, 59 (Mar 4, 2010).
31. M. Arumugam *et al.*, *Nature* **473**, 174 (May 12, 2011).
32. H. B. Nielsen *et al.*, *Nat Biotechnol*, (Jul 6, 2014).
33. N. Qin *et al.*, *Nature*, (Jul 23, 2014).
34. F. Plaza Onate *et al.*, *BMC Genomics* **16**, 1406 (Dec, 2015).
35. P. Brandtzaeg *et al.*, *Science* **160**, 789 (May 17, 1968).
36. U. Klein *et al.*, *Blood* **89**, 1288 (Feb 15, 1997).
37. A. L. Kau *et al.*, *Sci Transl Med* **7**, 276ra24 (Feb 25, 2015).
38. J. D. Planer *et al.*, *Nature* **534**, 263 (Jun 9, 2016).
39. G. D'Auria *et al.*, *Sci Rep* **3**, 3515 (2013).
40. M. Dzidic *et al.*, *J Allergy Clin Immunol*, (Aug 13, 2016).
41. H. Sokol *et al.*, *Proc Natl Acad Sci U S A* **105**, 16731 (Oct 28, 2008).
42. K. Machiels *et al.*, *Gut* **63**, 1275 (Aug, 2014).
43. H. Sokol *et al.*, *Inflamm Bowel Dis* **15**, 1183 (Aug, 2009).
44. E. Elinav *et al.*, *Cell* **145**, 745 (May 27, 2011).
45. B. van den Bogert *et al.*, *PLoS One* **9**, e114277 (2014).
46. I. Mashima *et al.*, *Anaerobe* **28**, 54 (Aug, 2014).
47. M. Strach *et al.*, *J Clin Microbiol* **44**, 2655 (Jul, 2006).
48. K. Atarashi *et al.*, *Science* **358**, 359 (Oct 20, 2017).
49. K. Moor *et al.*, *Nature* **544**, 498 (Apr 27, 2017).
50. S. Okai *et al.*, *Nat Microbiol* **1**, 16103 (Jul 04, 2016).
51. A. Mathias *et al.*, *J Biol Chem* **286**, 17239 (May 13, 2011).
52. G. Magri *et al.*, *Immunity* **47**, 118 (Jul 18, 2017).
53. S. F. Jorgensen *et al.*, *Mucosal Immunol* **9**, 1455 (Nov, 2016).
54. S. Agarwal *et al.*, *J Allergy Clin Immunol* **124**, 658 (Oct, 2009).
55. M. Veldhoen *et al.*, *Nature* **453**, 106 (May 1, 2008).

56. T. Klemola *et al.*, *Gut* **37**, 519 (Oct, 1995).
57. M. Perreau *et al.*, *J Exp Med* **211**, 2033 (Sep 22, 2014).
58. C. Chenevier-Gobeaux *et al.*, *Clin Chim Acta* **450**, 97 (Oct 23, 2015).
59. D. M. Rissin *et al.*, *Nat Biotechnol* **28**, 595 (Jun, 2010).
60. J. P. Pais de Barros *et al.*, *J Lipid Res* **56**, 1363 (Jul, 2015).
61. M. Miyara *et al.*, *Immunity* **30**, 899 (Jun 19, 2009).
62. M. Miyara *et al.*, *Proc Natl Acad Sci U S A* **112**, 7225 (Jun 09, 2015).
63. C. Juste *et al.*, *Gut*, (Jan 16, 2014).
64. J. J. Godon *et al.*, *Appl Environ Microbiol* **63**, 2802 (Jul, 1997).
65. E. R. Mardis, *Trends Genet* **24**, 133 (Mar, 2008).
66. B. Langmead, *Curr Protoc Bioinformatics* **Chapter 11**, Unit 11 7 (Dec, 2010).
67. M. A. Dillies *et al.*, *Brief Bioinform* **14**, 671 (Nov, 2013).
68. E. Le Chatelier *et al.*, *Nature* **500**, 541 (Aug 29, 2013).
69. K. H. Wilson *et al.*, *J Clin Microbiol* **28**, 1942 (Sep, 1990).
70. J. G. Caporaso *et al.*, *Nat Methods* **7**, 335 (May, 2010).
71. W. Li *et al.*, *Bioinformatics* **22**, 1658 (Jul 01, 2006).
72. J. R. Cole *et al.*, *Nucleic Acids Res* **37**, D141 (Jan, 2009).

Table 1. Demographic and clinical cohort summary.

	Healthy donors	Selective IgA deficiency	P-value*
Number of individuals	34	21	
Median age	32.9 [23-61]	36 [18-67]	0.1289
Sex Ratio F:M	18:16	13:8	0.4239
IMC (index de masse corporelle)	21.9 [18.7 – 33.9]	22 [19.1-36]	0.3599
Ethnic origin**			0.1149
Caucasian	23 (68%)	18 (86%)	
North Africa	4 (11%)	3 (14%)	
Africa	1 (3%)	0 (0%)	
Middle eastern	4 (11%)	0 (0%)	
Asiatic	2 (5%)	0 (0%)	
Recurrent infections		**9 (43%)**	
Upper and lower respiratory tract		7 (33%)	
Intestinal		3(14%)	
Vaginal		6/14 (42%)	
Antibiotics (>1 per year)***		**12 (57%)**	
Auto immune condition		**11 (52%)**	
Cytopenias		5 (24%)	
SLE		4 (19%)	
Thyroiditis		4 (19%)	
Celiac disease		2(10%)	
Biermer anemia		1(5%)	
Vitiligo		2(10%)	
Type 1 diabetes		1(5%)	
Ankylosing spondylitis		1(5%)	
Intestinal symptoms (chronic diarrhea and/or chronic abdominal pain leading to digestive endoscopy)		**9(43%)**	
Immunomodulatory therapy		**3 (14%)**	
Steroids****		3 (14%)	
Methotrexate		1 (5%)	
Hydroxychloroquine		3 (14%)	
Associated IgG4 subclass deficiency		**2 (10%)**	
Median IgG levels (mg/mL)	10.10 [6.7-15.2]	11.8 [7.14-18]	0.2475

* Continuous and discrete variables were assessed statistically with Mann-Whitney and chi-squared test for homogeneity, respectively.

** All donors domesticated in France.

*** 4 IgA deficient patients having received antibiotics within 3 months from sampling were excluded for gut microbiota analysis, but were included in the immunological phenotype analysis (cf. figure 6).

**** Median dose 10mg/day

Figures:

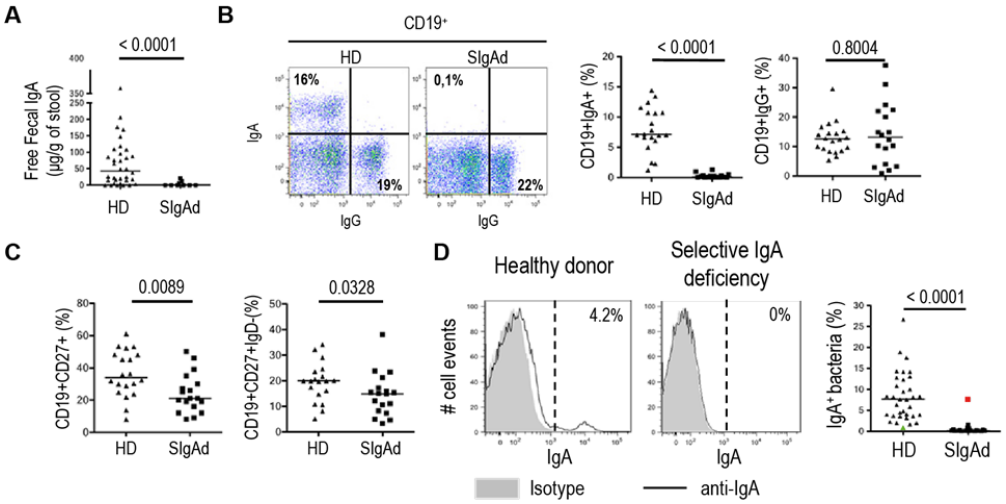

Figure 1: IgA deficient patients

(A) Free IgA levels in fecal water were measured by ELISA in healthy donors (HD, n=34) and selective IgA deficient patients (SIgAd, n=21). (B) $CD19^{+}IgA^{+}$ and $CD19^{+}IgG^{+}$ (surface staining) cells were detected by flow cytometry as shown in the FACS plot example as well as (C) peripheral $CD19^{+}CD27^{+}$ (memory B cells) and $CD19^{+}CD27^{+}IgD^{-}$ (switched memory B cells) cells. (D) SIgA-binding levels of *ex-vivo* purified and fixed microbiota are measured by a flow cytometry based assay. Grey histograms represent isotype control and black line anti-IgA stained microbiota. In the scatter dot plot, each symbol represents a donor (34 HD and 21 SIgAd). In all sections, horizontal bars represent medians. Indicated p-values were calculated using a Mann Whitney test. A healthy donor with non-detectable IgA binding of gut microbiota and an IgA deficient patient with detectable IgA binding of gut microbiota are indicated with a green and red symbol, respectively.

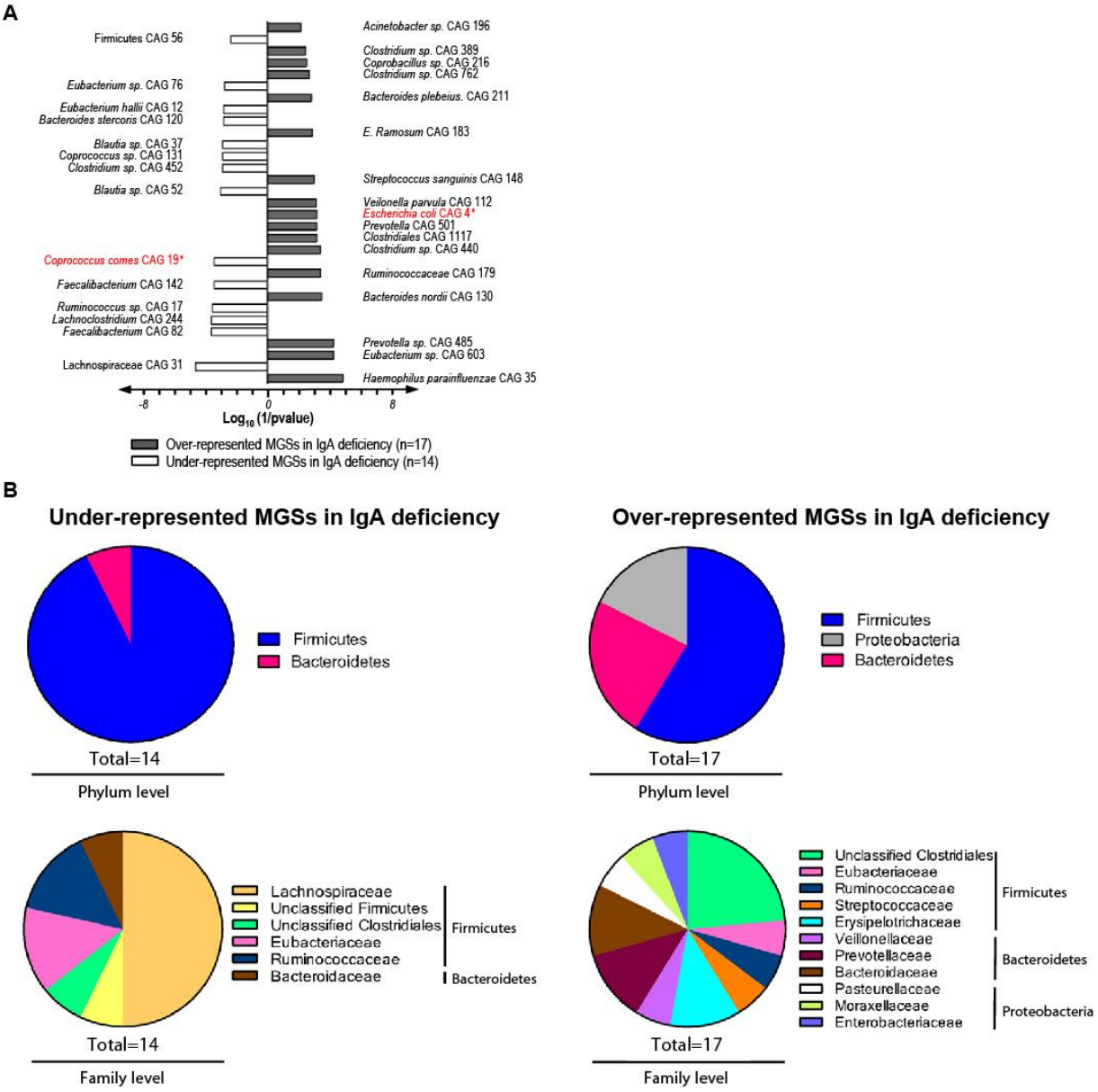

Figure 2: Bacterial repertoire shift associated with IgA deficiency.

(A) Differential MGS (n=31) between IgA deficient patients and healthy donors assigned at their lowest taxonomic level and ranked by statistical difference. White histograms represent MGS that are underrepresented in IgA deficiency (n=14) whereas grey histograms represent MGS that are overrepresented in IgA deficiency (n=17), compared to healthy donors. (B) Taxonomic distribution at phylum (first row) and family (second row) level of under-represented MGS (left column) and over-represented (right column) MGS in IgA deficiency.

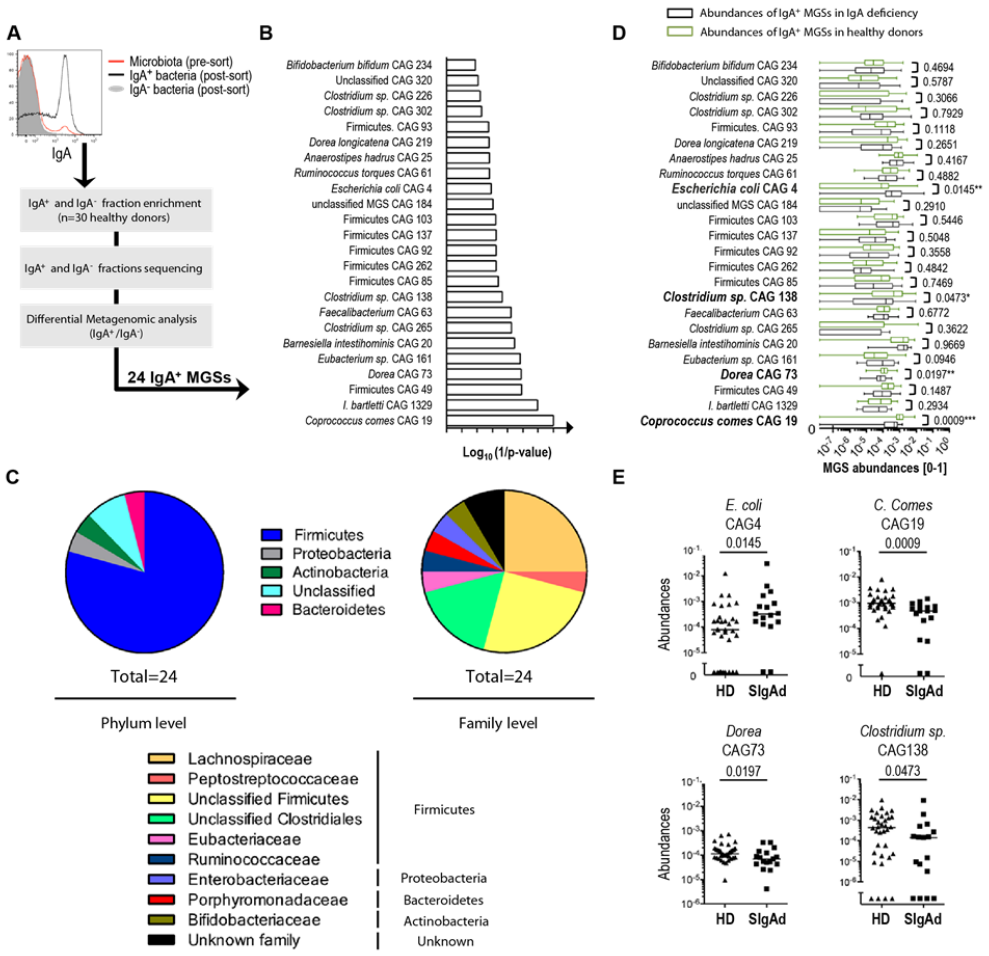

Figure 3: IgA$^+$ bacterial repertoire in healthy donors at MGS level

(A) IgA$^+$ and IgA$^-$ fractions of healthy donor gut microbiota (n=30) were enriched by magnetic sorting, sequenced and analyzed. (B) 24 indicated metagenomic species were overrepresented in the IgA$^+$ fractions. Bar diagram shows the 24 MGSs assigned at their lowest taxonomic level and ranked by the statistical significance with which it is found overrepresented in the IgA$^+$ fraction. (C) IgA$^+$ MGS represented at phylum (left) and family level (right). (D) IgA$^+$ MGS frequencies in healthy donors compared to IgA deficient patients. Ends of whiskers plots represent minimum and maximum values, vertical line represents medians and boxes represent data spreading (25% and 75% percentiles). (E) Abundances of the 4 IgA$^+$ MGS significantly different between healthy donors and IgA deficient patients. Horizontal lines represent medians and p-values were calculated with a Mann-Whitney test.

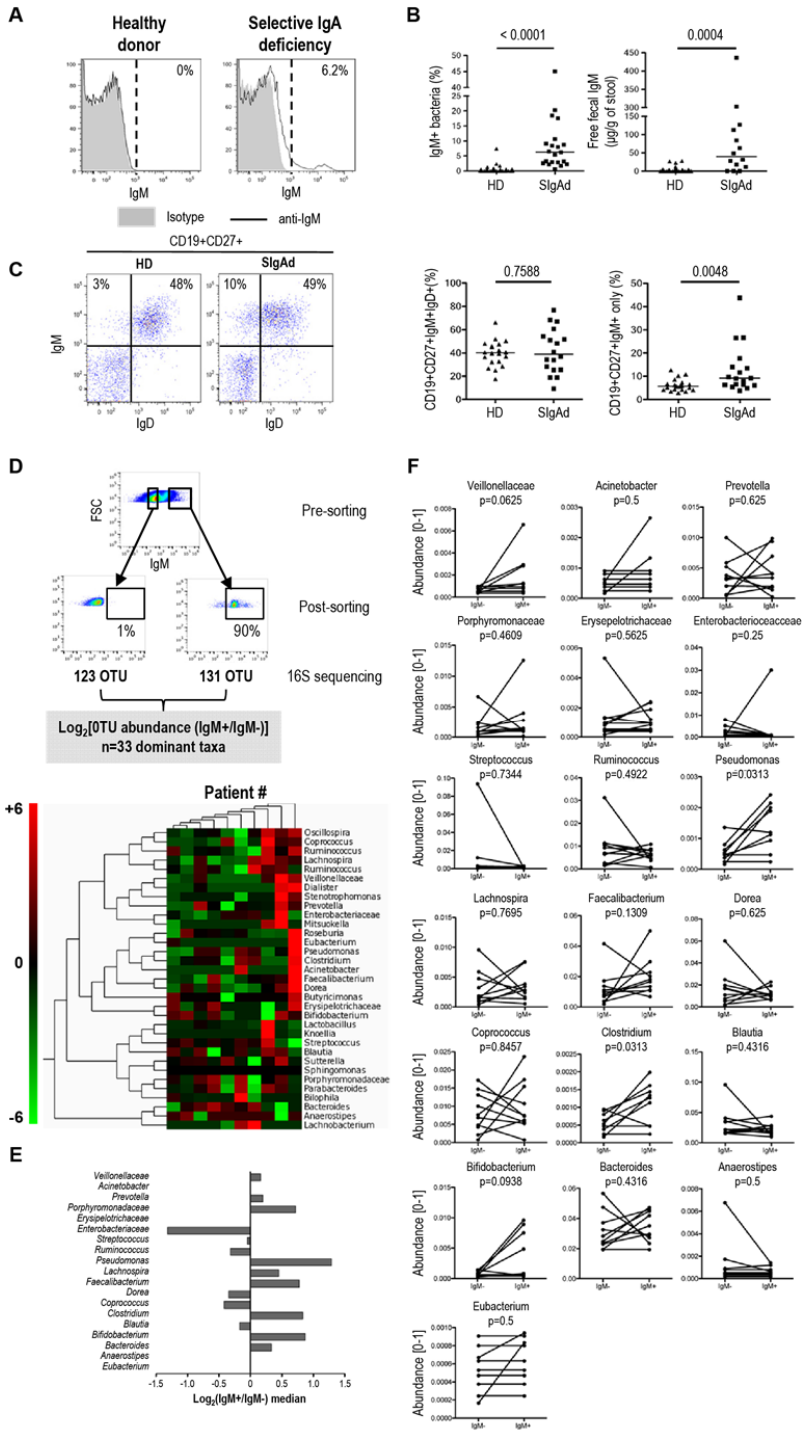

Figure 4: IgM-binding to IgA-deficient microbiota.

(A) Representative flow cytometry-based IgM detection on purified microbiota in a healthy donor and an IgA deficient patient. Grey histograms represent isotype controls and black lines IgM surface staining. (B) IgM binding levels in healthy donors (n=34) and SIgAd patients

(n=21) measured by flow cytometry (left). Free IgM levels in fecal waters measured by ELISA (right). (C) Peripheral $CD19^+CD27^+IgM^+IgD^+$ (marginal zone B cells) and $CD19^+CD27^+IgM^+IgD^-$ ("IgM only" B cells) frequencies in healthy donors and patients. (D) IgM^+ and IgM^- fractions of gut microbiota were sorted by flow cytometry in 10 patients and their composition was analyzed by 16S rRNA gene sequencing. Taxa enrichment in IgM^+ fraction is measured as the $\log_2[IgM^+/IgM^-]$ ratio of OTU abundances. The taxa enrichment measures of 33 dominant OTUs (rows) were grouped by hierarchical cluster analysis according to Ward's method and plotted as a heatmap (each column represents a patient). Green and red colors represent the lowest and highest ratios, respectively (range of the ratios: -6 to +6). (E) Medians of the IgM^+ enrichment ratio. Each horizontal bar represents the median ratio of the 10 donors by OTU. (F) Paired analysis of the abundances in IgM^+ and IgM^- fractions for each OTU of interest. In subfigure B and C horizontal bars represent medians and Mann-Whitney test was used to calculate p-values. In subfigure F p-values were calculated with a paired-rank Wilcoxon test.

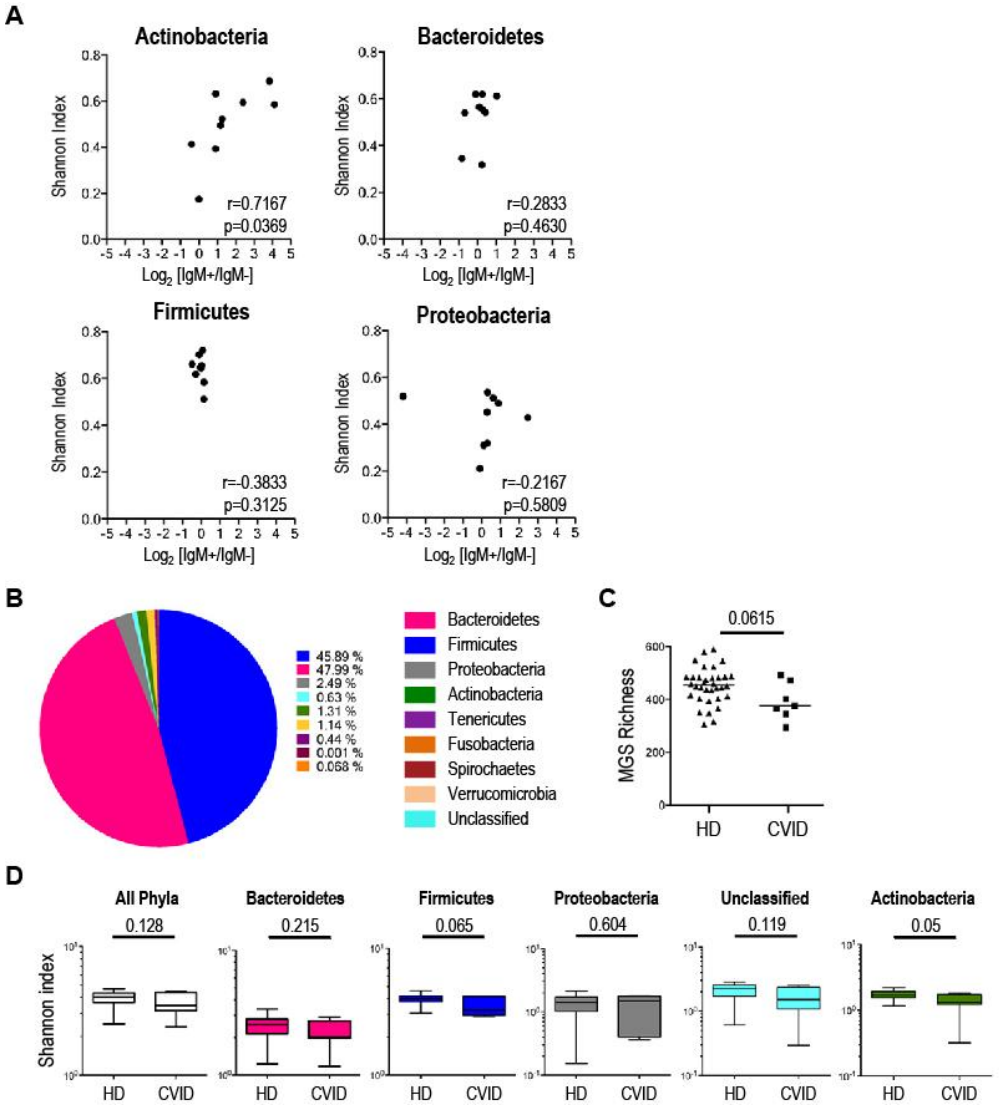

Figure 5: IgM binding and microbial diversity

(A) Scatter graphs represents the correlation between phylum enrichment in the IgM^+ fraction ($log_2([IgM^+]/[IgM^-])$, where [IgM*] represents phylum abundance in the IgM* fraction) and microbial diversity within a given phylum (Shannon diversity index) calculated from metagenomic sequencing (cf. Figure 1) in n=9 patients with selective IgA deficiency. Only dominant phyla are shown. Spearman coefficient (r) and p-values (p) are indicated. (B) Phyla distribution, (C) MGS richness and (D) diversity are shown for n=7 patients with common variable immunodeficiency. Diversity is calculated with Shannon's diversity index either for all the MGS (white whiskers plot) or within each phylum (pink: Bacteroidetes, blue: Firmicutes, cyan: unclassified, grey: Proteobacteria, green: Actinobacteria). Ends of whiskers represent the

minimum and maximum of all the data and boxes represent data spreading (1 standard deviation). Horizontal bars represent medians and p-values were calculated with a Mann Whitney test.

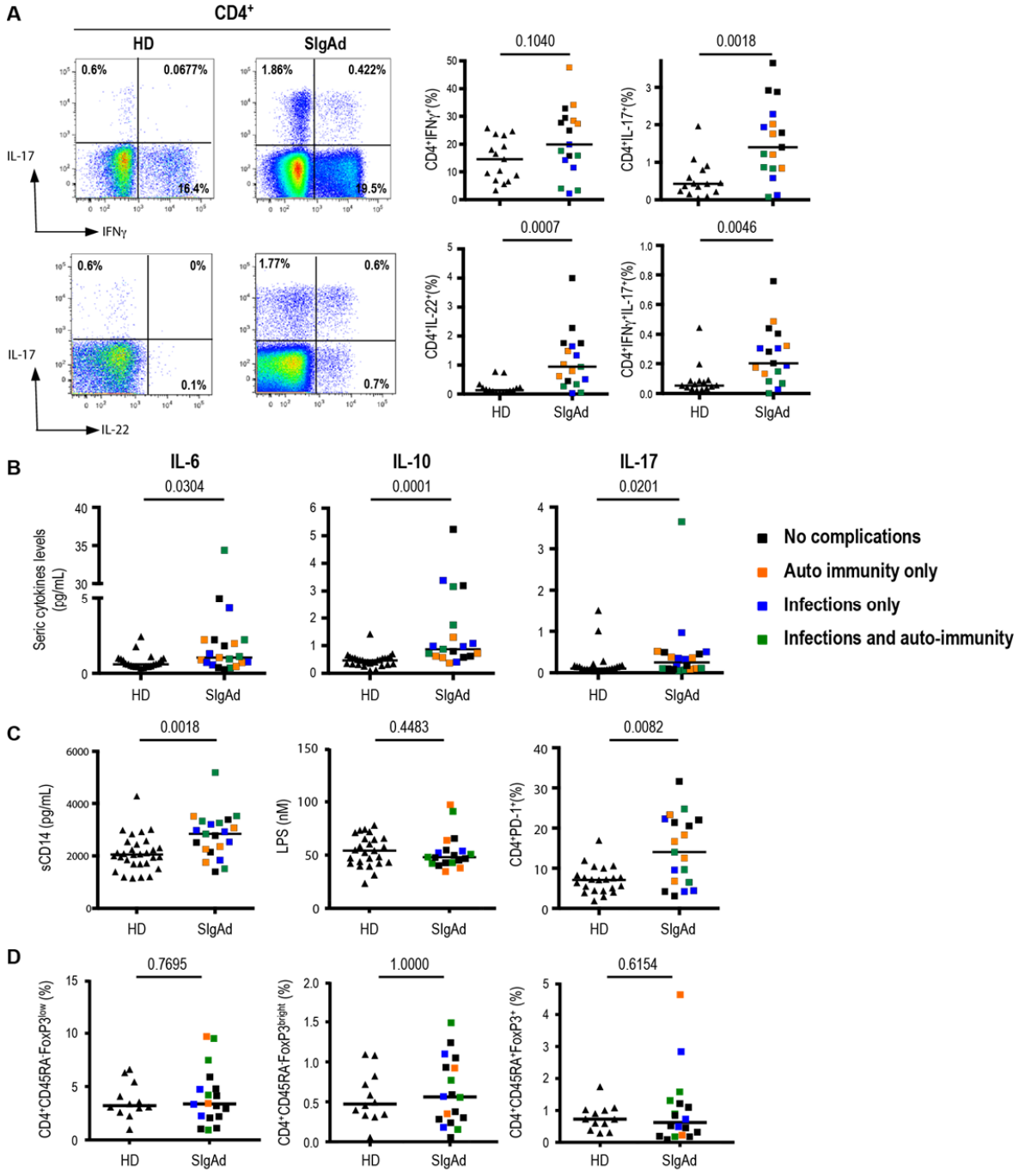

Figure 6: Systemic inflammatory and bacterial translocation markers

(A) Peripheral cytokine-secreting $CD4^+$ T cells after PMA-Calcium Ionophore stimulation (6h). On the left, intra-cellular detection of IFN-γ/IL-17 or IL-17/IL-22 in a healthy donor (left column) and a patient (right column). On the right, scatter dot plots showing percentages of peripheral $CD4^+IFN\text{-}\gamma^+$, $CD4^+IL\text{-}17^+$, $CD4^+IL\text{-}22^+$ and $CD4^+$ $IFN\text{-}\gamma^+IL\text{-}17^+$ cells in both groups.

(B) Seric IL-6, IL-10 and IL-17 levels (pg/mL) measured by SIMOA stratified according to donor status. (C) Seric soluble CD14 levels (pg/mL) were measured by ELISA (left), seric LPS levels (nM) were measured by mass spectrometry (middle) and peripheral PD-1 expressing $CD4^+$ T cells were detected by flow cytometry (right). (D) $CD4^+$ T regulatory cells were measured by flow cytometry, naïve Treg cells ($CD4^+CD45RA^+FoxP3^+$), effector Treg cells ($CD4^+CD45RA^-FoxP3^{bright}$), as well as activated conventional $CD4^+$ T cells ($CD4^+CD45RA^-FoxP3^{low}$). For all sections, horizontal bars represent medians and p-values calculated with a Mann-Whitney test. A color code indicates the main clinical features associated with each SIgAd case.

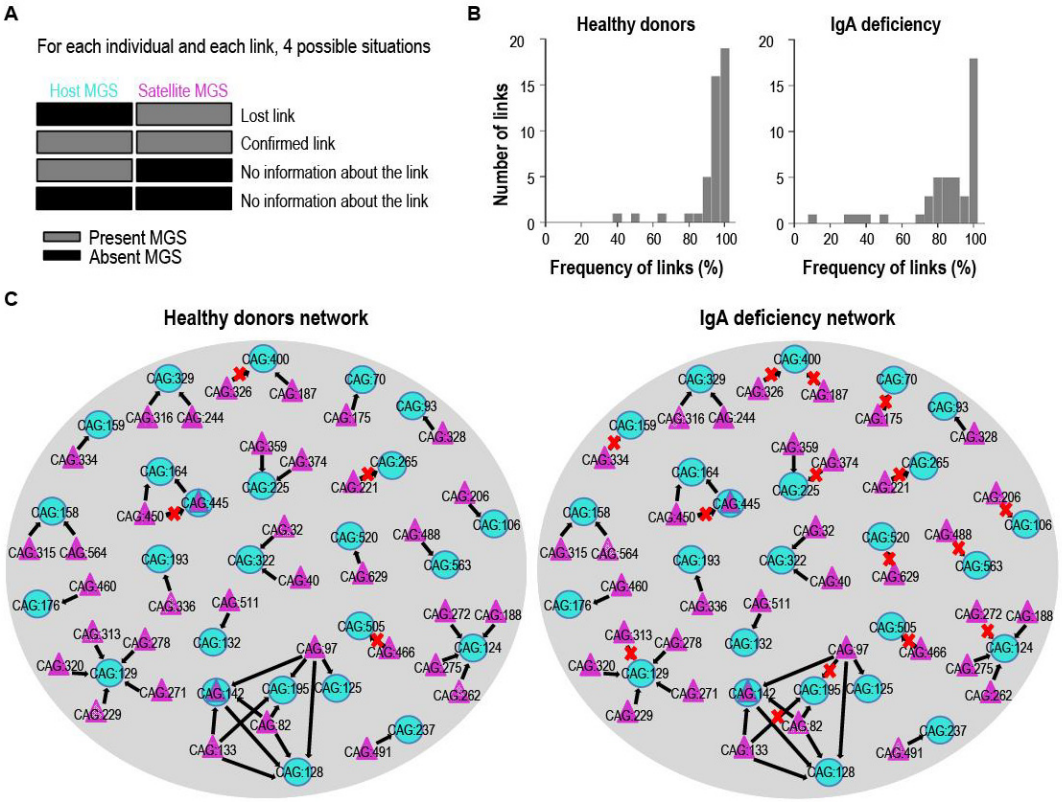

Figure 7: Bacterial dependency networks in IgA deficiency and healthy donors

(A) For each individual and each MGS-MGS dependency there are four possible situations. If the satellite MGS is present without corresponding host MGS, the link is considered lost. If the satellite and host are both present, the link is considered conserved. If the host is present in the absence of the satellite, or if both host and satellite are absent, no information is given about the link. (B) Distribution of the percentage of confirmed MGS-MGS dependency links. (C) MGS-MGS dependency-network compared between healthy donors (left) and IgA deficiency patients (right). Each of the two network maps shows the expected 45 highly significant and directional dependencies among 60 MGSs, as defined by Nielsen *et al.*[32]. Blue circles represent host MGSs. Purple triangles represent satellite MGSs. A purple triangle within a blue circle indicates both satellite and host status. Black arrows indicate the confirmed dependencies among MGSs (directional arrow from the satellite MGS to the host MGS). Red crosses indicate lost dependencies in healthy donors and IgA deficiency.

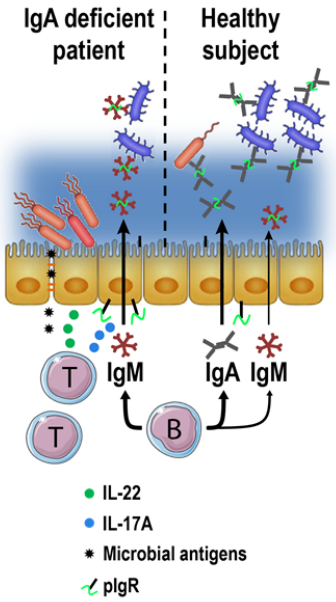

Figure 8: Ecological impact of secretory IgA (sIgA) binding on commensals (model). Healthy subjects (left) mainly bind fecal commensals with sIgA (residual sIgM found in the healthy intestine, this study). Healthy subjects also bind a diverse mucus-embedded commensal repertoire with IgA and IgM[52]. Selective IgA deficient patients (right) replace intestinal sIgA content with sIgM. sIgM binds efficiently species of the Firmicutes and Actinobacteria phyla, whereas Proteobacteria, such as E. coli, fail to be recognized by sIgM. In line with these findings selective IgA deficient patients display an increased abundance of E. coli (red bacteria) and reduced abundance of certain Firmicutes (blue bacteria), such as *Coprococcus comes*. Dysbiosis in IgAd patients is associated with a systemic inflammation status (increased Th17 response and pro-inflammatory cytokine secretion). We hypothesize that sIgA, as a first line check-point, may organize the spatial distribution of gut commensals and notably restrict pathobionts, such as *E. coli* from intimate contact with gut epithelium, thus prohibiting inflammatory responses, as well as aggregate beneficial microbes within the mucus layer in collaboration with IgM.

Title: Microbial ecology perturbation in human IgA deficiency

Authors: Jehane Fadlallah[1,2,†], Hela El Kafsi[1,2,†], Delphine Sterlin[1,2], Catherine Juste[3], Christophe Parizot[2], Karim Dorgham[1], Gaëlle Autaa[1], Doriane Gouas[1,2], Mathieu Almeida[4], Patricia Lepage[3], Nicolas Pons[5], Emmanuelle Le Chatelier[3,5], Florence Levenez[5], Sean Kennedy[5], Nathalie Galleron[5], Jean-Paul Pais de Barros[6,7], Laurent Lagrost[6,7], Lionel Galicier[8], David Boutboul[8], Alexis Mathian[2], Makoto Miyara[2], Eric Oksenhendler[8], Zahir Amoura[1,2], Joel Doré[3], Claire Fieschi[8], S. Dusko Ehrlich[3,5], Martin Larsen[1,2,‡,*], Guy Gorochov[1,2,‡,*]

Affiliations:
[1]Sorbonne Universités, UPMC Université Paris 6, INSERM, CNRS, Centre d'Immunologie et des Maladies Infectieuses (CIMI Paris UMRS 1135), 75013, Paris, France.
[2]AP-HP, Groupement Hospitalier Pitié-Salpêtrière, Département d'Immunologie, 75013, Paris, France.
[3]UMR1319 Micalis, INRA, Jouy-en-Josas, France.
[4]Center for Bioinformatics and Computational Biology, University of Maryland, Paint Branch Road, College Park, MD 20742, USA.
[5]INRA, Institut National de la Recherche Agronomique, US1367 MetaGenoPolis, 78350 Jouy en Josas, France.
[6]INSERM, LNC UMR866, University Bourgogne Franche-Comté, F-21000 Dijon, France
[7]LipSTIC LabEx, Fondation de Coopération Scientifique Bourgogne-Franche Comté, F-21000 Dijon, France
[8]Université Paris Diderot Paris 7, Department of Clinical Immunology, Hôpital Saint-Louis, Assistance Publique Hôpitaux de Paris (APHP), EA3518, 75010, Paris, France
[†]These authors contributed equally to this work
[‡]These authors jointly directed this work

*To whom correspondence should be addressed: Martin Larsen (martin.larsen@upmc.fr) or Guy Gorochov (guy.gorochov@upmc.fr).

Supplementary Materials:

Materials and Methods

Serum Ig levels:

Seric Ig levels were assessed by nephelometry using the Siemens BN2 Nephelometer (Siemens, Camberley, UK). IgA detection threshold was 0.07 mg/mL, this value being the admitted threshold to define IgA deficiency[28].

Cytokine quantification:

Cytokine levels were measured using Simoa technology[59], reagents and procedures were obtained from Quanterix Corporation. In the first step of the assay, capture beads coated with an anti-cytokine antibody were combined with the serum. After washing, the biotinylated detector antibody was added to the reaction to bind the captured cytokine. Following a second wash, the streptavidin-β-galactosidase (SBG) was added to bind the detector antibody resulting in enzyme labeling of captured cytokine. After washing, the beads were resuspended in a resorufin β-D-galactopyranoside (RGP) substrate solution and transferred immediately into the array of a Simoa disc to be individually captured into the microwells. The ß-galactosidase on captured cytokine hydrolyzed the RGP substrate into a fluorescent product that provides the signal for measurement. At low cytokine concentration, the percentage of bead-containing wells in the array that have a positive signal is proportional to the amount of cytokine present in the sample (digital measurement). At higher concentration, when most of the bead-containing wells have one or more labeled cytokine molecules, the total fluorescence signal is proportional to the amount of cytokine present in the sample (analog measurement). Concentration of cytokines in serum samples was interpolated from standard curves.

Table S1. Individual clinical description of selective IgA deficient patients.

Patient ID	Phenotype	Associated diseases	Infections	Digestive symptoms	Diagnosis onset	Subclass deficiency	IgG (mg/ml)
SIgAd01	No complications	No	No	No	2012	No	14.6
SIgAd02	Infections only	No	Respiratory	No	2011	No	11.4
SIgAd03	Auto immunity	Type 1 diabetes/Celiac disease	No	Yes, diarrhea	2008	No	14.6
SIgAd04	Auto immunity	Celiac disease/Vitiligo	No	Yes, diarrhea	2012	IgG4	8.43
SIgAd05	Infections only	No	Respiratory	No	2012	IgG4	12.9
SIgAd06	Auto immunity and infections	SLE	Digestive	Yes, diarrhea	2007	No	15.9
SIgAd07	Auto immunity and infections	Thyroiditis	Respiratory	Yes, diarrhea	2013	No	11.8
SIgAd08	Auto immunity	Thyroiditis	No	No	2013	No	10.3
SIgAd09	No complications	No	No	No	Not available	No	7.14
SIgAd10	No complications	No	No	No	Not available	No	10.7
SIgAd11	Auto immunity and infections	Thyroiditis/Spondylarthitis	Respiratory	No	Not available	No	7.56
SIgAd12	Auto immunity	SLE/cytopenia	No	No	2011	No	16.7
SIgAd13	No complications	No	No	No	2010	No	10
SIgAd14	No complications	No	No	No	2008	No	15.3
SIgAd15	Auto immunity and infections	Cytopenia/Gastritis/Thyroiditis	Respiratory	Yes, diarrhea	>20 years	No	7.94
SIgAd16	Infections only	No	Digestive	Yes,Diarrhea	2012	No	17
SIgAd17	No complications	No	No	No	2013	No	18
SIgAd18	Auto immunity	Thyroiditis/Gastritis	No	Yes, diarrhea	2012	No	7.61
SIgAd19	Auto immunity and infections	SLE	Respiratory	No	2008	No	7.09
SIgAd20	Auto immunity and infections	SLE/cytopenia	No	Yes, diarrhea	1990	No	4.68
SIgAd21	No complications	No	No	Yes, diarrhea	2012	No	12.5

SIgAd18-21 were excluded from metagenomic analysis due to recent anti-biotherapy (within past 3 months), but were phenotyped immunologically (cf. figure 6).

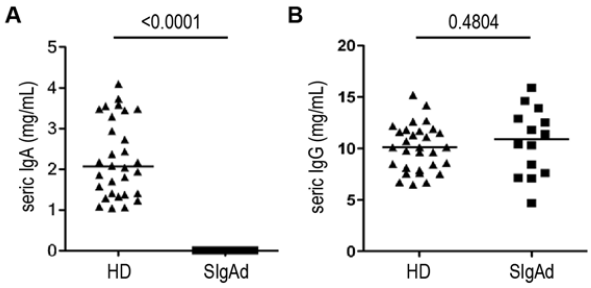

Figure S1: Serum IgA and IgG levels in IgA deficient patients

(A) IgA and (B) IgG levels were assessed in healthy donors (HD) and selective IgA deficient patients (SIgAd) by nephelometry, with a detection threshold of 0.07 mg/mL for IgA. Horizontal lines represent median values and p-values are calculated with a Mann-Whitney test.

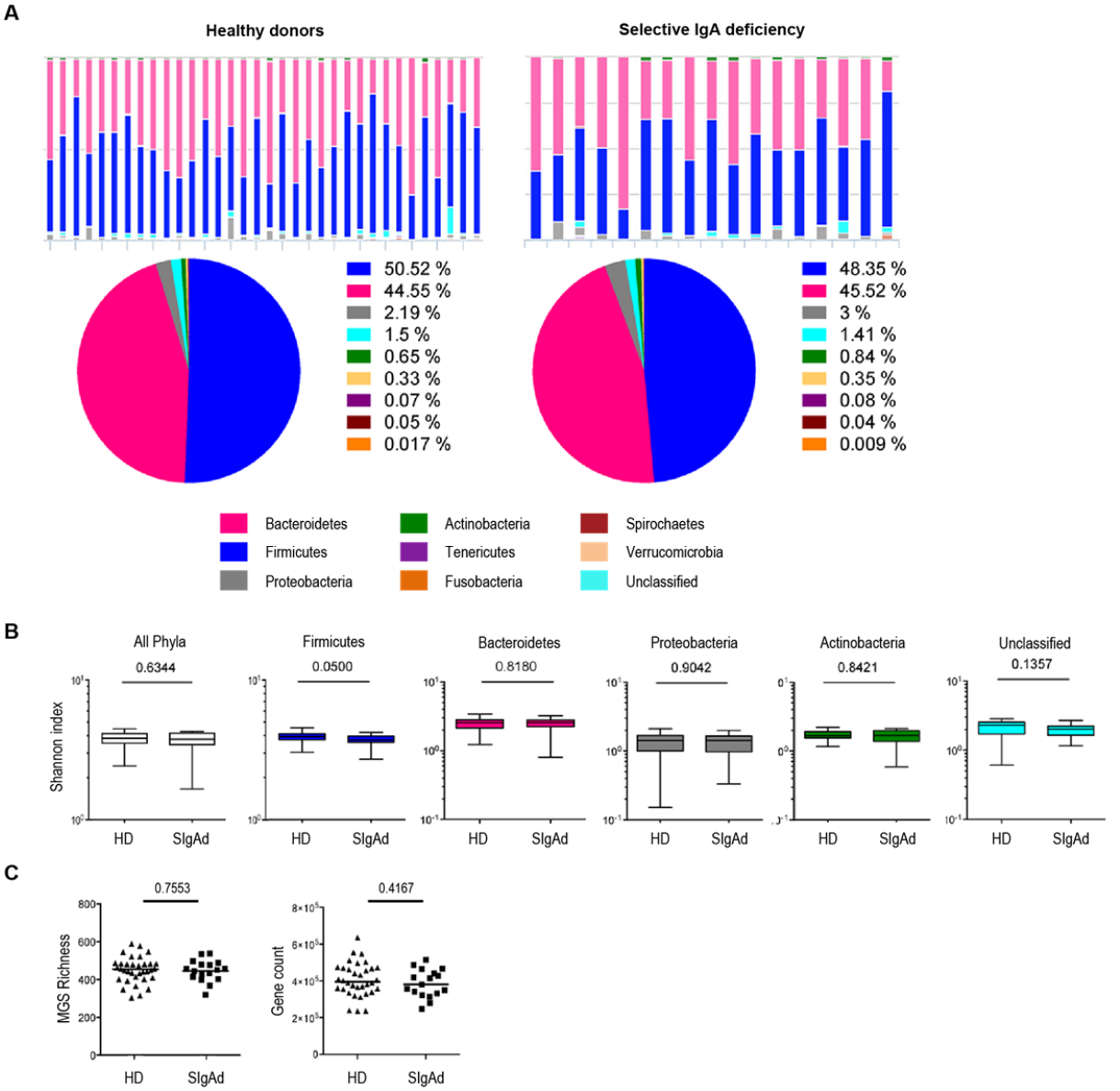

Figure S2A: Global gut microbiota composition

(A) Metagenomic analysis performed at phylum level performed on stool samples from healthy donors (n=34) and selective IgA deficient patients (n=17). Individual donor histograms and pie charts stratified to donor status. Each color represents a major phylum (pink: Bacteroidetes, blue: Firmicutes, grey: Proteobacteria, green: Actinobacteria, cyan: unclassified). (B) Global diversity at metagenomic species level (white whiskers plots) and specific diversity within each dominant phylum were calculated as the Shannon diversity (Shannon index) in both donor groups. Ends of whiskers represent minimum and maximum values and boxes represent data spread (25% and 75% percentiles). (C) Metagenomic species (MGSs) richness (left) and gene richness (gene

count, right) were calculated for each group as the number of different MGSs or genes for each donor (MGSs were considered present if at least one read was detected in a given donor). Horizontal bars represent medians. Indicated p-values were calculated using a Mann Whitney test.

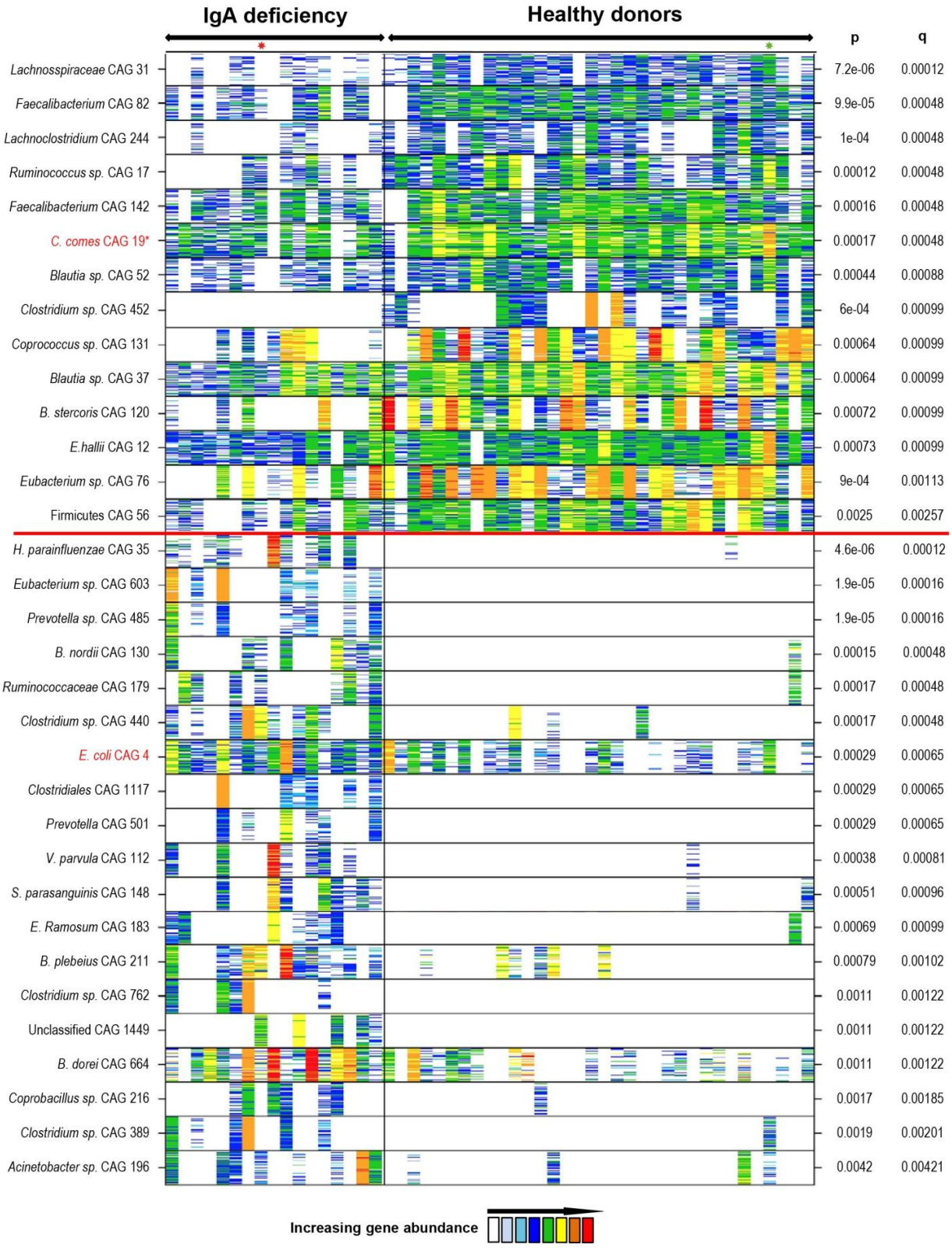
IgA deficiency
Healthy donors
p
q
Lachnosspiraceae CAG 31 7.2e-06 0.00012
Faecalibacterium CAG 82 9.9e-05 0.00048
Lachnoclostridium CAG 244 1e-04 0.00048
Ruminococcus sp. CAG 17 0.00012 0.00048
Faecalibacterium CAG 142 0.00016 0.00048
C. comes CAG 19* 0.00017 0.00048
Blautia sp. CAG 52 0.00044 0.00088
Clostridium sp. CAG 452 6e-04 0.00099
Coprococcus sp. CAG 131 0.00064 0.00099
Blautia sp. CAG 37 0.00064 0.00099
B. stercoris CAG 120 0.00072 0.00099
E.hallii CAG 12 0.00073 0.00099
Eubacterium sp. CAG 76 9e-04 0.00113
Firmicutes CAG 56 0.0025 0.00257
H. parainfluenzae CAG 35 4.6e-06 0.00012
Eubacterium sp. CAG 603 1.9e-05 0.00016
Prevotella sp. CAG 485 1.9e-05 0.00016
B. nordii CAG 130 0.00015 0.00048
Ruminococcaceae CAG 179 0.00017 0.00048
Clostridium sp. CAG 440 0.00017 0.00048
E. coli CAG 4 0.00029 0.00065
Clostridiales CAG 1117 0.00029 0.00065
Prevotella CAG 501 0.00029 0.00065
V. parvula CAG 112 0.00038 0.00081
S. parasanguinis CAG 148 0.00051 0.00096
E. Ramosum CAG 183 0.00069 0.00099
B. plebeius CAG 211 0.00079 0.00102
Clostridium sp. CAG 762 0.0011 0.00122
Unclassified CAG 1449 0.0011 0.00122
B. dorei CAG 664 0.0011 0.00122
Coprobacillus sp. CAG 216 0.0017 0.00185
Clostridium sp. CAG 389 0.0019 0.00201
Acinetobacter sp. CAG 196 0.0042 0.00421
Increasing gene abundance

Figure S2B: Differential CAG abundance analysis between healthy subjects and IgA deficient patients.

CAGs enriched in IgA deficiency (bottom left quadrant, n=17 patients) and in healthy donors (top right quadrant, n=34) fractions are separated by a horizontal red line. Genes are in rows, abundance is indicated by color gradient (white, not detected; red, most abundant); the enrichment significance is shown (p-values were calculated with a Wilcoxon test and corrected p-values (q) were calculated with the method of Benjamini and Hochberg). Individuals are shown in columns. IgA deficient patients and healthy donors are separated by a vertical black line. The differentially abundant CAGs are ordered by decreasing p and q values. The healthy donor with non-detectable IgA binding of gut microbiota and the IgA deficient patients with detectable IgA binding of gut microbiota are indicated with a green and red star respectively.

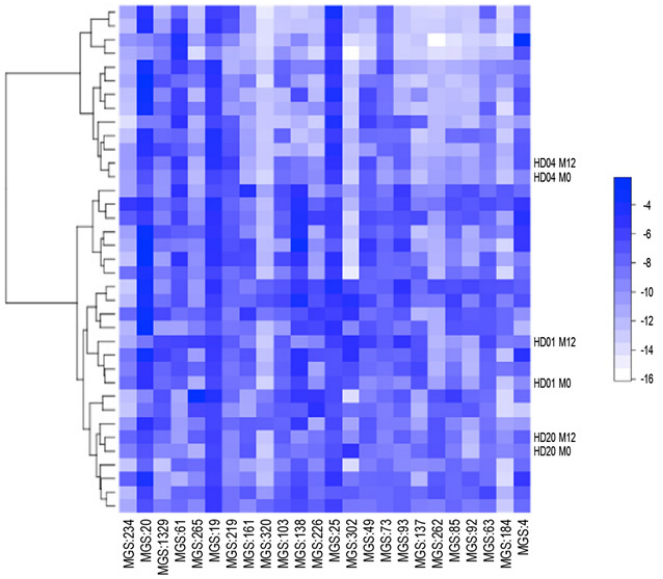

Figure S2C: Longitudinal stability of 24 IgA^{+} MGS abundances in 3 healthy subjects.

Heatmap depicting the abundance of the 24 IgA-bound MGS (columns) in 37 healthy donor samples (rows) including 3 donors sampled twice over an interval of 12 months. Samples were grouped by hierarchical cluster analysis according to Ward's method (dendrogram). White-to-blue color gradient represents low and high ratios (cf. color code).

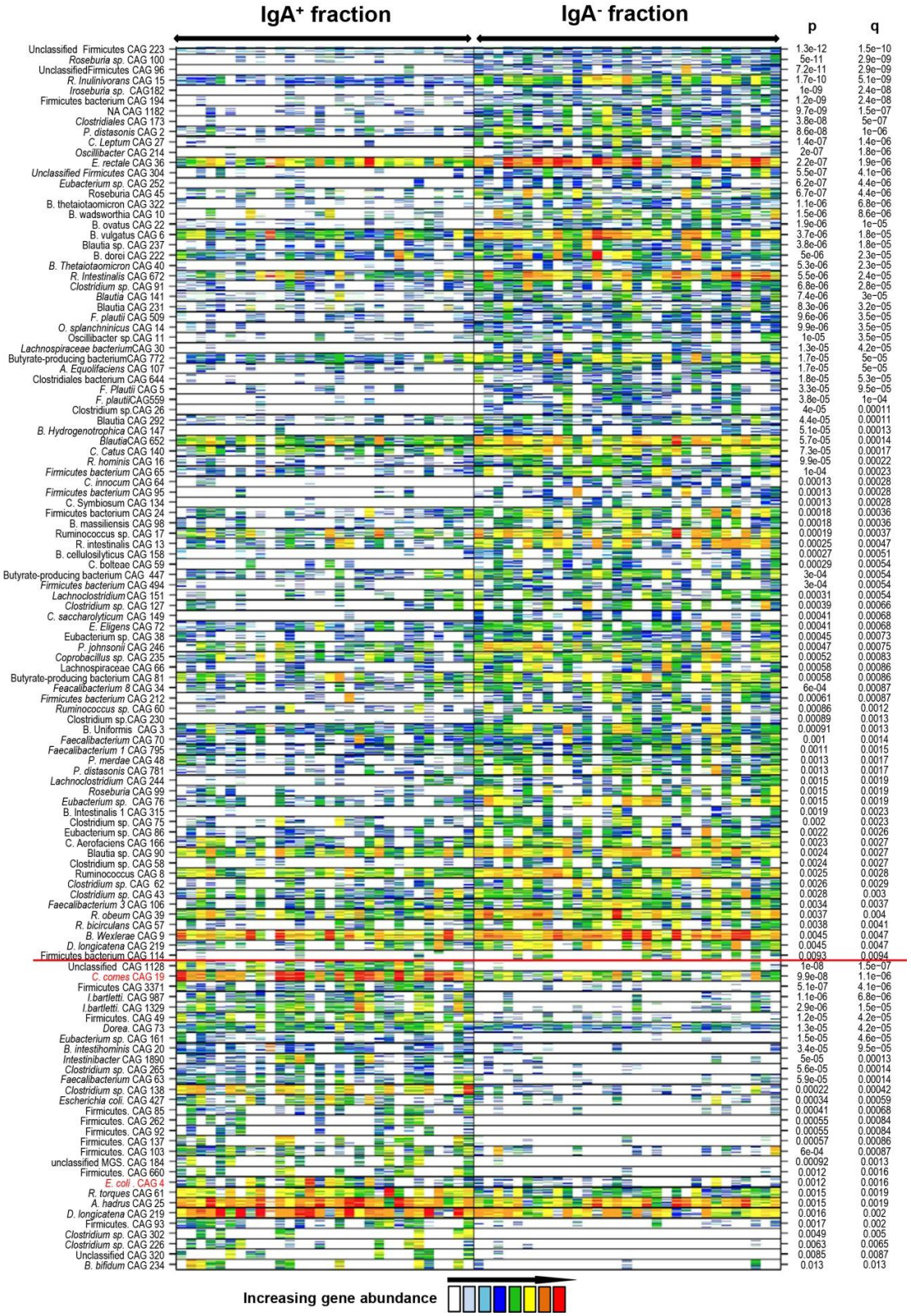

Figure S3: Differential CAG abundance analysis between IgA$^+$ and IgA$^-$ microbes.

Abundance of 50 CAG ‘tracer’ genes in the IgA^+ (n=30) and IgA^- (n=30) fractions. Genes are in rows, abundance is indicated by colour gradient (white, not detected; red, most abundant); the enrichment significance is shown (p-values were calculated with a Wilcoxon test and corrected p-values (q) were calculated with the method of Benjamini and Hochberg). Individuals are shown in columns. IgA^+ and IgA^- fractions are separated by a vertical black line. The differentially abundant CAGs are ordered by decreasing p and q values. CAGs enriched in IgA^+ (bottom left quadrant) and in IgA^- (top right quadrant) fractions are separated by a horizontal red line.

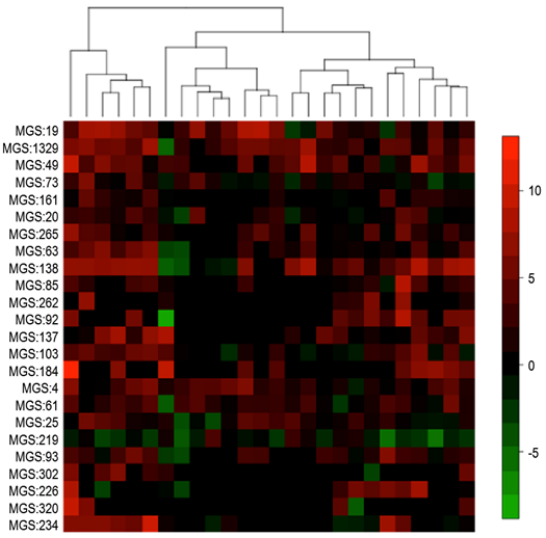

Figure S4: MGS enrichment analysis between IgA$^+$ and IgA$^-$ microbes.

Abundance of 24 MGS bound by sIgA in stool samples from healthy donors (cf. Figure 2). MGS enrichment in IgA$^+$ fraction is measured as the log2[IgA+/IgA-] ratio of MGS abundances and depicted as a heatmap with each line representing an MGS and each row a donor. Donors were grouped by hierarchical cluster analysis according to Ward's method. Green and red colors represent low and high ratios, respectively (cf. color code).

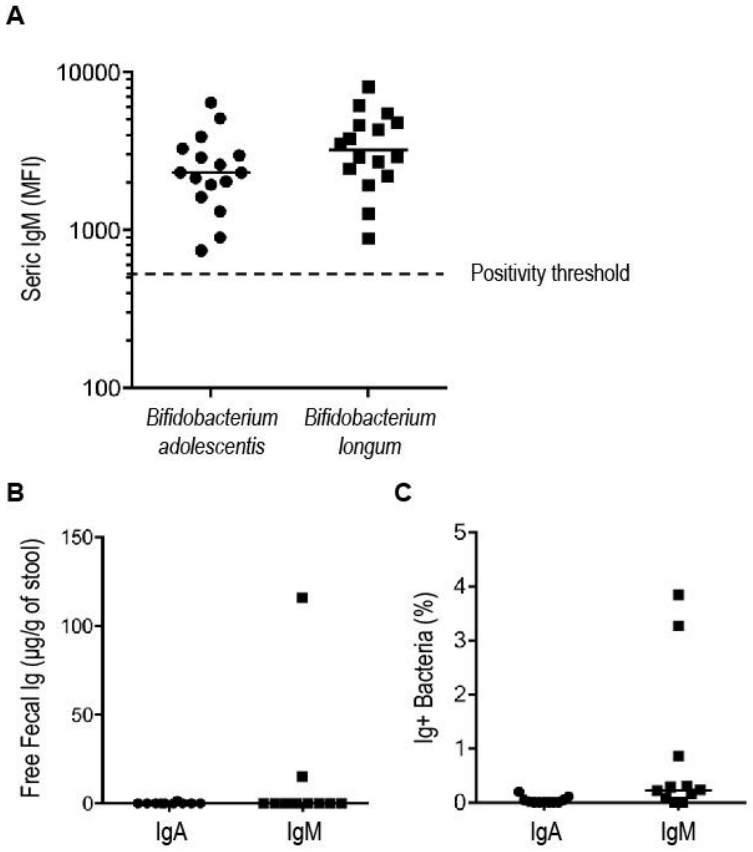

Figure S5: Gut microbiota IgA and IgM binding in CVID patients with complete IgA deficiency.

(A) Serum IgM titers against *Bifidobacterium adolescentis* and *Bifidobacterium longum* were measured by flow cytometry in 16 healthy donors. Positivity threshold is defined as the isotype control staining MFI level. (B) Free IgA and IgM levels in fecal water measured by ELISA. (C) Flow cytometry detection of IgA and IgM binding in CVID patients with complete IgA deficiency (n=15). Horizontal lines represent medians, p values are calculated with a Mann-Whitney test.